Membrane Proteomics

Characterization of Brush Border membrane proteins of mice intestinal mucosa.
Case study: cholesterol absorption

Dissertation
zur Erlangung des Doktorgrades
der Naturwissenschaften

vorgelegt beim Fachbereich
Biochemie, Chemie und Pharmazie
der Johann Wolfgang Goethe-Universität
in Frankfurt am Main

von Eirini Tsirogianni
aus Athen (Griechenland)

Frankfurt (2008)

Vom Fachbereich Biochemie, Chemie und Pharmazie der
Johann Wolfgang Goethe-Universität als Dissertation angenommen.

Dekan: Prof. Dr. D. Steinhilber

Gutachter: Prof. Dr. M. Karas
                          PD Dr. H. Langen

Datum der Disputation: 18.05.2009

To Axel and Nephelie

(Where there is a will, there is a way)

Table of content

Summary

The epithelial absorbing cells of the small intestinal villi, the enterocytes, are the main protagonists for the transport of nutrients from the intestinal lumen to the interstitial fluids. The oriented flow of nutrients is carried out by different and complementary transport systems present in the apical and the basolateral domains of the enterocyte's plasma membrane. One of the distinctive characteristics of those intestinal cells is the presence of numerous structurally distinct protrusions (referred as microvilli) on the apical surface of the plasma membrane. They confer the brush-like appearance of the microvillus border (commonly referred to as the "brush border") typically observed in the light microscope.

Over the years, there has been considerable interest to study the molecular mechanisms driving the transport of molecules across the intestinal brush border membrane (BBM). Defects have been described to cause a variety of pathological conditions, such as disorders in the metabolism of saccharides (glucose and galactose malabsorption, lactose intolerance), amino acids (Hartnup disease, aminoacidurias), ions (sodium and potassium in the case of familiar diarrhea), metals (zinc in acrodermatitis enteropathica) and cholesterol lipids (cardiovascular diseases). In particular, the essential role of the BBM in regulating the delicate balance between cholesterol influx and efflux from the lumen to the enterocyte has been recently highlighted through the genetic analysis of individuals suffering of cholesterol disorders as well as in several clinical studies involving the use of dietary plant sterols (phytostrerols) or specific protein inhibitors blocking essential components of the cholesterol absorption/resorption pathway.

Most studies aimed at studying the enterocyte BBM have been conducted through gene analysis, activity tests and immunology assays. Its characterization at the protein level, however, has been hampered by the difficulty to isolate a pure fraction from the small intestine and by the nature of the targets of interest, being most of them transmembrane integral proteins. Predictably, the few proteomics studies reporting the identification of proteins localized in the BBM membrane failed to identify transporters and receptors that are known to be located in this fraction. Also, the very aggressive environment to which the enterocyte is exposed combined with the class of proteolytic enzymes present in this membrane imply the use of targeted strategies to maintain the structural integrity of the BBM proteins in the isolation protocol.

The primary goal of this study was the development and the evaluation of mass spectrometry-based analytical strategies for the analysis and the identification of hydrophobic proteins, such

as found in the enterocyte BBM. Analysis of membrane proteins in a complex mixture has always represented an obstacle in the proteomics field. Their hydrophobic nature requests a detergent for their solubilization and makes the use of traditional proteomics techniques less compatible (e.g. two-dimensional gel electrophoresis) while their low abundance in complex biological mixtures (in comparison to soluble proteins) requires additional purifications steps for their enrichment. Even at equal abundance, the identification rate of membrane proteins is lower than that of soluble proteins. The development of an optimized sample preparation protocol combined with a robust analytical method was expected to enable the investigation of proteins located in the BBM and so to confirm the findings of several other studies, namely, that the BBM contains proteins that participate in cholesterol regulation in the enterocytes. Finally, a third goal of this study was to investigate conditions for which a quantitative mass spectrometric experiment using a label-free strategy could be used to investigate biological samples, as technical variability and sample stability of the BBM preparation were not known. The reproducibility of the technical steps had to be monitored for the whole workflow, from the BBM preparation until the mass spectrometric analysis of the samples, so to pinpoint difficulties and limitation of the workflow and preparing the way for a label free quantification strategy using only the information from the LC-MS data.

In this study, an analytical strategy enabling for the first time (to my knowledge) the direct characterization of BBM proteins previously described to participate in cholesterol absorption is presented. Key points of this strategy consisted of an improved protocol to reproducibly isolate and purify BBM preparations from the intestinal tissue, the design of an inhibitor cocktail specifically aimed at minimizing the proteolytic activity of the BBM endogeneous proteases, a targeted strategy to enrich the BBM transmembrane proteins from common cytosolic contaminants, and the use of a very robust and sensitive instrumentation using capillary liquid chromatography and tandem mass spectrometry. In particular, the original protocol from Kessler et al for purifying and enriching the proteins contained in the BBM fraction was significantly improved and extended to selectively remove cytosolic and basolateral contaminants. Whole intestine, or scrapped mucosa thereof for higher purity, was first lysed and a crude membrane fraction was isolated by differential centrifugation. The resulting pellet was then resuspended and the basolateral membrane fraction was specifically removed from the preparation by $CaCl_2$ precipitation. The enriched BBM vesicles were then pelleted by centrifugation and subjected to additional washes in high salt and high pH to remove the remaining cytosolic and membrane-associated proteins. The performance of the protocol and the purity of the obtained BBM vesicles were monitored by Western Blot

analysis following a known basolateral (Na^+/K^+ ATPase a1) and BBM markers (FATP-4). The new protocol provided a 30- to 50-fold enrichment factor (versus about a 10-fold enrichment factor following the Kessler protocol) compared to the crude lysate while a known basolateral marker, Na^+/K^+ ATPase a1, remained undetected in the purified BBM fraction. Simultaneously, a significant part of the protocol optimization was devoted to the inhibition of the endogenous BBM vesicles proteases, most of them of the His-Zn dependent metalloprotease superfamily that are not inhibited by common Serine, Cysteine proteinase inhibitors. Partial inhibition (approximately 90%) was achieved by adding amastatin, a known partial inhibitor, and several peptide substrates in all buffers used in the BBM purification protocol. In addition, excess calcium (a known activator of this family of proteases) was removed by adding EDTA in the first wash step immediately following the calcium precipitation step.

A tripilicate analysis of a purified mouse BBM fraction resulted in the reliable identification of 1460 proteins, of which 260 proteins were predicted to be transmembrane integral proteins. A detailed GO analysis revealed that the proteins identified in this study were equally distributed between plasma membrane, the ER/Golgi/endosome compartments, cytoskleleton and mitochondria while only a quarter of the BBM proteins were annotated as cytoplasmic. Using this protocol, a number of proteins known to play a critical role in cholesterol absorption were identified directly for the first time at the protein level. For example, the Niemann-Pick C1-like 1 protein (NPC1L1), an abundant protein of the BBM preparation, is a 13-transmembrane segments protein described to play a critical role in cholesterol absorption. Similarly, the two half-size, 6-transmembrane segments ABC transporters ABCG5 and ABCG8 are involved in the biliary secretion of cholesterol and plant sterols. The complex formed by Caveolin-1 and Annexin-2 has been suggested as key element for the cholesterol trafficking from the BBM to the endoplasmic reticulum while SR-BI, a 4-transmembrane segments receptor, has been found to be involved in cholesterol uptake. The role of many of the proteins mentioned above and of several other proteins, such as CD36, Galectin-4, and ABCB1, is the subject of hot debates. Several studies have supported their involvement in cholesterol absorption but their precise mechanisms of action have remained unclear so far. Finally, it is worth mentioning here the presence of three lipoproteins, ApoA-I, ApoA-IV and ApoE, within the identified proteins of the BBM preparation. Their confident identification in this membrane preparation was rather unexpected as lipoproteins are by nature small, soluble proteins secreted by the liver and, therefore, they are not considered as constituents of the

BBM. Rather, their presence might be due to their tight interaction with some BBM constituents, such as LRP-1 and Cubilin, which are known to interact with apoliproproteins.

The complexity of the present proteomics workflow raised the question whether it was possible to reproducibly and quantitatively survey hundreds of membrane proteins simultaneously in the enterocyte BBM vesicles. Some key parts of the study were therefore analyzed in greater details using the protein identification information (comparing the successful MS/MS analysis between bands) and comparative analysis of the precursor ion signals to investigate in a more systematic manner the factors weighting in the reproducibility of the overall analysis. Using both strategies, most of the technical steps, such as the LC-MS/MS identification strategy, the separation of complex protein mixtures by 1-D-SDS-PAGE, or the in-gel digestion of proteins using trypsin, were not identified as major contributors to the overall variability of the experiment if appropriately controlled. Taken as a whole, the systematic evaluation of the overall analytical process unambiguously confirmed the high reproducibility achieved by the LC-MS/MS process. Rather, the BBM isolation protocol itself was identified as a potential source of variability due to its relative length and complexity and due to the inclusion of several steps that might have been difficult to carry on quantitatively, such as the $CaCl_2$ precipitation step. Also, the remaining proteolytic activity of the abundant BBM proteases, if unchecked, could also contribute to sample degradation and add extensive variability to the protein identification process. Finally, the variability in the number of commonly identified proteins was significantly lower when samples were compared to each other using a proper design of experiment.

In conclusion, this study demonstrates the feasibility to reproducibly and quantitatively analyze membrane proteins in complex mixture such as isolated from the enterocyte BBM. Key elements for a successful analysis were a robust sample preparation protocol yielding highly enriched BBM vesicles, a tightly controlled analytical strategy, and a statistically driven data analysis scheme. In particular, and most importantly, in this BBM vesicle preparation highly enriched for membrane proteins, the number of different peptides and the average sum of peptide counts were reflective for the relative abundance of a given membrane protein in the preparation, independently of its number of transmembrane segments. However, those identified tryptic peptides were exclusively located within the loops or in the cytoplasmic regions of membrane proteins. Assuming that trypsin cleaves off transmembrane helices, that is, the transmembrane tryptic peptide is generated during the digestion procedure, the rather hydrophobic nature and the length (30-40 amino acids in average) of most of those peptides may hinder their extraction from the gel band or they may stick tube walls during

peptide extraction. Finally, very long and/or hydrophobic peptides might not be amenable to the standard RP-LC-MS/MS conditions used in this study. In this respect, the use of a different stationary phase, such as HILIC, might provide additional and complementary peptide information.

Finally, this study opens the way for additional proteomics experiments focused on the BBM biology that will contribute and complement existing studies about inhibition of cholesterol absorption that have been focused so far at the gene expression level. In a preliminary experiment, the BBM preparations from an ApoE knockout mice (one of the most widely used mouse models to study dislipidemia in which the targeted deletion of the *apoE* gene leads to severe hypercholesterolemia and spontaneous atherosclerosis) and a wild type mouse of the same genetic background were compared. In the absence of biological replicates, only "black and white" differences were considered. Most interestingly, the Ileal Bile Acid Transporter (IBAT) protein and the ApoAI protein, which were robustly identified in the wild type animals, couldn't be detected in the ApoE knockout mice. This finding was not described in any of the earlier published studies and strongly suggests a disruption of the bile acids metabolism in the knockout animal. In conclusion, the analytical strategy described in this study was shown sufficiently mature to perform comprehensive comparative analysis of mice that have, for example, been treated with specific compound or subjected to different diets. In due course, this study could has been followed by a full fledge proteomics study in which a much more comprehensive biological experiment could have been investigated, such as a control mouse vs statins vs ezetimibe treatment, and where the impact of those drugs in the BBM (unknown at present) could have been investigated in more detail.

Zusammenfassung

Die absorbierenden Epithelzellen der Dünndarmzotten, die Enterozyten, spielen beim Transport der Nährstoffe vom intestinalen Lumen zu den interstitiellen Flüssigkeiten eine zentrale Rolle. Der gerichtete Fluss von Nährstoffen wird durch verschiedene komplementäre Transportsysteme gewährleistet, welche in den apikalen und basolateralen Domänen der Plasmamembrane der Enterozyten angesiedelt sind. Eine der besonderen Eigenschaften dieser Zellen ist das Vorkommen von zahlreichen Ausbuchtungen mit charakteristischer Struktur, sogenannten Mikrovilli, auf der apikalen Oberfläche der Plasmamembran. Diese verleihen dem Mikrovillusrand (auch als „Bürstensaum" bezeichnet) das bürstenartige Aussehen, das im Lichtmikroskop typischerweise sichtbar ist.

Es besteht seit Jahren ein starkes Interesse an der Erforschung der molekularen Mechanismen, die für den Transport der Moleküle durch die intestinale Bürstensaum-Membran verantwortlich sind. Defekte in diesen Transportmechanismen können eine Vielzahl von pathologischen Problemen auslösen, wie zum Beispiel Störungen im Metabolismus von Sacchariden (Glukose- und Galaktose-Fehlabsorption, Lactoseunverträglichkeit), von Aminosäuren (Hartnup Krankheit, Aminoazidurie), von Ionen (Natrium und Kalium bei Durchfall), von Metallen (Zink bei acrodermatitis enteropathica) und von Cholesterinlipiden (Herzgefäßkrankheiten). Insbesondere wurde vor Kurzem die wesentliche Rolle der Bürstensaum-Membran bei der Regulierung des empfindlichen Gleichgewichts zwischen Cholesterinaufnahme und -abgabe zwischen Lumen und Enterozyten hervorgehoben, einerseits durch die genetische Analyse von Einzelpersonen, die an Cholesterinstoffwechsel-Störungen leiden, andererseits durch einige klinische Studien, die die Effekte von diätetischen Pflanzensterolen (phytosterols) oder spezifischen Proteininhibitoren auf wesentliche Komponenten der Cholesterinaufnahme-Systeme untersucht haben.

Die Untersuchung der Bürstensaummembran ist in den meisten Studien mittels genetischer Analysen, Aktivitätstests und immunologischer Methoden durchgeführt worden. Deren Charakterisierung auf der Proteinebene wurde bisher durch den Umstand erschwert, dass eine Membran-Präparation von Dünndarm im geeigneten Reinheitsgrad sehr anspruchsvoll war, und weil die meisten Zielproteine integrale Membranproteine sind. In den wenigen existierenden Proteomik-Studien der Bürstensaummembran wurden bezeichnenderweise die meisten Transportproteine und Rezeptoren nicht identifiziert, die in dieser Fraktion erwartet würden. Das aggressive Umfeld und die proteolytischen Enzyme, denen die Enterozyten in

dieser Membran ausgesetzt sind, erfordern eine eigene Herangehensweise, um die strukturelle Integrität der Bürstensaum-Membranproteine im Isolationsprotokoll zu erhalten.

Das Hauptziel dieser Studie war die Entwicklung und die Evaluierung von massenspektrometrischen Vorgehensweisen für die Analyse und Charakterisierung von hydrophoben Proteinen, wie sie zum Beispiel in den Enterozyten der Bürstensaummembran vorkommen. Die Analyse von Membranproteinen in einem komplexen Gemisch ist seit jeher eine Herausforderung in der Proteomik. Die Hydrophobizität von Membranproteinen erfordert den Einsatz von Detergenzien für ihre Solubilisierung, was den Gebrauch von traditionellen Proteomik-Techniken (z.B. der zweidimensionalen Gelelektrophorese) erschwert. In komplexen biologischen Proteingemischen ist ihre Konzentration im Vergleich zu löslichen Proteinen niedrig, weshalb zusätzliche Reinigungsschritte für ihre Anreicherung nötig sind. Selbst bei ähnlichen Konzentrationen ist die Identifikationsrate für Membranproteine niedriger als diejenige für lösliche Proteine. Die Entwicklung eines optimierten Isolationsprotokolls in Kombination mit einer robusten analytischen Methode sollte demnach die Untersuchung von Bürstensaum-Membranproteinen ermöglichen. Dadurch könnten die Ergebnisse anderer Studien bestätigt werden, die zeigen, dass die Bürstensaum-Membranproteine enthält, die an der Cholesterinregelung in den Enterozyten beteiligt sind. Ein drittes Ziel dieser Studie war herauszufinden, unter welchen Bedingungen biologische Proben ohne den Einsatz von Isotopen-markierten Standards massenspektrometrisch quantifiziert werden können. Da weder die technische Reproduzierbarkeit noch die Stabilität der Bürstensaummembran-Präparation bekannt waren, musste die Reproduzierbarkeit der technischen Schritte für den vollständigen Arbeitsablauf überwacht werden, von der Bürstensaummembran-Präparation bis zur massenspektrometrischen Analyse der Proben, um rechtzeitig Schwierigkeiten und Einschränkungen zu identifizieren, die eine reine LC-MS-Quantifizierungsstrategie beeinträchtigen könnten.

In dieser Untersuchung wird (meines Wissens nach) erstmals ein analytisches Vorgehen beschrieben, welches die direkte Charakterisierung von Bürstensaum-Membranproteinen ermöglicht, deren Beteiligung bei der Cholesterinsynthese bereits beschrieben worden ist. Die Hauptelemente dieses Vorgehens sind ein verbessertes Protokoll, um reproduzierbar Bürstensaummembran-Präparation aus Dünndarmgewebe zu isolieren und aufzureinigen, das Design eines Proteaseninhibitorencocktails, der spezifisch die proteolytische Aktivität der endogenen Proteasen der Bürstensaum-Membran minimiert, ein gezieltes Vorgehen, um Bürstensaum Transmembranproteine aus den üblichen zytosolischen Kontaminanten aufzureinigen, und schliesslich die Verwendung eines robusten und empfindlichen

Instrumentariums auf der Basis von Kapillar-Flüssigchromatographie und Tandem-Massenspektrometrie. Insbesondere wurde das Protokoll von Kessler et al für die Aufreinigung und Anreicherung der Proteine in der Bürstensaum-Membran deutlich verbessert und ausgedehnt auf die gezielte Entfernung von zytosolischen und basolateralen Verunreinigungen. Ganze Dünndärme, oder zur Erhöhung der Reinheit davon abgeschabte Dünndarmschleimhaut, wurde zuerst lysiert und eine Rohmembranfraktion wurde durch differentielle Zentrifugierung isoliert. Das Pellet wurde resuspendiert und die basolaterale Membranfraktion gezielt durch $CaCl_2$-Fällung entfernt. Die angereicherten Bürstensaummembran-Vesikel wurden durch Zentrifugation sedimentiert und bei hoher Salzkonzentration und hohem pH-Wert gewaschen, um verbleibende zytosolische und membranassoziierte Proteine zu entfernen. Die Effizienz des Protokolls und die Reinheit der erhaltenen Bürstensaummembran-Vesikeln wurden durch Western-Blot-Analyse des bekannten basolateralen Markerproteins Na^+/K^+ ATPase a1 und des Bürstensaummembran-Markers FATP-4 überprüft. Das neue Protokoll ergab einen 30- bis 50-fachen Anreicherungsfaktor (im Vergleich zu einem 10-fachen Anreicherungsfaktor nach Anwendung des Kessler-Protokolls) bezogen auf das Rohlysat, wobei der basolaterale Marker, Na^+/K^+ ATPase a1, in der gereinigten Bürstensaummembran-Fraktion nicht mehr nachgewiesen werden konnte. Besondere Beachtung wurde bei der Protokoll-Optimierung der Inhibition der endogenen Bürstensaum-Membran Proteasen geschenkt, von denen die meisten der His-Zn-abhängigen Metalloproteasen-Superfamilie angehören, welche nicht auf gewöhnliche Serin- und Cystein-Protease-Inhibitoren reagieren. Durch die Zugabe von Amastatin, eines bekannten Teilinhibitors, und verschiedener Peptidsubstrate in alle Puffer des Bürstensaummembran-Aufreininigungsprotokolls wurde eine partielle Inhibition von etwa 90% erzielt. Ausserdem wurde das überschüssige Kalzium (ein bekannter Aktivator dieser Protease-Familie) durch Zugabe von EDTA im ersten Waschschritt direkt im Anschluss an die Kalzium-Fällung entfernt.

Eine Analyse mit Triplikaten der aufgereinigten Maus Bürstensaummembran-Fraktion ergab 1460 zuverlässig identifizierte Proteine, von welchen 260 Proteine als integrale Transmembranproteine vorhergesagt wurden. Eine detaillierte Aufschlüsselung gemäss Genontologie ergab, dass sich die Proteine gleichmässig über die Plasmamembran, ER/Golgi/Endosom Kompartimente, das Zellskelett und die Mitochondrien verteilen, während nur ein Viertel der Bürstensaummembran-Proteine als zytoplasmatisch annotiert werden. Mittels dieses Protokolls wurden mehrere Proteine, welche bei der Cholesterinaufnahme eine wichtige Rolle spielen, erstmals auf Proteinebene direkt

identifiziert. Ein Beispiel dafür ist das Niemann-Pick C1-like 1 (NPC1L1) Protein, ein Protein mit 13 Transmembran-Domänen, welches in der Bürstensaummembran-Präparation stark vertreten war und das in der Literatur als zentral bei der Cholesterinaufnahme beschrieben wird. Gleiches gilt für die beiden ABC Transporter ABCG5 und ABCG8, die an der biliären Sekretion von Cholesterin und planzlichen Sterolen beteiligt sind. Dem aus Caveolin-1 and Annexin-2 bestehenden Komplex wurde ebenfalls eine zentrale Rolle für den Cholesterintransport von der Bürstensaummembran zum endoplasmatischen Retikulum zugewiesen, während SR-BI, ein 4-Transmembrandomänen-Rezeptor, an der Cholesterinaufnahme beteiligt sein soll. Die Rolle vieler der oben erwähnten und weiterer Proteine wie zum Beispiel CD36, Galectin-4, und ABCB1 ist Gegenstand intensiver Diskussionen. Verschiedene Studien belegen ihre Beteiligung an der Cholesterinaufnahme, aber ihr genaues Wirkprinzip ist noch unklar. Bemerkenswert ist ausserdem der eindeutige Nachweis von drei Lipoproteinen, ApoA-I, ApoA-IV and ApoE, unter den identifizierten Proteinen der Bürstensaummembran-Präparation. Ihr Nachweis war kaum zu erwarten, da Lipoproteine kleine, lösliche, von der Leber sekretierte Proteine sind und deshalb nicht als Bestandteile der Bürstensaummembran gelten. Ihr Vorkommen dürfte einer engen Interaktion mit bestimmten Bürstensaummembran-Komponenten wie z.B. LRP-1 und Cubilin zuzuschreiben sein, von welchen bekannt ist, dass sie mit Apolipoproteinen interagieren.

Angesichts der Komplexität der genutzten Proteomik-Vorgehensweise stellte sich die Frage, ob es möglich ist, reproduzierbar und quantitativ Hunderte von Membranproteinen gleichzeitig zu untersuchen. Einige Schritte wurden deshalb besonders gründlich analysiert, einerseits anhand der Protein-Identifikationen (Vergleich der MS/MS-Resultate zwischen Banden) und andererseits anhand des Peptidsignals, um systematisch die massgeblichen Faktoren für die Reproduzierbarkeit der Gesamtanalyse festzustellen. In beiden Fällen konnte gezeigt werden, dass die meisten technischen Schritte, wie z.B. das Verfahren für LC-MS/MS Identifikationen, die Auftrennung komplexer Proteingemische anhand von 1D-SDS-PAGE oder der In-Gel-Verdau von Proteinen mittels Trypsin, nur wenig zur Gesamtvariabilität beitrugen, sofern sie angemessen kontrolliert wurden. Die hohe Reproduzierbarkeit des LC-MS/MS-Prozesses konnte dabei bestätigt werden. Hingegen konnte das Bürstensaummembran-Protokoll als mögliche Quelle der Variabilität identifiziert werden, da es relativ lang und komplex ist und Schritte enthält, deren Reproduzierbarkeit schwer zu gewährleisten ist, z.B. die $CaCl_2$-Fällung. Ausserdem könnten Überreste proteolytischer Aktivität aus den stark vertretenen Proteasen Probenabbau bewirken und dadurch die Reproduzierbarkeit des Protein-Identifikationsprozesses beträchtlich verringern. Die

Variabilität der Anzahl der Proteine, die in allen Proben gemeinsam identifiziert wurden, war aber bei der Verwendung einer optimierten Versuchsplanung deutlich geringer.

Die vorliegende Arbeit zeigt, dass Membranproteine in einem komplexen Gemisch wie z.B. einem Isolat aus der Enterozyten-Bürstensaummembran reproduzierbar und quantitativ analysiert werden können. Entscheidend für den Erfolg der Analyse sind ein robustes Probenaufbereitung-Protokoll, um stark angereicherte Bürstensaummembran- Vesikeln zu erhalten, eine engmaschige Kontrolle der Analyseschritte und ein statistisch abgesichertes Vorgehen bei der Auswertung. Besonders wichtig zu bemerken ist, dass in diesen stark mit Membranproteinen angereicherten Bürstensaummembran-Vesikeln die Anzahl unterschiedlicher Peptide und die durchschnittliche Summe der Anzahl Peptide die relative Abundanz jedes Proteins widerspiegelt, unabhängig von der Anzahl seiner Transmembran-Segmente. Allerdings stammen die identifizierten tryptischen Peptide ausschliesslich von Schleifen oder zytoplasmatischen Bereichen des Membranproteins. Wenn man annimmt, dass Trypsin Transmembranhelices spaltet, d.h. dass Transmembranpeptide während des Verdaus entstehen, so ist es gut möglich, dass die hydrophoben Eigenschaften und die Länge der meisten dieser Peptide (durchschnittlich 30-40 Aminosäuren) die Extraktion aus der Gel-Bande erschweren, oder sie könnten während der Peptid-Extraktion an der Gefässwand haften bleiben. Ausserdem ist nicht klar, ob sehr lange oder hydrophobe Peptide unter den üblichen RP-LC-MS/MS-Bedingungen analysierbar sind. Unter diesem Gesichtspunkt könnte der Einsatz einer anderen stationären Phase wie z.B. HILIC zusätzliche, komplementäre Peptidinformation liefern.

Diese Arbeit schafft auch eine Grundlage für zukünftige Proteomik-Experimente mit Schwerpunkt auf der Bürstensaummembran-Biologie, welche bereits existierende, auf Genexpressions-Daten basierte Studien zur Inhibition von Cholesterinabsorption vertiefen und ergänzen. In einem Vorexperiment wurden Bürstensaummembran-Präparationen einer ApoE Knockout-Maus und einer Wildtyp-Maus mit demselben genetischen Hintergrund untersucht. Die ApoE Knockout-Maus ist eines der am meistverbreiteten Mausmodelle um Dislipidämie zu untersuchen, wobei eine gezielte Deletion des apoE-Gens eine schwere Hypercholesterämie und spontane Atherosklerose auslöst. Da biologische Replikate fehlen wurden nur "schwarz-weisse" Unterschiede betrachtet. Interessanterweise konnten das Ileal-Bile-Acid-Transporter(IBAT)-Protein und das ApoAI-Protein in der ApoE-Knockout-Maus nicht nachgewiesen werden, während es bei Wildtyp-Mäusen durchwegs nachweisbar war. Dieser Befund ist in keiner der früher publizierten Untersuchungen beschrieben und deutet stark auf eine schwere Störung des Gallensäuren-Metabolismus im Knockout-Tier hin.

Zusammenfassend kann gesagt werden, dass das analytische Vorgehen in der vorliegenden Arbeit eine umfassende vergleichende Analyse von Mäusen erlaubt, die z.B. mit bestimmten Wirkstoffen oder Futterzusammensetzungen behandelt wurden. Dieser Untersuchung könnte eine breit angelegte Proteomik-Studie folgen, in welcher ein viel umfassenderes biologisches Experiment untersucht würde, z.B. ein Vergleich zwischen Mäusen mit Statin-Behandlung, Ezetimibe-Behandlung und einer Kontrollgruppe von unbehandelten Mäusen. Der bislang unbekannte Effekt dieser Wirkstoffe auf die Bürstensaum-Membran könnte damit detaillierter untersucht werden.

1. INTRODUCTION

1.1 Brush Border Membrane BBM

1.1.1 Physiology of the small intestine

The small intestine is the longest section of the digestive tube and consists of three segments forming a passage from the pylorus to the large intestine. The duodenum is a short section starting immediately after the stomach and which receives secretions from the pancreas and liver via the pancreatic and common bile ducts. It is followed by the jejunum, considered to be roughly 40% of the small intestine in man, but closer to 90% in animals, and by the ileum, which connects to the large intestine. The ileum is considered to be about 60% of the intestine in man. However, veterinary anatomists usually refer to it as being only the short terminal section of the small intestine.

The structure of the small intestine looks on the first sight quite similar to other regions of the digestive tube. However, three features account for its huge absorptive surface area. The inner surface of the small intestine is not flat but wrinkled into circular folds (mucosal folds), which increase its surface area several-folds. The mucosa itself is composed of multitudes of projections (villi) which protrude into the lumen and are covered with epithelial cells. Finally, the lumenal plasma membrane of those absorptive epithelial cells is also folded and densely-packed in microdomains named "microvilli", whose border is commonly referred to as the **"brush border"** due to its appearance in the microscope (see Fig. 1.1).
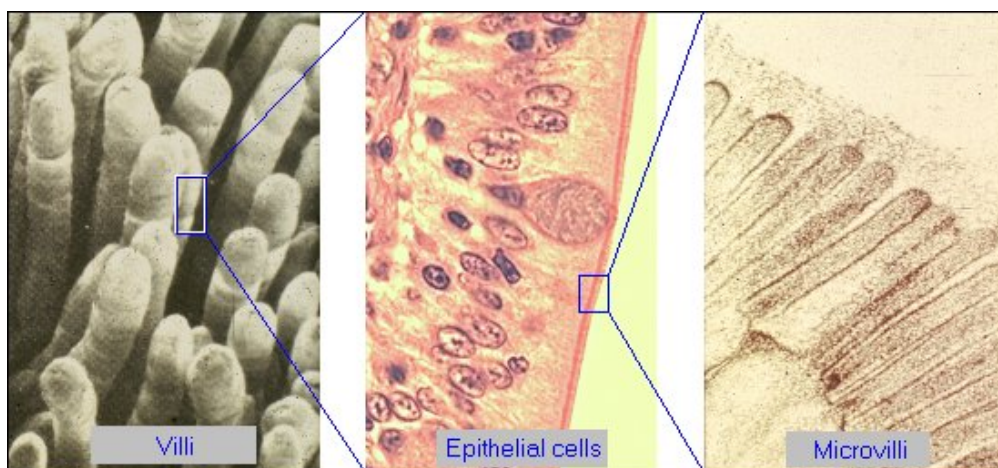
Figure 1.1: *The small intestine inner surface. The panels above depict the bulk of the small intestine surface area expansion, showing villi, the epithelial cells that cover the villi, and the microvilli of the epithelial cells (downloaded from web site http://www.vivo.colostate.edu:80/hbooks/pathphys/digestion/smallgut/anatomy.html).*

The epithelial cells of the small intestine mature into absorptive epithelial cells that cover the villi. These are the cells that take up and deliver to the blood stream virtually all nutrients from the diet. Two other major cell types populate the small intestinal epithelium: the enteroendocrine cells which, as part of the enteric endocrine system, sense the lumenal environment and secrete hormones such as cholecystokinin and gastrin into blood; and the Goblet cells, which secrete lubricating mucus into the intestinal lumen.

1.1.2 Brush Border Membrane: Location and Function

Intestinal epithelial cells are polar in their cellular organization. The intestinal brush border (synonyms: microvillus, luminal, apical) membranes of the enterocytes differ in protein and lipid composition from the inner side of the plasma membrane, the basolateral membrane (BLM) (see Fig. 1.2). The apical surface of polarized intestinal epithelial cells (the surface facing the intestinal lumen) is characterized by structurally distinct cell protrusions referred as microvilli or brush border membranes (BBMs), responsible for digestion and absorption of nutrients.
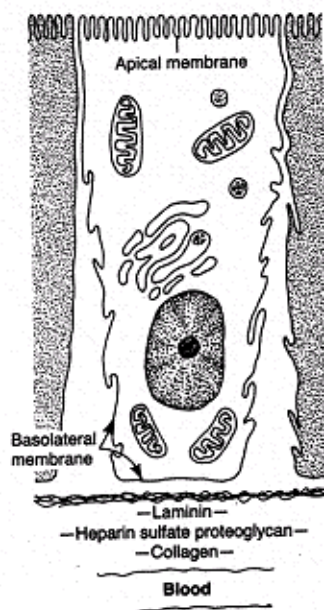
Figure 1.2: *Schematic representation of a typical intestinal epithelial cell. The apical membrane (BBM) has a different protein and lipid composition from the basolateral membrane. The BBM can be isolated from the BLM using protocols that take advantage of the difference in polarity between the two membranes (figure downloaded from web site: http://www.vivo.colostate.edu:80/hbooks/pathphys/digestion/smallgut/anatomy.html).*

The processing capacity of enterocytes is directly proportional to the surface of absorptive epithelia BBM. BBM are supported by cytoskeletal actin filaments which are organized into

both more or less permanent and rapidly rearranging bundles. Cytoskeleton bundles are in turn interconnected with transmembrane protein complexes forming a highly organized import–export membrane interface specialized for a variety of digestive and absorptive functions, such as protein and peptide degradation, absorption of minerals, amino acids, sugars, lipids and cholesterol (1). Shortcomings in these mechanisms may cause a variety of pathological conditions such as disorders in the metabolism of saccharides (glucose galactose malabsorption, lactose intolerance) amino acids (Hartnup disease, aminoacidurias), ions (sodium and potassium in the case of familiar diarrhea), metals (zinc in acrodermatitis enteropathica) and cholesterol lipids (cardiovascular diseases).

Recently, several proteomics studies have reported the identification of proteins localized in the BBM membrane (2, 3). Until now, however, these approaches have failed to identify transporters and receptors that are known to be located in the BBM membrane based on kinetic studies, immunological assays and in gene data, probably because of the complexity of the analyzed samples.

1.1.3 Brush Border Membrane and Lipid Rafts microdomains

Recent studies have suggested that plasma membranes might be organized into heterogeneous functional microdomains. One type of these microdomains, called lipid rafts, is stated to be enriched in glycosphingolipids/cholesterol and in typical sets of proteins, among them also cholesterol transporters (4). Lipid rafts can be isolated by taking advantage of their resistance to nonionic detergent extraction at cold and by their differential buoyancy on a density gradient ultracentrifugation. The lipid rafts hypothesis was originally proposed to explain how proteins and lipids were sorted to the apical surface of polarized cells. However, in recent years, several functions including signaling, cholesterol homeostasis, cell trafficking or even docking sites on mammalian cells for certain pathogens and toxins have also been attributed to lipid rafts (5, 6). Despite accumulated experimental data from biophysical, biochemical, and fluorescent microscopy studies supporting the fact that lipid rafts may exist in vivo, the lipid rafts hypothesis remains controversial at least for their size, stability and the mechanism of their formation (7).

Lipid rafts isolated from the BBM have also been the subject of several recent proteomics studies (3, 8, 9). While these studies have reported the identification of proteins that were localized in the lipid rafts, almost none of these proteins were described to be involved in cholesterol absorption, a major area of interest for the analysis of the lipid rafts.

1.2 Cholesterol homeostasis

The view of cholesterol as a nasty substance clogging arteries and causing heart disease is probably the one aspect that is better known to the general public. However, besides its unflattering reputation, cholesterol fulfills many other roles and is a vital component of cell membranes without which the cell would not function. It is also the precursor to all steroid hormones, bile acids and oxysterols, which by themselves are important regulatory molecules in many metabolic pathways.
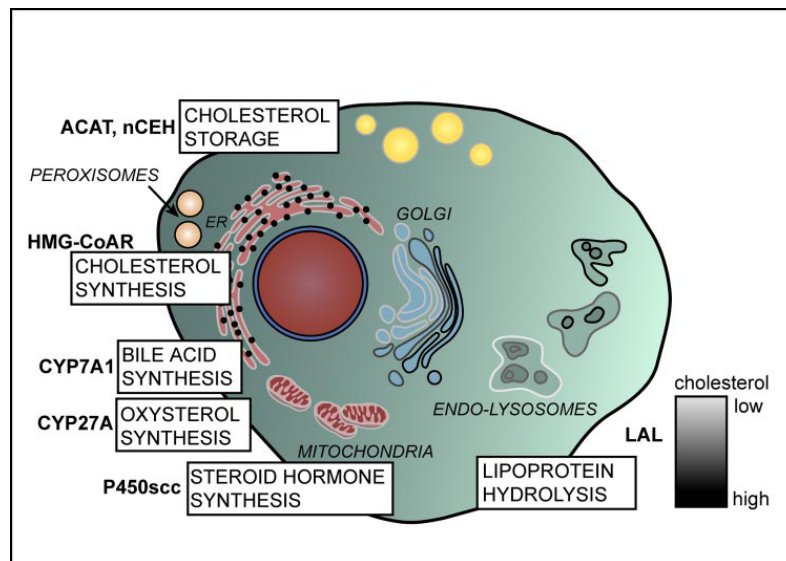
Figure 1.3: *Cellular cholesterol distribution and key enzymes of cellular cholesterol metabolism. The approximate cholesterol content of the membrane is indicated by shades of gray. The main processes of cholesterol metabolism, key enzymes involved, and their subcellular locations are indicated. Key enzyme in the cholesterol synthesis is hydroxymethylglutaryl COA reductase (HMG-CoAR). The 3'-OH group of cholesterol is esterified by the enzyme acyl-CoA: cholesterol acyltransferase (ACAT). The enzyme responsible for cholesterol esters degradation in lipid droplets is neutral cholesterol sterol hydrolase (nCEH). Cholesterol can be converted to bile salts via two pathways: the classic pathway, involving the key regulatory enzyme CYP7A1 hydrolase, and the alternative pathway, probably related to the oxysterol synthesis, involving the key enzyme sterol 27-hydroxylase (CYP27A), located in the mitochondria. Cholesterol is an obligatory precursor for steroid hormone production (figure adapted from Ikonen Elina, (10))*

Endogenous cholesterol is synthesized mainly in the liver in a regulated pathway. The 27-carbon tetracyclic cholesterol molecule is synthesized from acetate in a series of ~30 enzymatic reactions. The ER is the primary production site of cholesterol and the key rate-limiting enzyme of this pathway is the ER-located hydroxymethylglutaryl COA reductase (HMG-CoAR) (figure 1.3). The sub-compartmentalization of the cholesterol biosynthetic pathway remains poorly understood as of today.

In general body cholesterol is primarily of endogenous origin and its homeostasis involves the movement of cholesterol between peripheral tissues and the liver (11). The liver regulates the *de novo* synthesis of cholesterol and the excretion of cholesterol into bile (directly or after conversion to bile acids), the secretion of cholesterol into blood as very low-density lipoproteins (VLDL), the modulation of receptor-mediated cholesterol uptake, the formation of esterified cholesterol (CE) and the storage of cholesterol. The intestine regulates cholesterol absorption and excretion into feces (12).

1.2.1 Cholesterol absorption in the Small Intestine

Intestinal cholesterol absorption is a complex process that involves multiple interrelated sequential degradative and synthetic pathways, many of them not yet clearly defined. This biological process has attracted the interest of many pharmaceutical companies because it might provide multiple therapeutic targets in the management of patients with hypercholesterolemia.

Dietary cholesterol is absorbed from bile salt micelles with fatty acids and lysophospholipids in the proximal part of the small intestine. Key proteins involved in dietary cholesterol uptake by the enterocytes have been identified during the past few years and, in particular, the important role that the NPC1L1 protein plays in this process (13). The NPC1L1 protein is localized in the brush-border membrane of enterocytes and has been shown to be required for intestinal uptake of both cholesterol and plant sterols (14). Recent evidence suggests that this protein is the target of the cholesterol-lowering drug ezetimibe (15). Whether NPC1L1 functions as a genuine cholesterol transporter, promoting cholesterol transfer through the plasma membrane, or is indirectly involved in the process is not yet known. In addition, the ABC transporter family half-transporters ABCG5 and ABCG8 (sterolin-1 and sterolin-2) constitute a functional heterodimeric unit limiting sterol absorption (16). The role of ABCG5 and ABCG8 in dietary cholesterol absorption may not be direct, that is, by inhibiting dietary cholesterol uptake by enterocytes; rather, this transporter appears to stimulates hepatic sterol excretion into the bile and thereby modulates the bile-acid/sterol ratio, possibly to promote the secretion of absorbed sterols from the intestinal epithelium back into the gut lumen (17). Fig 1.4 summarizes in more detail the cholesterol and sterols absorption in the enterocytes and possible proteins and pathways that are associated to this process.
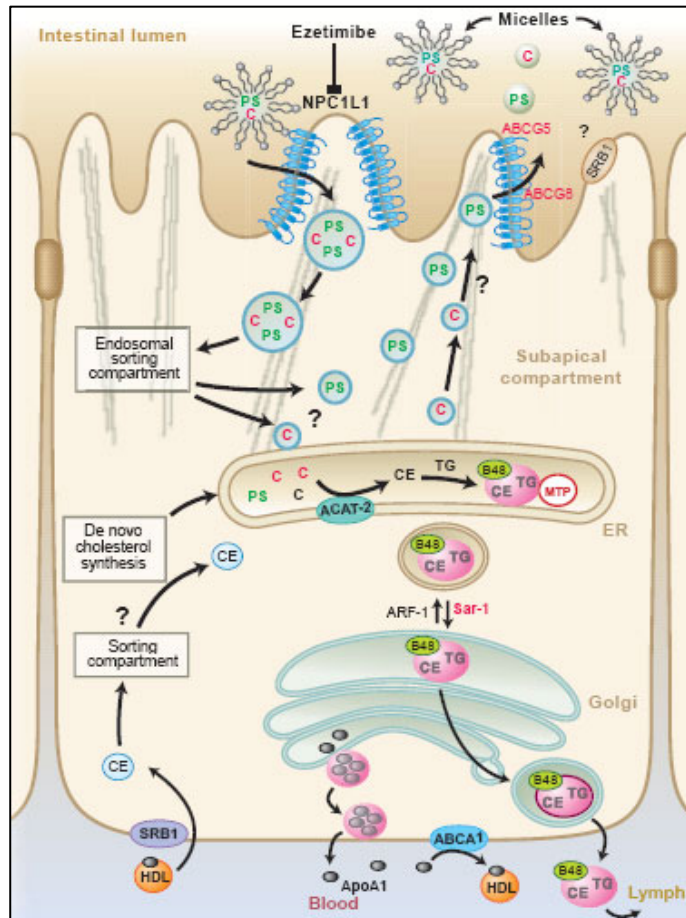
Figure 1.4: *Absorption of dietary cholesterol and noncholesterol sterols in enterocyte. NPC1L1, expressed at the apical surface of enterocytes, may be the transporter that selectively absorbs dietary cholesterol (C) from micelles in the lumen of the small intestine, a step that is blocked by the drug ezetimibe. In this model, the NPC1L1 transporter permits the uptake of cholesterol (and noncholesterol sterols) into vesicles that then move through a sub-apical endosomal sorting compartment. Mutations in either of the transporters ABCG5 or ABCG8 cause the hyperabsorption of dietary plant sterols (PS) and other noncholesterol sterols from the small intestine, resulting in the human disease sitosterolemia. The endosomal sorting compartment allows cholesterol to progress to the endoplasmic reticulum (ER), where it is esterified (CE) by ACAT-2 and then transferred to chylomicrons (pink) ready for secretion into the bloodstream; plant sterols are shunted through a pathway resulting in their transport back to the gut lumen via ABCG5 and ABCG8. Cholesterol that is synthesized de novo is also esterified by ACAT-2 and enters chylomicrons (figure adapted from Klett E.L. et al,(18))*

1.3 Membrane proteomics

1.3.1 Proteomics – Definition and workflow

The genomic sequencing of numerous organisms has radically transformed biological and medical research, providing the foundation for the large-scale interpretation of gene and cellular function. In this context, the term "proteome", coined in 1994 by Marc Wilkins (19), describes the entire **prote**in complement of the gen**ome**. Proteomics, the studies of the proteome, encompass the identification, characterization and quantification of the complete set of proteins expressed in the lifetime of a given cell, tissue or organism, including isoforms, polymorphisms and modifications, protein-protein interactions and the structural description of proteins and their complexes. Most biological functions are carried out by proteins, and to understand how cells work, one must study which proteins are present, what they do, and how they interact with one another. If the genome represents the words in a dictionary, then the proteome provides the definitions, while the interactions of the proteins with one another and with the other molecules in their environment provide the grammar to form a meaningful language.

Proteomics would not be possible without the previous achievements of genomics, which provide the information about the large, but finite number of gene products that are the focal point of proteomics studies. The challenges of proteomics are larger and far more complex than the huge but basically straightforward task of mapping the genome. In contrast to the static nature of the genome, which is essentially identical in every cell of an organism, the proteome is dynamic, constantly changing and responding to internal and external stimuli. Proteomics must deal with unavoidable problems of limited and variable sample material, sample degradation, vast dynamic range (more than 10^{12} orders of magnitude for protein abundance in plasma), a multitude of post-translational modifications, almost endless tissue, developmental and temporal specificity, and disease and drug perturbations.

Proteomics represents nowadays a large family of partially overlapping areas of interest evolving along with technology breakthroughs, bioinformatics advances, and certainly also with the personal interests of investigators. Some of those areas of application are mass spectrometry-based proteomics, proteome-wide biochemical assays, systematic structural biology and imaging techniques, proteome informatics, and clinical applications of proteomics. The divisions between these areas are somewhat arbitrary, not least because technological breakthroughs often find immediate application on several fronts. More important, biologically useful insights into protein function often emerge from the

combination of different proteomic approaches. This study is mostly focused in mass spectrometry-based proteomics and this area will be discussed in more details in the following paragraphs.

All proteomics experiments aim, in an ideal setting, to monitor quantitatively a full proteome at any time point of an experiment. However, the sheer complexity and dynamic range of an unfractionated proteome makes it technically impossible to address all its constituents simultaneously by any direct analytical means. As a result, it is often one of the most critical steps of a proteomics experiments to knowingly restrict the scope of the experiment to a biochemical-relevant sub-proteome that can be effectively monitored by the chosen analytical approach. A criterion often used in this process is to take advantage of some prior knowledge to "bias" the experiment towards a protein population of interest, for example by limiting the proteomics analysis towards a specific cell compartment (organelle-based proteomics), by isolating proteins with specific physical-chemistry characteristics (cytosolic proteins, membrane proteins, cytoskeleton, etc.), or even to only consider proteins captured through specific interaction ("affinity" proteomics, chemical proteomics, etc.). The achieved analytical level of precision increases generally in pair with the degree of fractionation obtained, albeit sometimes at the danger of excluding an important (and usually unknown) aspect of the experiment to follow. Conversely, a comprehensive survey of a broadly-chosen proteomics experiment might require the analysis of so many different fractions (to ensure the monitoring of a significant portion of the proteome of interest) that such an experiment might not be practicable anymore from a technical point-of-view.

Almost all mass spectrometry-based proteomic approaches are performed at the peptide level as the MS analysis of whole proteins (the so-called top-down approach) is less sensitive and the deconvolution of the generated multiply charged species is very difficult to handle from a bioinformatics point-of-view. A protein mixture of interest, isolated from a cell lysate, tissues, or enriched by a biochemical fractionation or affinity selection, is very often analyzed as a final fractionation step by 1D or 2D gel electrophoresis. Proteins are then in-gel digested (in the rare case an electrophoretic step is omitted, the proteins are then directly digested) and the extracted peptides can then be fractionated again by liquid chromatography before being analyzed by mass spectrometry. The types of chromatography and mass spectrometer that are used for an experiment depend mostly on the complexity and the type of questions that need to be answered for given analytical strategy.

The ability of mass spectrometers to quantitatively analyze ever smaller amounts of proteins from increasing complex mixture at a very high level of precision has been a primary driving force in this proteomics approach. In the last 2 years, new bioinformatics tools and the extended use of statistics in data processing and data analysis have enabled the design and execution of proteomics experiments that would not have been thinkable even 5 years ago. In particular, the high mass accuracy (below 2 ppm via Lock Mass injection, (20)) and the extended dynamic range shown by the newest generation of mass spectrometers now allow the statistical evaluation of the quality and the significance of a protein identification (via its constitutive peptides), while simultaneously being able to measure ion current of peptide ions at an unprecedented level of precision.

1.3.2 Membrane Proteins - Importance and Characteristics

Membranes play a critical role in cellular structure by providing a physical barrier between the cell and its environment and the various subcellular compartments within eukaryotic cells. Although the basic structure and function of biological membranes is provided by the lipid bilayer, membrane-spanning proteins confer unique compartment-specific functions and communication between separated environments.

The plasma membrane provides a physical boundary between the cell and its environment, playing important roles in many fundamental biological processes such as cell-cell interactions, signal transduction, and material transport. The plasma membrane components have been extensively targeted for drug design; in particular, plasma membrane proteins may account for up to 70% of all known drug targets (e.g., HER2- and G protein-coupled receptors). For example, identification of overexpressed plasma membrane proteins in diseased cells could provide protein targets for the design of either therapeutic monoclonal antibodies or small-molecule drugs.

Membrane proteins are by definition proteins that are associated with the membrane. However, in this context, the concept "associated" relates to several different situations. In a first case, the polypeptide chain spans the lipid bilayer a certain number of times and proteins belonging to this category are defined as "integral" or "intrinsic" membrane proteins. In a second case, the membrane-associated protein might be physically coupled to the membrane, whereas the association is mediated by a post-translation modification of the polypeptide, for example the grafting of a fatty acid, a polyisoprenyl chain, or through glycolipid anchors, such as the glycosylphosphatidyl inositol modification common in eukaryotic species. Such proteins are defined as "membrane-anchored" proteins. Finally there are proteins that are

associated to a membrane due to their interaction with other membrane proteins or some specific lipids, but which contain neither transmembrane domains nor lipid modifications. Such proteins are typically referred to as "membrane-associated".

The transmembrane domains of integral proteins can typically fold to form either an α-helix or a β-sheet secondary structure. α-helices are formed by the consecutive joining of mostly non-polar amino acids, with typically 15-25 amino acids required to span the membrane bilayer. These amino acids exhibit positive hydropathy values and hence are the major contributors to the hydrophobic character of the membrane proteins. β-sheet transmembrane domains are formed by alternating polar and non-polar amino acids in the amino acid sequence. Polar amino acid side chains face the aqueous channel while the side chain of the non-polar amino acids interface with the lipid bilayer. This type of proteins is considerably less hydrophobic than α-helical integral proteins. Proteins with β-sheet transmembrane segments tend to form a β-barrel structure that allows the passage of defined molecules. Structural prediction software for membrane proteins predicts nowadays quite successfully proteins with transmembrane α-helix structure but typically don't consider the β-sheet type.

1.3.3 Proteomic approaches for Membrane Proteins

Although the analysis of soluble proteins by mass spectrometry-based proteomics technologies has made rapid progress in recent years, the analysis of membrane proteins has lagged behind and their identification is typically underrepresented in datasets. Thus, the portion of membrane proteins reported in existing analysis is much lower than the 20–30 % predicted by the human genome (21).

Traditionally the analysis of complex protein mixtures has been carried out using gel-based methods (22). Unfortunately, the well-known, highly-resolving two-dimensional IEF/SDS gel electrophoresis method has never been successfully applied for the separation of membrane proteins (23). Many hydrophobic proteins do not solubilize well in the non-ionic detergent required for the isoelectric focusing step and the few that survived this process tend to precipitate at their isoelectric point. In many studies, this separating method has been replaced by the more robust SDS-based one-dimensional gel coupled with mass spectrometry (24). Alternatively, several authors have described alternate two-dimensional gel electrophoresis separation technique, such as the so-called two dimensional blue native/SDS electrophoresis method (25) or diagonal SDS-PAGE electrophoresis (26).

Membrane proteins (and their derived hydrophobic peptides) tend to show the same trends if a reverse-phase liquid chromatography separation step is used instead. Several approaches have

attempted to overcome this problem by using either strong organic acid-cyanogen bromide (27), detergents (28), acid labile surfactants (29), organic solvents (30), salts, or high pH conditions (31) to solubilize membrane proteins. Although these methods proved to be efficient one way or the other, they were usually bound with other disadvantages that limited their usefulness. The presence of detergents affects the performance of chromatographic separation and also leads to mass spectral signal suppression. Commonly-used proteases, like trypsin, cannot be used for digestion of proteins when organic acids are used due to low pH conditions, or the enzyme activity is significantly reduced in presence of high percentage of organic solvents. High pH-based methods use proteinase K to cleave proteins non-specifically at random amino acid sequences, which makes the resulting peptide mixture extremely complex to analyze. In general, all methods described above require extensive sample handling to make the sample compatible for mass spectral analysis.

In summary, it is clear that there is a strong need for a simple and efficient method to analyze membrane proteins. The first step in this direction is the development of a suitable protocol for the enrichment of membrane proteins. As membrane proteins are typically lower in abundance then compared to soluble proteins, a biochemical fractionation method taking advantage of their unique physical-chemical properties is required to overcome their otherwise rather limited dynamic range in the analysis.

1.4 Mass spectrometry

Mass spectrometry has been the analytical workhorse tool for biologists over the last quarter century. As its name implies, the mass spectrometer is an instrument to measure the mass of a substance. Initially used for the analysis of small, volatile molecules, its use has quickly become ubiquitous in the "biological" world with the invention of ionization sources compatibles with the analysis of the "big" molecules in the 1980s.

Mass spectrometers have been used over the last hundred years for a wide variety of applications, ranging from estimating the masses of elements and their isotopes to small molecule identification and characterization to modern day proteomics analysis. According to the time period and the need, mass spectrometry and mass spectrometers have evolved. Various types of mass spectrometers exist today to address a wide variety of applications ranging from analyzing the soil of alien worlds in space research to structural characterization of complexes of proteins.

All mass spectrometers consist of three basic components (Fig. 1.5). An ion source ionizes first the molecules to analyze, then a mass analyzer separates the generated ions according to their mass-to-charge ratio (m/z) and a detector measures the ion beam current. Each of these elements exists in several different forms so that a wide variety of mass spectrometers may exist to fulfill different needs.

Ion source   **Mass analyser**   **Detector**
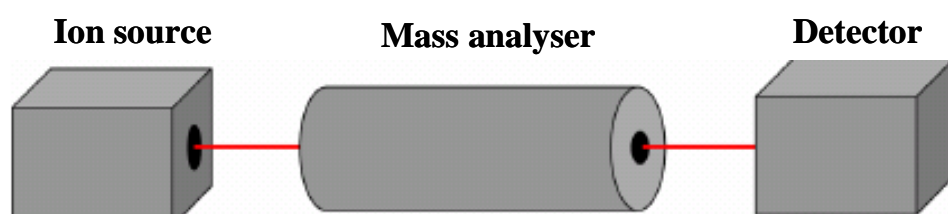
Figure 1.5: Schematic representation of the components of a mass spectrometer (Adapted from Lottspeich and Zorbas (32)).

Mass spectrometers are operated under vacuum (ranging from 10^{-4} for ion traps to 10^{-10} Torr for FT-based instruments) to prevent the loss of the ions by collision with a gas molecule.

1.4.1 A brief history of MS in biology

First attempts to analyze intact peptides by mass spectrometry were achieved using fast atom bombardment (FAB) ionization, which was first described in 1981 by Barber and co-workers (33). This ionization method was able to desorb (for the time) rather large molecules in the mass range of 2000–17000 Da depending on the sensitivity required. The first method that

was able to ionize high-molecular mass molecules such as proteins was achieved using plasma desorption (PD) ionization in 1982 (34), based on an earlier method described by Macfarlane and co-workers (35).

It is however only with the inventions of the electrospray ionization (ESI) and the matrix-assisted laser desorption/ionization (MALDI) technique that peptides and proteins became really amenable to mass spectrometry analysis. In the eighties Fenn and co-workers (Yamashita & Fenn, 1984) developed electrospray ionization as a technique to ionize intact large molecules in solution. One of the peculiarities of ESI is to generate ions of differing charge states for the same analyte leading to spectra with numerous peaks. The nature of ESI has the advantage of being directly compatible with liquid chromatography and capillary electrophoresis systems so that peptides or proteins mixtures analyzed by one of those two methods can be investigated online by ESI-MS. During the same period Karas and Hillenkamp (36) discovered matrix assisted laser desorption/ionization (MALDI). Tanaka was able to obtain protein molecular ions of masses up to typically 25000 Da with a matrix made of an ultra fine metal powder mixed with glycerol (37) when Karas and Hillenkamp were investigating the polypeptide mellitin, 2843 Da, and the oligosaccharide stachyose, 666 Da. It is however the type of matrix that Karas and Hillenkamp used, a UV-light absorbing organic compound, that is the basis for most of the now existing MALDI applications (38).

1.4.2 Ionization technique

Matrix-assisted laser desorption/ionization (MALDI) and Electrospray ionization (ESI) are nowadays the most commonly used ionization processes to analyze proteins and peptides by mass spectrometriy. Their principles and characteristics are described below.

1.4.2.1 Matrix-assisted laser desorption/ionization (MALDI)

The generation of a protonated molecule in the gas phase using the MALDI ionization process is achieved by mixing the analyte of interest with a large excess of a matrix material and to let them co-crystallize onto a planar surface, typically a metallic target. The resulting crystal is then irradiated by nanosecond laser pulses, typically using a nitrogen laser at a wavelength of 337 nm or, more recently, solid-state Nd: YAG laser at a wavelength of 355 nm. The matrix plays a particular role in the MALDI process (Fig. 1.6). First, the matrix absorbs the incoming energy of the laser pulse (therefore protecting
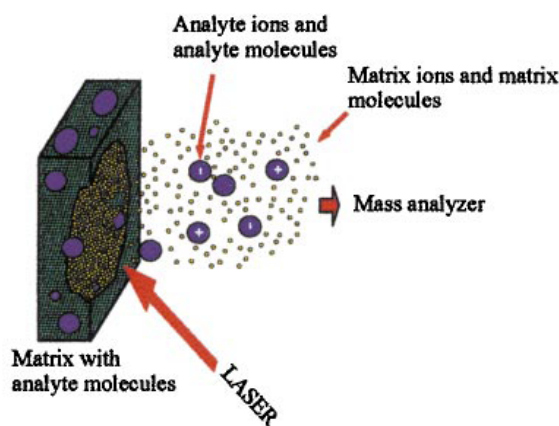
Figure 1.6: *Schematic representation of MALDI process and instrument. A sample cocrystallized with the matrix is irradiated by a laser beam, leading to sublimation and ionization of peptides (Adapted from Mann et al.,(39)).*

the analyte from a direct "hit"), which is then followed by a thermal explosion leading to desorption of the analyte and subsequent ionization by charge transfer from the matrix to the analyte in the gas phase. Matrices are typically small organic molecules with an absorbance maximum matching the wavelength of the laser employed, but differing in the amount of energy they impart to the biomolecules during desorption and ionization and hence the degree of fragmentation (unimolecular decay) that they may cause. Thus, peptides are typically analyzed using so-called hot matrices such as α-cyano-4-hydroxycinnamic acid or dihydrobenzoic acid while protein work would use "colder" matrices such as sinapinic acid. The MALDI ionization process is characterized by the formation of mostly singly-charged ions, therefore generating simple ion spectra, but putting a special requirement on mass analyzers to analyze the wide variety of ions (in respect to their m/z ratio) that can be generated.

The precise nature of the ionization process in MALDI still remains speculative in its nature and the obtainment of high signal intensities with good resolution is determined to some extent by trial and error, depending on incorporation of the analytes into crystals, their likelihood of capturing and/or retaining a proton during the desorption process and a number of other factors including suppression effects in peptide mixtures. Proteins generally undergo fragmentation to some extent during the MALDI process, resulting in broad peaks and loss in sensitivity; therefore MALDI is mostly applied to the analysis of peptides.

1.4.2.2 Electrospray ionization (ESI)

The electrospray ionization technique is performed using a liquid interface. A suitable solvent containing the analyte of interest is passed through a hypodermic needle set at high voltage to electrostatically disperse, or electrospray, small, micrometer-sized charged droplets, which rapidly evaporate and which impart their charge onto the analyte molecules (Fig. 1.7). The ionization process takes place at atmospheric pressure and is therefore very gentle (without fragmentation of analyte ions in the gas phase). The generated ionized molecules are transferred into the mass spectrometer with high efficiency for analysis. Depending of the flow rate (mid μl to ml/min), the electrospray process need to be assisted using nebulizer gas.
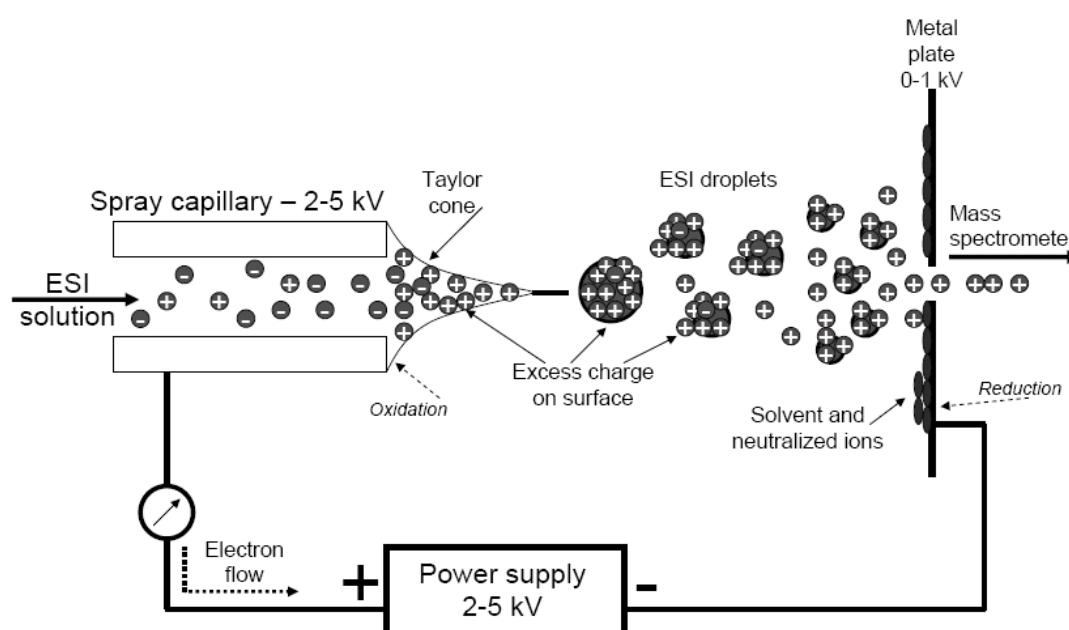
Figure 1.7: *Overview of the mechanics of the electrospray ionization technique. The high voltage applied to the spray sample causes positive charge to build up at the spray tip. Due to the charge and the pressure a so-called Taylor cone is formed. From the Taylor the charged droplets well decrease in size and split, until eventually single proteins or peptides with multiple charged are desorbed and transferred to the mass spectrometer inlet (Adapted from Lottspeich and Zorbas(32)).*

A wide range of compounds can be analyzed by ESI-MS; the only requirement is that the molecule should be sufficiently polar to allow attachment (in the positive ionization mode) or removal (in the negative ionization mode) of a charge. This includes proteins, oligonucleotides, sugars (with less sensitivity, as sodium rather than hydrogen is the charging agent), and polar lipids. For a given compound, the signal strength (peak height in the spectrum) increases linearly with the analyte concentration over a wide range until saturation occurs. Similarly, very low flow rate (below the μl/min range) also favor high sensitivity as

37

the inner diameter of the electrospray needle can be reduced to produce smaller droplets with enhanced ionization potential for the analyte.

There does not seem to be an upper mass limit for analysis by ESI-MS. Large ions are typically multiply charged (proteins and peptides by added protons in the positive mode and abstracted protons in the negative mode), which brings them into the mass-to-charge (*m/z*) measurement range of most mass analyzers. The distribution of charges for an analyte gives rise to the typical multiple charge envelop. The resulting spectrum can then be simplified by deconvolution, an algorithm that sums up the signal intensity into a single peak at the molecular weight of the analyte. Very complex mixtures can be analyzed by ESI-MS, but spectra become increasingly difficult to interpret as the molecular weight of the components and their number increases.

Electrospray can be performed either in infusion mode, the so-called nanoelectrospray format, or in combination with a liquid separating technique, such as high-performance liquid chromatography. When these two techniques are coupled in the so-called LC-MS configuration, the eluting components of the sample can be analyzed online by the mass spectrometer. In this scenario, sample cleanup, separation, and concentration are all achieved in a single step.

1.4.3 The Mass Analyzer

The mass analyzer is the part of the mass spectrometer in which ions of given m/z ratio can be separated from each other, therefore conferring to the mass spectrometer its analytical power. In the field of proteomics, critical parameters are mass accuracy, sensitivity, resolution, and the ability to generate fragment spectra (tandem MS) informative enough to infer the identity of the parent ion.

MALDI and ESI ionization technique have been typically coupled to five types of mass analyzers: the ion trap, the time-of-flight (TOF), the quadrupole, the Fourier transform ion cyclotron (FT-MS) and the electrostatic ion trap (Orbitrap) analyzers. All these analyzers present different characteristics and performance profiles; most of them can be used either stand-alone or in combination with additional analyzers, either to combine the analytical strengths of two complementary ion selection strategies or to perform structural analysis of molecules by using successive ion selection processes. A summary of the most common mass analyzer configurations used in the proteomic field is shown in Fig. 1.8.
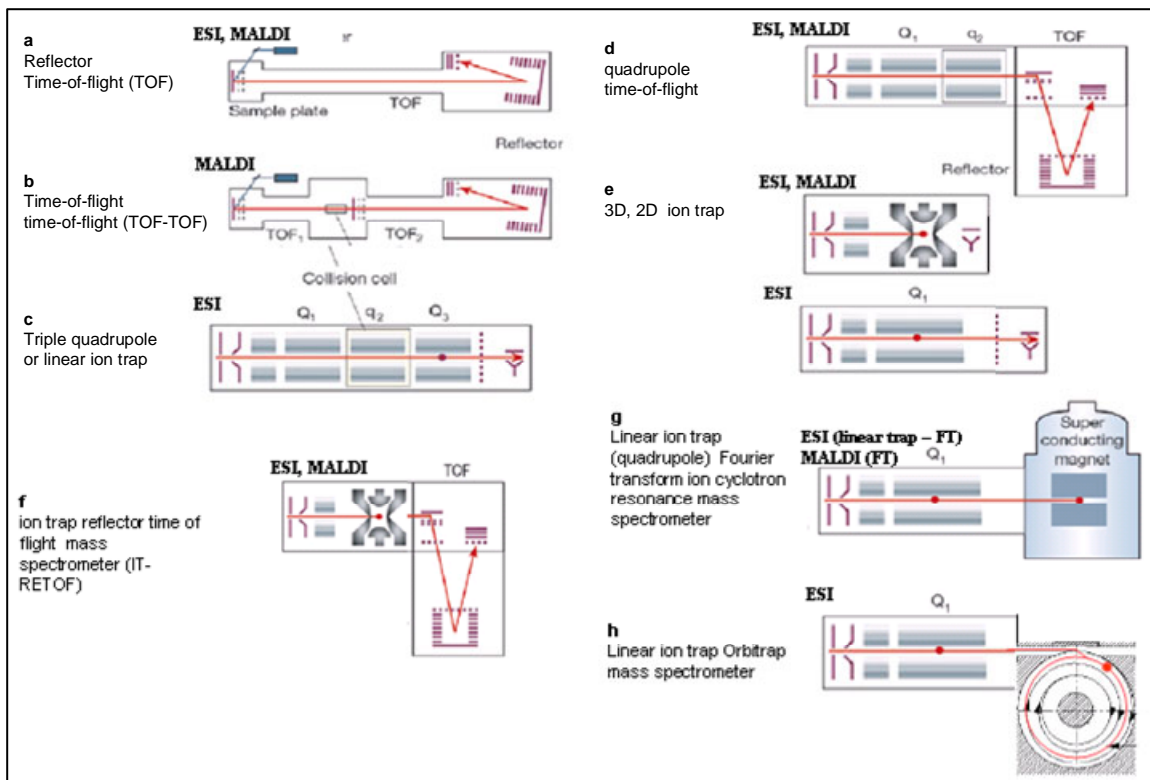
Figure 1.8: *Instrumental configurations of the commonly used mass spectrometers in proteomics. The typical ionization technique that is used for each of those mass spectrometers is indicated (Adapted from Aebersold, R and Mann, M. (40)).*

All ion trap mass analyzers share the ability to store ions for a certain time before ejecting them sequentially to a detector or to another mass analyzer. Their trapping capability make them very sensitive; however, only a limited number of ions can be accumulated at their point-like centre before space-charging distorts their distribution, which results in relatively low mass accuracy. The 'linear' or 'two-dimensional ion trap' (41) is a recent development where ions are stored in a cylindrical volume that is considerably larger than that of the traditional, three-dimensional ion traps. Linear ion traps can be operated either as a quadrupole (continuous) or as an ion trap (pulsed) analyzer, allowing increased sensitivity (1-2 times), higher trapping capacity (40 times) and trapping efficiency (11-14 times) than conventional 3D traps. The FT-MS and the Orbitrap mass analyzers are also ion trapping instruments but they capture ions under high vacuum in a high magnetic or electrostatic field respectively, which enables exquisite mass accuracy, high resolution, and dynamic range. In their commercial versions, however, these two mass analyzers have been used as ion detectors and have been coupled with other types of analyzers, such as a linear ion trap or a quadrupole, for ion isolation and selection.

Time-of-Flight (TOF) mass analyzers are one of the simplest separating devices in the mass spectrometric field. Incoming ions are accelerated to high kinetic energy and are separated along a flight tube as a result of their different velocities before being detected. In a more complex configuration, the so-called "reflector time of flight" (reTOF) mass analyzer, incoming ions are turned around in a reflector (an electrostatic mirror) which compensates for slight differences in kinetic energy which might otherwise severely limit the resolution obtained by such mass analyzers.

Quadrupole are versatile mass analyzers that can be used either as focusing lens (all ions go through at any time), as a time-dependent ion scanning device (ions of a given m/z ration are transmitted one after the other), or as a static mass filter (only ions of a given m/z ratio can be transmitted at any time). Quadrupole's selectivity relies on an electric field of various intensity rotating between four hyperbolic rods, allowing a stable trajectory only for ions of a particular desired m/z range. While quadrupole-based mass spectrometers used to be quite common a decade ago, the analyzer's rather low resolution has limited them to special mass spectrometric applications, such as quantification by multiple-reaction monitoring, or in combination with other mass analyzers, such as in the hybrid quadrupole-Time-of-Flight (Q-TOF) mass spectrometer.

Structural information of peptides is typically gained by collision-induced dissociation using tandem mass spectrometers. In this technique, an ion of interest is selected in the first mass spectrometer and is then dissociated by collision with an inert gas, such as argon or nitrogen, in a collision cell. The resulting fragments are then transmitted to the second mass spectrometer, producing the tandem mass, or MS/MS, spectrum. In most instruments in use today, multiple collisions impart energy onto the molecule until it fragments. (This is low-energy fragmentation, in which any single hit is not sufficient to break a peptide bond. In high-energy fragmentation, the molecules have higher velocity and a single hit can break bonds). The most common ion types produced in the low-energy fragmentation mode are b and y ions, after fragmentation of the amide bond with charge retention on the N or C terminus, respectively (Fig. 1.9). Several of the mass spectrometers depicted in Fig. 1.8 are suitable for tandem mass spectrometry. For example, a TOF-TOF instrument (Fig. 1.8, panel b) includes a collision cell between two TOF sections. Ions of a given mass-to-charge (m/z) ratio are selected in the first TOF section, fragmented in the collision cell, and the masses of the fragments are separated in the second TOF section. In triple-quadrupole instruments, (Fig. 1.8, panel c), ions of a particular m/z are selected in a first section (Q1), fragmented in a collision cell (q2), and the fragments are separated in Q3. The quadrupole TOF hybrid

instrument combines the front part of a triple quadruple instrument with a reflector TOF section for measuring the mass of the product ions. This instrument combines the versatility
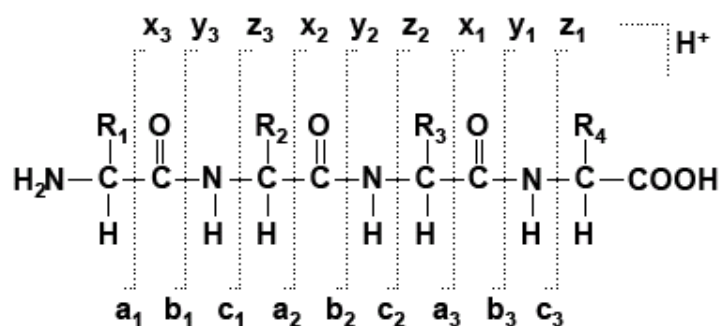
Figure 1.9: *Product ions generated in the fragmentation of peptide molecules by tandem mass spectrometry. Collision-activated dissociation (CID) causes single cleavage to occur more or less randomly at the various amide bonds in the collection of peptide molecules. This process generates a series of fragments that differ by a single amino acid residue. Ions of type y contain the C terminus plus one or more additional residues. Ions of type b contain the N terminus plus one or more additional residues. Additional ion types such as a, c and x and z types correspond to cleavages at different positions in the backbone (Adapted from Lottspeich and Zorbas (32)).*

of a triple quadrupole instruments with the mass accuracy and resolution expected from a TOF analyzer. Finally, ion traps (Fig. 1.8, panels e, g, h) represent a special type of tandem mass spectrometer as ion traps can be operated both as a trapping device and a collision cell. Ions of interest with specific m/z ratio are first accumulated in the trap and then excited by resonance. The resulting fragments ions are trapped again in the ion trap analyzer until they are sequentially ejected from the trap for detection.

1.4.3.1 The Linear Ion Trap-Orbitrap Mass Spectrometer

The ever increasing demands in mass accuracy and resolution from research areas such as proteomics and metabolomics have stimulated the development of novel types of mass spectrometers combining high performance with moderate size at an affordable price. The Orbitrap-based mass spectrometer, manufactured by Thermo Electron (now Thermo Fischer Scientific) based on the invention of Makarov (42, 43), is the first novel mass analyzer introduced in the market in decades.

The Orbitrap can be considered as a modified Kingdom trap or, more simply, as a modified form of a quadrupole ion trap using static electrostatic fields while the quadrupole ion trap uses a dynamic electric field (44). A cross-view of the Orbitrap mass analyzer is shown in Fig. 1.10.
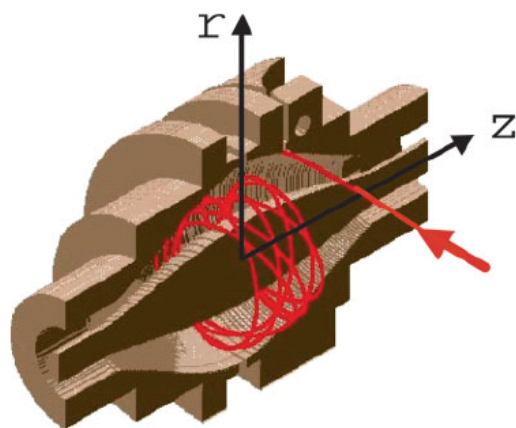
Figure 1.10: *Cross-view of the Orbitrap mass analyzer. Ions are injected into the Orbitrap at the point indicated by the red arrow. The ions are injected with a velocity perpendicular to the long axis of the Orbitrap (the z-axis). Injection at a point displaced from z=0 gives the ions potential energy in the z-direction. Ion injection at this point on the z-potential is analogous to pulling back a pendulum bob and then releasing it to oscillate. Figure adapted from Hu et al., (45).*

Ions injected into the Orbitrap mass analyzer assume a circular movement in the trap due to the electrostatical field. The specially shaped electrodes produce an electrostatic potential containing no cross-terms in *r* and *z* so that, when ions are injected at a point displaced from *z*=0, the resulting potential in the *z*-direction is exclusively quadratic. Ion motion along the *z*-axis may be described as an harmonic oscillator and is completely independent of *r, φ* motion (φ is the angular coordinate). Therefore, an ion mass/charge ratio *m/z* is simply related to the frequency of ion oscillation ω along the *z*-axis (Eq. 1)

$$\omega = \sqrt{(z/m) \cdot k} \qquad (1)$$

The main components of the commercially-available LTQ-Orbitrap hybrid mass spectrometer (Thermo Electron) are illustrated in the figure 1.11 (see also Hu et al. 2005 for a detailed description of the first orbitrap prototype). An electrospray source creates ions, which are transferred using RF-only guide quadrupoles and octapoles into a a linear quadrupole ion trap. The ion guide brings the ions through several stages of differential pumping, from the atmospheric pressure of the ion source to approximately 10^{-5} Torr in the LTQ ion trap. The LTQ ion trap is required to couple the continuous electrospray ion source with the Orbitrap, which operates in a pulsed fashion. After a sufficient number of ions has been accumulated (typically taking 10–400 ms), the back lens of the linear ion trap is pulsed open and ions are collected in a C-shaped ion trap, which is used to store and collisionally cool ions before injection into the orbital trap as defined ion packets with very small temporal (100–200 ns)
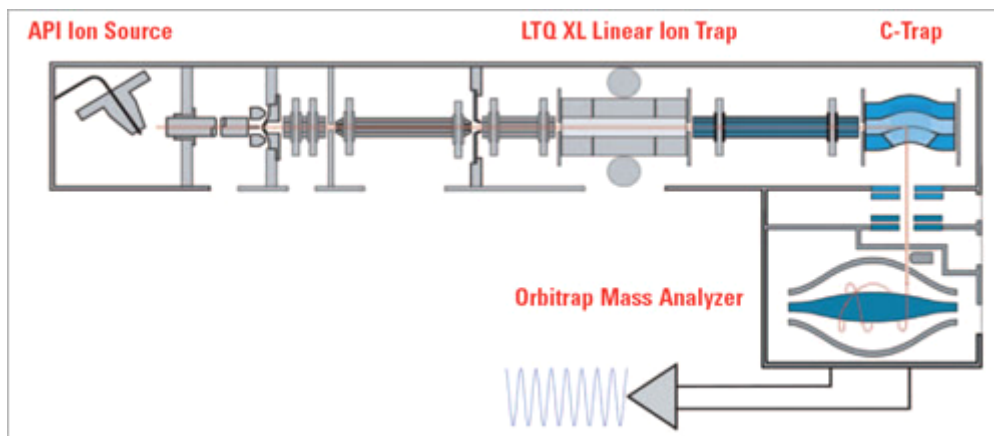
Figure 1.11: *The commercially-available LTQ-Orbitrap mass spectrometer. Ions are produced by the electrospray ion source and then proceed through the source, the focusing lenses and octapoles to pass into the linear ion trap (LTQ). The LTQ mass analyzer can be used either as a stand-alone instrument, submitting ions to MS and MSn analysis, or it can serve as an ion accumulator and buncher, allowing the pulsed Orbitrap mass analyzer to be coupled to the continuous electrospray ionization source. After accumulation of a sufficient number of ions, they are axially ejected from the LTQ and collected in a C-shaped ion trap (C-Trap) from which they are passed into the Orbitrap mass analyzer. The trapped ions assume circular trajectories around the center electrode and their axial oscillations, along the center electrode, are detected and transformed back to mass spectrum using fast Fourrier Transform (Figure downloaded from the Thermo Fisher web site (http://www.thermo.com/) on July 19, 2008.*

and spatial spread. Once injected into the Orbitrap at a position offset from its equator, ion bunches start coherent axial oscillations without the need for any additional excitation. All ions have exactly the same amplitude, although ion packets of different mass/charge ratios will execute their axial oscillations at their respective frequencies. The exquisite resolution attained in the orbitrap mass analyzer is only possible if the ion packet remains spatially coherent during the analysis time (up to several seconds); beside superior manufacturing, the performance of the instrument is made possible by keeping the orbitrap analyzer at very high vacuum, approximately 10^{-10} Torr. The detection of an ion image current due to motion along the Orbitrap axis is made possible by splitting the outer electrode in half at $z=0$. The ion current is differentially amplified from each half of the outer electrode and then undergoes analog-to-digital conversion before processing by customized control and acquisition software. The so-called resulting transient is the converted back to a mass spectrum using fast Fourrier transform.

The superior performance of the LTQ-Orbitrap mass spectrometer was critical to carry on the protein identification of the BBM membrane and to enable the initial steps aiming at developing a label free quantification approach based on LC-MS/MS data. This instrument's high mass accuracy (1-3 ppm), excellent resolution (up to 100,000 depending on the

acquisition time) and extended dynamic range (more than 3 orders of magnitude) made the Orbitrap the ideal analytical tool for this study.

1.5 Data analysis and Bioinformatics tools

The analysis of the large amount of data generated in mass spectrometry-based experiments is a significant challenge and is currently a bottleneck in many proteomics approaches. Nevertheless, a significant progress has been made in the form of software package for proteome analysis and their application to biological and clinical research.

The vast majority of proteomic data nowadays are being generated by mass spectrometry, more specifically, by tandem mass spectrometers of ever increasing performance. These instruments and the diverse workflows they support have in common that they generate hundreds to tens of thousands of fragment ion spectra per hour of data acquisition. The assignment of these fragment ion spectra to peptide sequences, the inference of the proteins represented by the identified peptides and the determination of their abundances in the analyzed sample present complex computational and statistical challenges. It is essential for proteomics to develop and generally apply tools and solutions to these problems to provide accurate and reproducible results.

1.5.1 Protein Identification

The protein identification process follows a well-defined analytical scheme requiring the consideration of several parameters to return a successful and unambiguous answer. Not surprisingly, the type of sample preparation, choice of instrument, data acquisition and peak peaking parameters define which computational identification process can be selected and the scope of results that can be obtained.

All protein identification algorithms used in high-throughput mass spectrometry-based proteomics experiments share in common the characteristic to compare an experimentally recorded mass spectrum with the protein/peptide information derived from a sequence database. The quality and the significance of the match are then provided back to the user in the form of more-or-less elaborate statistics, a topic which has seen many progresses recently. A few years ago, the experience of the mass spectrometer's operator was often determinant to define at which thresholds a protein or a peptide hit was relevant in an experiment. The development of probabilistic approaches to rank the likeliness of protein/peptide identification and the use of decoy databases to estimate the false positive identification rate has been determinant to define generally applicable rules to protein identification.

There are two strategies commonly used for high-throughput protein identification based on mass spectrometry, the Peptide Mass Fingerprint (PMF) and Fragmentation Mass Fingerprinting (FMF).

1.5.1.1 Peptide Mass Fingerprint (PMF method)

Peptide mass fingerprinting (PMF) was the first available method for large-scale protein identification using mass spectrometry, and is still widely used (46, 47). This strategy compares the signals observed in a MS analysis with theoretical spectra obtained from an *in silico* digestion of each protein sequence from a reference database (Fig. 1.12). One of the main limitations of the PMF method consists in the basic assumption of the algorithm that the signals in the experimental spectrum must originate from one or from very few proteins, making this approach incompatible with the identification of proteins in complex mixtures. As a result, protein identification using PMF has been mostly carried out in proteomic approaches based on 2D gel electrophoresis, for fractionation of a biological sample, combined with MALDI TOF MS analysis of the usually rather pure 2D gel spots.
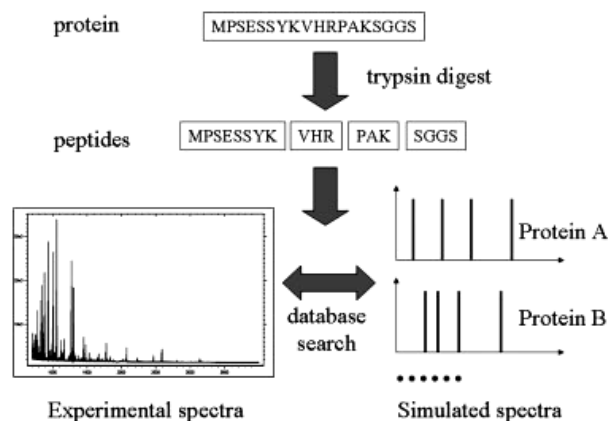
Figure 1.12: *PMF protein identification scheme. The protein in the sample is digested into peptides, whose mass-to-charge ratios are shown in the PMF spectra. The PMF spectra are compared with the simulated spectra of the in silico-digested proteins. Figure adapted from Song, Z. et al., (48).*

Used with the proper data, PMF remains popular and works well in practice because of its speed to compute PMF scores against a database. PMF yields in many cases protein identifications with high confidence, especially with organisms with smaller genomes. Interestingly, a confident PMF identification score can usually be obtained even in the presence of point mutations or occasional post-translational or chemical modifications, as long as a minimal number of peptides characteristic for a protein remains detectable in the spectrum. A non-confident identification score may pinpoint to splice variants, individual

sequence variants and errors in the database, or more prosaically, to the presence of a protein mixture in the sample or to incomplete digestion. As more sophisticated methods for scoring PMF are being developed, more proteins may now be identified with confidence (49).

Another limitation of the PMF strategy is its sensitivity to database size. There is a direct correlation between the statistical confidence a PMF algorithm can ascribe to protein identification and the size of a sequence database. A larger database has an elevated chance of experimental masses randomly matching theoretical peptide masses in these databases, thereby decreasing the confidence of protein identifications using PMF. The same observation is also valid if mass spectra are searched against a database using very relaxed search criteria (for example, by relaxing the enzymatic specificity a peptide has to display for being considered in the search, or by increasing the error by which theoretical masses are allowed to match experimentally-determined peptide masses) or if many dynamic modifications are considered simultaneously (oxidation, phosphorylation, etc., which increase the actual size of the search database).

Popular search engines to mine PMF data include Mascot, MS-Fit and ProFound. All these tools allow the user to set up their searching parameters, such as for example database choice, constant and variable modifications, mass accuracy error, and miss cleavages. Each engine returns the result of the search with a measure of confidence gauging how significant is the obtained identification. For example, Mascot (50) uses a proprietary algorithm based on the MOWSE scoring system (47). By calculating the distribution of peptide lengths across the entire search database, a probability can be calculated for each observed peak for this match being purely random. While, most of the time, all PMF search engines returns similar answers, it is also not uncommon that multiple search of PMF data by using different tools leads to different identifications because of the different scoring systems that each of these tools use.

1.5.1.2 Fragmentation Mass Fingerprinting (FMF)

The currency of information for tandem mass spectrometry-based proteomics is the fragment ion spectrum (MS/MS spectrum) of a specific peptide ion that was fragmented, typically, in the collision cell of a tandem mass spectrometer. The correct assignment of such a spectrum to a peptide sequence is a first and central step in proteomic data processing. A large number of computational approaches and software tools have been developed to automatically assign peptide sequences to fragment ion spectra. These can be classified into three categories: (i) Fragmentation Mass Fingerprinting (FMF), by correlating un-interpreted fragment ion spectra with theoretical spectra predicted for each peptide contained in a protein sequences database,

or by correlating acquired fragment ion spectra with libraries of experimental MS/MS spectra identified in previous experiments (spectral library searching); (ii) De novo sequencing, where peptide sequences are explicitly read out directly from fragment ion spectra; and (iii) hybrid approaches, such as those based on the extraction of short sequence tags of 3–5 residues in length, followed by 'error-tolerant' database searching.

The FMF strategy is to date the most frequently used peptide identification method for large-scale proteomics studies; it was also used in this study and, for that reason, it will be described in more details in the following paragraph.

Several MS/MS database search engines have been developed to support the FMF approach and their basic functionality is illustrated in the Fig. 1.13. Mascot and Sequest are the most commonly used programs for peptide identification using the FMF strategy. Mascot is based on the probabilistic MOWSE algorithm which uses the parent mass and the relative abundance of the derived fragment masses for that parent as constraints on the search space. The Sequest peptide identification process starts with the selection of candidate sequences based on the length of a continuous y/b fragment ions correlating with the input spectrum. A second, more sensitive and more computational-intensive correlation function returns scores based on the presence of significant peaks in the experimental spectrum corresponding to the expected y/b ion peaks in each theoretical spectrum in the database. Sequest outputs a ranked list of candidate peptide identifications.
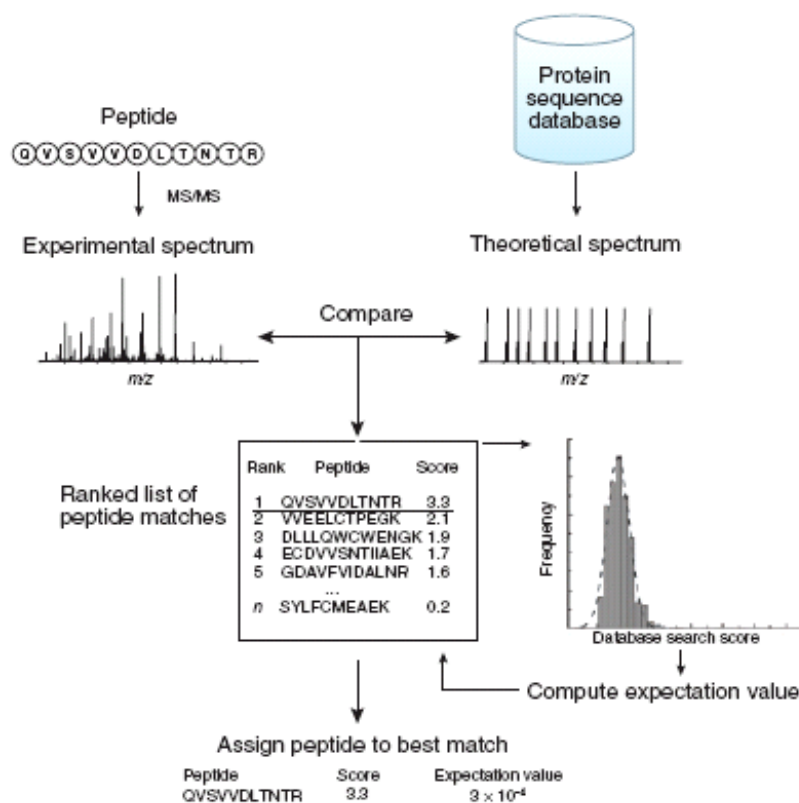
Figure 1.13: *Peptide identification by MS/MS database searching*. *An acquired MS/MS spectrum is correlated against theoretical spectra constructed for each database peptide that satisfies a certain set of database search parameters specified by the user. A scoring scheme is used to measure the degree of similarity between the spectra. Candidate peptides are ranked according to the computed search score, and the highest scoring peptide sequence is selected for further analysis. The figure was adapted from Nesvizhskii et al., (51).*

Regardless of the search engine used, the fraction of spectra identified is usually quite low, with 10–20% being typical, although some high-quality experiments may yield as high as 50% identification rates. This is largely dependent on the quality of the sample, the type of mass spectrometer used, and the manner the mass spectrometer has been set up.

One of the main weaknesses of this identification strategy is that a top-scoring peptide, as determined by the search engines, may not necessarily be the correct candidate. The remaining challenge is to determine whether the putative identifications are in fact correct. Five years ago, a standard procedure might have been to apply an arbitrary cutoff score and to manually inspect spectra and judge correctness. This was extremely labor intensive and subjective, thus not reproducible. Recently several computational strategies have emerged for validating the search results and assigning a false discovery rate at the peptide level for a given threshold. One of those strategies is to repeat the identification search using a reverse or a shifted database by applying the same searching criteria, thus allowing the estimation of a

false discovery rate for a given experiment (52). An alternate strategy consists of storing MS/MS data in a library at an early stage of the identification process. Comparison of MS/MS data occurs by comparing experimental spectra with those previously measured and stored in a database using a spectra-matching algorithm. Such an approach was proven to be very effective in the small molecule area (53).

1.5.2 Validation of peptide and protein identification

The purpose of most proteomic experiments is not the identification of peptides, but the identification of the proteins present in the sample before digestion. Thus, the peptide sequences which were identified from the fragment ion spectra need to be grouped accordingly to their corresponding protein, and a confidence measure needs to be recomputed at the level of proteins. This process is by far not straightforward owing to several challenges, and it has not been satisfactorily resolved up to today. First, many correctly identified peptides tend to cluster into a relatively small number of proteins, especially when there were very abundant proteins in the analyzed sample. At the same time, incorrect spectral identifications match randomly to a much larger number of proteins in the searched sequence database. Thus, almost every high scoring, incorrect spectral assignment introduces one additional incorrect protein identification, resulting in a significant increase in the false discovery rate when going from the peptide to the protein level. Second, an important number of peptides are not unique to a single protein and hence will point to more than a single entry in the protein sequence database. Such cases most often are the result of the presence of homologous proteins, splicing variants, or redundant entries in the sequence database. This problem is particularly serious in the case of higher eukaryotes, or when protein sequence databases are derived from genomic databases. Therefore it becomes critical to have an appropriate tool that is able to assess the validity of the protein inference and associate a probability to it. A tool such as ProteinProphet combines probabilities assigned to peptides identified by MS/MS to compute accurate probabilities for the proteins present (54).

In order to increase the confidence level for peptide or protein identification, technical replicates of an analyzed sample might be necessary. In an LC-MS/MS proteomic approach, unique peptide coordinates (such as retention time and accurate mass) could be useful for validating identification.

Another method to confirm the presence of a protein via the characterization of one (or several) of its constitutive peptides is to use Multiple Reaction Monitoring (MRM). In the MRM-targeted strategy, which is typically performed the same way as for a quantification

experiment (see section 1.6.2), one sets the mass spectrometer to specifically detect signature ions of the peptide of interest, usually in presence of its isotopically labeled counterpart to pinpoint the exact retention time in the analysis and to validate the chosen transitions. Performed on a triple quadrupole instrument (or hybrid quadrupole/linear ion trap instrument), the data acquired by setting both mass analyzers to predefined m/z values achieve increased selectivity while the non-scanning nature of this experiment accounts for an increased sensitivity. It thus allows detection and structural confirmation of low abundance analytes in complex biological samples.

1.6 Quantification in Proteomics

The determination of the relative levels of protein abundance in organisms or tissues exposed to different physiological or environmental conditions is essential to the study of disease processes and cellular responses to stress. Until recently, a quantitative proteomics experiment would have typically been performed by differential display analysis using two-dimensional gel electrophoresis (2DE), whereas protein spot intensities on associated gels would have been compared and the differences quantified. While this method compares favorably to the precision achieved by modern mass spectrometers in terms of sensitivity and reproducibility (especially in the DIfferential Gel Electrophoresis [DIGE] strategy pioneered by Amersham BioSciences, today GE Healthcare), its dependence on the two-dimensional gel electrophoresis platform limits its applicability and throughput, especially when fractionation is required. In the last few years, especially with the popularization of high-performance mass spectrometers around laboratories, several alternative strategies have been proposed for relative quantification via mass spectrometry (55).

In this section three conceptually different strategies to perform quantitative LC-MS experiments will be presented and are summarized in the Fig. 1.14. In the first approach, quantification is achieved by spectral counting of fragment ion spectra assigned to a particular peptide; in the second approach, quantification is achieved by the differential detection of peptides or proteins compared to an internal standard in the form of their isotopically-labeled counterpart; and in the third approach, peptides and proteins are quantified "label-free" (that is, without internal standard) based on the precursor ion signal intensities.
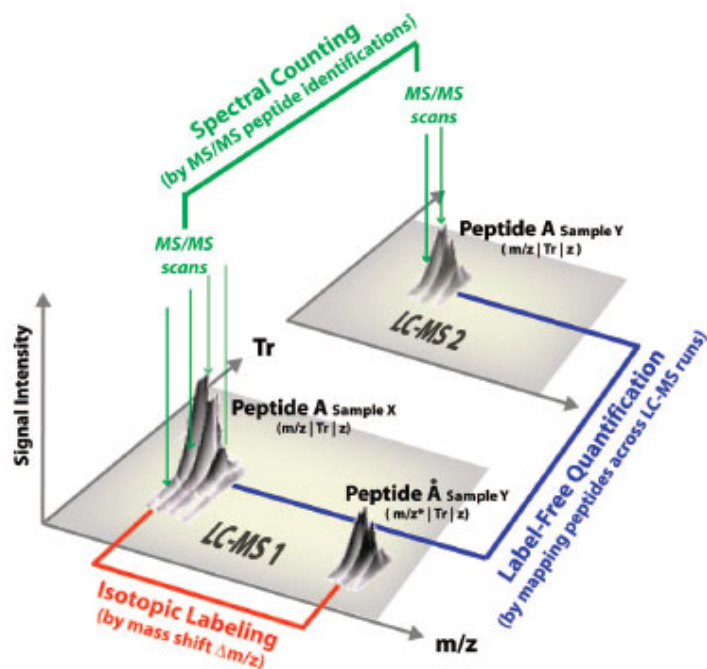
Figure 1.14: *Representation of LC-MS based quantification strategies in proteomics. In the first approach, spectral counting estimates abundance values based on the number of times a peptide was successfully identified by tandem mass spectrometry (MS/MS) and compare these across experiments (green). In the second approach using isotopic labeling, quantification is based on the differential analysis of two samples, X and Y, in the same LC-MS run whereas the peptide A (in sample X) and its stable isotope labeled counterpart A[*] (in sample Y) are detected by their characteristic mass difference Δm/z (red). In the third approach using label-free quantification, peptide signals are extracted by tracking the isotopic patterns along their chromatographic elution profile. Relative quantification between the LC-MS run 1 and the LC-MS run 2 is achieved by comparing the coordinates m/z, Tr and z of the peptide (blue). This picture is adapted from Mueller et al., (56).*

1.6.1 Semi-quantitative analysis based on Spectral Counting

This type of analysis was originally developed to evaluate semi-quantitatively shotgun proteomic approaches. In this strategy, wherein the main goal of the experiment is to identify a maximum number of peptides per LC-MS run, the mass spectrometer is set up to automatically select a number of peptide precursor ions from the full scan MS for collision induced dissociation (CID) analysis according to predefined criteria (typically the 1-5 most intense precursor ions of every full scan is selected for MS/MS analysis). The quantification strategy is based on the number of times a precursor ion peptide is selected for fragmentation and identified in a large data set. This value is related to the abundance of a peptide represented by its precursor ion in the sample mixture. Spectral counts of peptides associated with a protein are then averaged into protein abundance index (57-60).

Spectral counting approaches have most frequently been used for the analysis of low to moderate mass resolution LC-MS data as a convenient, fast, and intuitive quantification

strategy. A draw back in this type of approach is its dependency on the quality of MS/MS peptide identification, because errors in the assignment of peptides propagate directly into protein abundance indexes. The assembly of spectral counts of peptides into a protein index is unproblematic for peptides whose sequences belong to only one protein; however, it is difficult to resolve protein for which peptides map to multiple protein sequences (as, for example, for peptides from conserved protein regions). It has been suggested, therefore, that spectral counting ought to be used only with proteotypic peptides (61).

Another hurdle with spectral counting is the manner spectral counts are computed if only a small number of peptide identifications have been achieved for a given protein. This situation can occur, for example, for proteins of low abundance (only a few representative peptides might be detected), for proteins of low molecular mass or with specific physicochemical characteristics that limit peptide observability (these proteins might generate only a few representative peptides for LC-MS/MS measurements), if peptide identification becomes ambiguous or if the sample complexity overwhelms the mass spectrometer's dynamic range. As a result, the spectral counting quantification method should be used for high abundant proteins, based on a large number of peptide identifications. It becomes unreliable for proteins with one- to two-peptides identifications since low-abundance peptide identification is not reproducible between different LC-MS runs (precursor masses selection for MS/MS is skewed towards peptides of high abundance.

In a modified quantification approach, peptide abundance can also be estimated from the extracted ion chromatograms (XIC), which, in this way, partially alleviate the MS/MS problem (62). In this approach, it is sufficient that a peptide has been successfully analyzed by MS/MS in at least one of all the LC-MS runs. The mass to charge ratio m/z and the retention time T_R of the MS/MS of the precursor ion are then used for the extraction of the XIC of this peptide in the other LC-MS measurements. In this case, however, any fluctuation on the LC system will cause a drift on the retention time of the peptide (see also section 1.6.3).

In summary, quantification approaches based on spectral counting or on the extracted ion chromatograms of an assigned peptide represent valuable techniques for LC-MS/MS data of low- and medium-resolution. This strategy is not recommended for underrepresented peptides as the MS/MS analytical strategy is biased for high abundant peptides in a mixture.

1.6.2 Relative Quantification based on Differential Stable Isotope labeling

Differential stable isotope labeling in combination with mass spectrometry has become an extremely popular method for quantitative proteomics (40). Absolute quantification has been

demonstrated whereas the peptides of a sample are compared against their spiked-in isotopic labeled synthetic analogues (63). Differential isotopic labeling also allows the relative quantification of peptide intensity between multiple biological samples within a single LC-MS measurement.

A variety of MS-compatible labeling techniques are now available. They can be roughly divided into two categories, isobaric or isotopic labeling strategies. Fig. 1.15 illustrates the workflow of these two approaches.
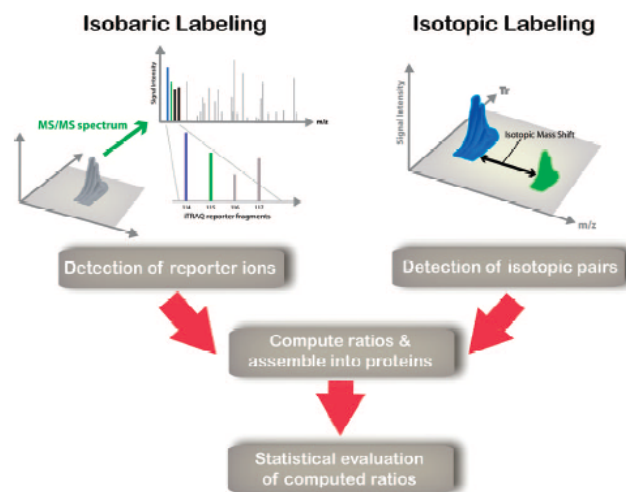
Figure 1.15: *Schematic representation of the quantitation principles of isobaric and isotopic labeling. Isobaric labeling generates in the MS/MS different reporter ions that are used to calculate peptide abundance between different samples. On Isotopic approaches differentially label peptides or proteins from two samples (green / blue) produce isotopic pairs of characteristic mass shifts. Common to both approaches are the next steps where peptide ratios are computed, assembled into protein ratios and then statistically assessed to evaluate the significance of the detected fold changes. Figure is adapted from Mueller et al.,(56).*

Isobaric labeling reagents, as exemplified by the iTRAQ reagents (64), are peptide tags that introduce an identical mass shift to all derivatized peptides (that is, these peptides are undistinguishable in the MS spectrum), but produce MS/MS-specific reporter ions that are used for quantification. Differentially labeled samples are combined and simultaneously analyzed by shotgun LC-MS/MS analysis, and peptide abundance values are compared via the reporter fragments in the MS/MS spectra. In contrast, isotopic labeling methods, for example the ICAT reagents (65) and the SILAC protein labeling strategy (66), generate pairs of peptides with characteristic mass differences introduced by the applied label. Typically, the isotopic forms of an MS/MS identified peptide are detected by their mass shifts and identical elution profiles and are used to compute a peptide abundance ratio between the "heavy"

labeled peptide from one condition and the "light" peptide version of the other condition (Figure 1.14, blue and green coloring).

Isobaric Labeling of Peptides can be achieved, for example, with the new 8-plex iTRAQ labeling reagent that allows the simultaneous quantification of up to eight biological samples (Choe et al., 2007). The isobaric reagent reacts with primary amino groups and produces in the MS/MS fragmentation spectrum eight different unique reporter groups, one per reagent flavor, at 113, 114, 115, 116, 117, 118, 119, and 121 *m/z*. iTRAQ labeling does not increase the sample complexity at the MS level because the reagent is based and relies on a fully MS/MS-dependent workflow. Therefore, the strength of the approach, and its limitation, is that only peptides subjected to CID fragmentation which could be successfully assigned to a peptide sequence are quantified.

The Isotopic Labeling Techniques commonly used in proteomics for relative quantification experiments can be divided into three general categories: chemical tagging of peptides and proteins, stable isotope labeling of peptides during enzymatic digestion, and metabolic labeling of living cells and animals.

The first strategy using chemical tagging of specific amino acids is the Isotope Coded Affinity Tag (ICAT) method. In this technique, two biological samples are labeled at the peptide or protein level using chemically identical but isotopic different reagents specifically targeting cysteine groups. The labeled peptides differ in their molecular weight by 8 mass units (the newer ICAT reagents now introduces 9 mass units). After the labeling step, the samples are combined and subjected to mass spectrometric analysis. In the following years, a number of related strategies have been developed in which the sets of reagents differ in specificity, structure, mass difference, and number of isotopic forms (67). All these labeling techniques have in common that they generate a complex sample mixture consisting of pairs of chemically identical peptides of different mass. Those pairs are then detected in the precursor ion mass spectrum, and the signal intensity is used to compute the relative abundance of the respective analyte in either sample.

The incorporation of isotopic analogues as replacement of atomic elements in the peptide sequence can be done through enzymatic reaction. ^{18}O labeling is based on the incorporation of heavy oxygen into the C-termini of peptides during tryptic digestion of proteins in presence of heavy water (68). The major advantage of ^{18}O labeling is that the method is not limited to a specific subpopulation of peptides. This method is generic and can be applied in any kind of peptide sequence. However the optimization of the tryptic digestion process can be quite

difficult and the ^{18}O incorporation may not be very specific for the C-termini as other amino acids might be labeled, making data analysis more complex.

Metabolic labeling has become very popular since the development of the SILAC approach. In its original version, cell cultures are grown in media containing either light ^{12}C- or heavy ^{13}C-labeled arginine and lysine that are then metabolically incorporated into proteins through the cell cycle. After mixing the two samples and digestion, the generated isotopic peptide pairs are detected by mass shifts of multitudes of 6 mass units for each lysine or arginine incorporated. Since the label is added at a very early stage of the experiment, this technology allows normalizing for experimental variations introduced in the sample processing steps. The main limitation of the SILAC labeling strategy is that is largely restricted to biological material grown in culture, as is metabolic labeling of tissues by ^{15}N (a popular method to isotopic label bacteria (69, 70)), and thus is not generally applicable to tissues, body fluids, or clinical applications. A promising, but rather expensive approach for ^{15}N labeling of all the amino acids with high level of incorporation in mammalian (non human!) organisms has been suggested by (71). In this approach, a young (typically a newborn or a fetus) animal is fed for several weeks a special diet in which the nitrogen source is isotopic labeled, resulting in high rates of labeling in organs.

Multiple Reaction Monitoring (MRM) represents a MS-based quantification approach that is taking fully advantage of the labeling strategies mentioned above. MRM experiments differ from conventional proteomics approaches in that they are hypothesis-driven (i.e. screening for known or putative entities). As a part of this targeted approach proteotypic peptides (i.e. sequences that are unique to one single protein), and their corresponding fragmentation patterns are used to define MRM transitions. In such approaches, one starts by generating a list of proteins of interest, the corresponding proteotypic peptides are derived, and the associated fragment ions are predicted (or extracted from a database) to define the MRM transitions. Relative estimation of the elution times might also be made and all that information will then be used to generate a list of targeted peptides. Relative and absolute peptide quantification can be achieved by adding at some stage of the experiment the isotopic (not the isobaric!) labeled peptides and to measure all samples in presence of the same internal standards. As MRM is per se a multiplex methods (it has been claimed that several hundreds of peptides can be quantified in the same LC-MS experiment), it is thus possible to detect and quantify a large number of peptides in consecutive single LC/MS runs with minimal effort (72).

In summary, isotopic labeling strategies allow for the highly accurate quantification of LC-MS experiments since the differential analysis is performed on single LC-MS runs where the peptide pairs can be very accurately detected by the distinct mass shifts characteristic to the utilized labels. The described isotopic labeling strategies have in common that the combination of multiple, differentially labeled samples increases drastically the complexity of the peptide population. This is particularly problematic for the identification of peptide–ions by MS/MS where only a limited number of peptide signals can be subject to CID fragmentation during an LC-MS experiment. The selection of precursor ions is biased toward high-intensity peptide signals and leads to a prominent undersampling of low-abundance peptides. This limits the range of isotopic and, up to a certain point, isobaric labeling techniques since the computation of peptide/protein ratios is based exclusively on peptides that were successfully identified by MS/MS. Similarly, the chemical labeling strategies (for both the isotopic and the isobaric methods) increase intrinsically the variability of the experimental design because of the additional sample processing steps required by the procedure.

1.6.3 Label free quantitation of LC-MS data

Newer mass spectrometers with high scanning rates and high mass accuracy, such as the new generation of ESI-TOF, the LTQ-FT and the LTQ-Orbitrap instruments, have opened the way for the label-free quantification of LC-MS data. Typically, peptide signals are detected in the full MS scan and distinguished from chemical noise through their characteristic isotopic pattern. These patterns can then be tracked across the retention time dimension and used to reconstruct a chromatographic elution profile of the monoisotopic peptide mass. The total ion current of the peptide signal is then integrated and used as a quantitative measurement of the original peptide concentration (see Fig. 1.15). In principle, every peptide signal within the sensitivity range of the MS analyzer can be extracted and quantified independently of a MS/MS acquisition (73). This strategy increases the dynamic range for peptide detection and minimizes the bias against low-abundance peptides commonly observed for MS/MS-based approaches. Full MS measurements performed by the mass spectrometers of the latest generation provide very high resolution power and mass precision in the low parts per million range. Thus, the peptide signal extracted from the full MS level has become so precise and specific that the quantification process may now be uncoupled from the identification process.

In contrast to a differential labeling experiment, biological samples assessed in a label-free experiment are measured separately. The extracted peptide signals are then mapped across

multiple LC-MS measurements using the coordinates "mass to charge" (m/z) and retention time (Tr). A successful peptide tracking depends on the available mass resolution of the utilized mass spectrometer and the stability of the LC system. In this paradigm, the Tr reproducibility of each peptide across the LC-MS runs becomes critical and requires the use of bioinformatics tools for the correction of chromatographic fluctuations. High mass accuracy MS instruments in combination with sophisticated computational methods for signal intensity normalization offer a platform for the automated label-free quantification of complex biological mixtures in a linear dynamic range of 3 to 4 orders of magnitude without the requirement of MS/MS information.

Biological information is not directly inferred from a list of differentially-regulated masses and requires structural information, typically by tandem mass spectrometry. In a label-free experiment, even when performed on a modern mass spectrometer, typically only a fraction of the detected and quantified peptide features are identified by an MS/MS peptide assignment (74). Fig. 1.16 schematically represents some of the processes to follow the identification of peptides showing interesting quantitative patterns.

Peptide signals are extracted from the full MS (gray dots) and then tracked across LC-MS multiple measurements to generate a list of aligned peptides (red dots), which serves as a framework for further annotation efforts (75, 76). Each peptide signal is defined by its accurate mass and time tag (AMT) (77) on the mass and retention time dimension (blue flag), and a subset of these peptide elements are selected according to a specific property of the peptide signal (for example, here the intensity). These AMT tags are then subjected to post annotation where the identity of an AMT (i.e., protein name and peptide sequence) is derived by comparing its accurate mass and retention time coordinates to a map of AMT tags from successfully identified peptide signals (green flags). AMT tags can also be obtained from peptide identifications either in a targeted LC-MS/MS runs (inclusion list run) or in a large repository of MS/MS peptide identifications compiled from multiple LC-MS experiments (78) (MS/MS repository).
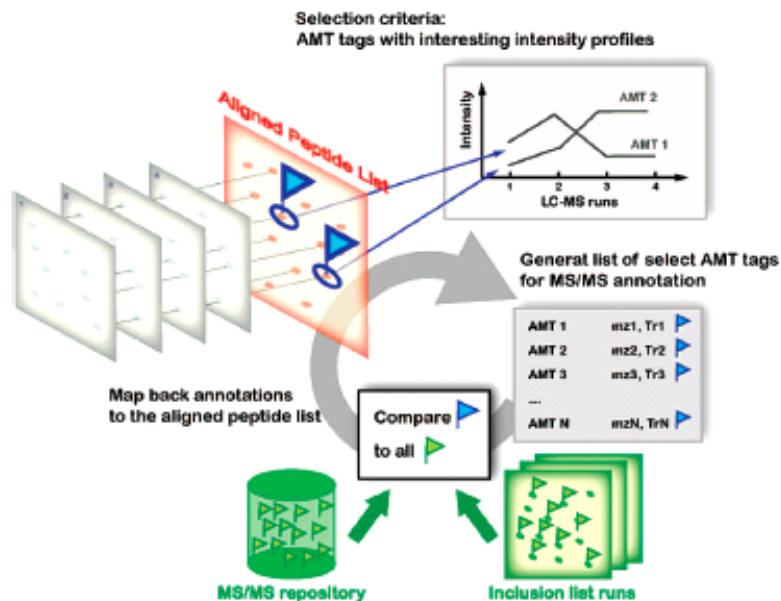
Figure 1.16: Schematic representation of post processing approaches to follow peptide characterization based on their accurate mass and retention time coordinates (AMT). An aligned peak list (red map) is generated by peak detection and alignment of multiple LC-MS runs (gray maps) AMT tags of peptide signals (blue flags) with interesting intensity profiles are then integrated into a list of AMT tags for annotation. Identities of AMT tags are derived from a map of identified peptides (green flags) by comparing AMT coordinates (mass and retention time) across LC-MS runs. These maps might be fed through a large repository of collected MS/MS peptide identifications or through targeted MS/MS via inclusion list runs. Figure is adapted from Mueller et al., (56).

In summary, label-free quantification represents a key technology for the analysis of complex biological samples due to its ability to exhaustively sample peptide signals detected directly at the full MS level. However, in order to be able to use this approach, there is a need for sophisticated computational methodology to process the acquired data.

In all the quantification approaches in order to be able to trust differential expressions of the quantified proteins it is very important to know what is the variability due to technical steps or biological replicates so that can be estimated what is significant different. Monitoring the technical variability and stability of the BBM preparation and the LC-MS system is part of this study, in order to develop the right tools and set the criteria that reflect what is common and what is different between the compared biological samples.

2. OBJECTIVES

The primary goal of this study was the development and the evaluation of analytical strategies for the analysis and the identification of hydrophobic proteins by mass spectrometry. Analysis of membrane proteins in a complex mixture was always an obstacle for the proteomics field. Their hydrophobic nature requests a detergent for their solubilization and makes the use of traditional proteomics techniques less compatible (e.g. two-dimensional gel electrophoresis). Their low abundance in complex biological mixtures (in comparison to soluble proteins) requires additional purifications steps for their enrichment. Even at equal abundance, the identification rate of membrane proteins is lower than that of soluble proteins. This study aimed at better understanding constrains linked with the analysis of membrane proteins and at investigating possible improvements in this process.

The Brush Border Membrane (BBM) from mice small intestine was chosen as an ideal source of membrane proteins. The investigation of proteins that are located in the BBM membrane and may play a key role in cholesterol absorption has been of increasing interest in the recent years for many pharmaceutical companies because of the correlation of cholesterol absorption malfunction with cardiovascular diseases. This proteomic approach should confirm the findings of several studies that were based on gene analysis, activity tests and immunology assays, namely that the BBM contain proteins that participate in cholesterol regulation in the enterocytes.

A third goal of this study was to investigate conditions for which a quantitative mass spectrometric experiment using a label-free strategy could be used to investigate biological samples, as technical variability and sample stability of the BBM preparation were not known. The reproducibility of the technical steps has to be monitored for the whole workflow, from the BBM preparation until the mass spectrometric analysis of the samples, so to pinpoint difficulties and limitation of the workflow and preparing the way for a label free quantification strategy using only the information from the LC-MS data.

An additional aim of this study was to set the grounds for the analysis of gene expression along the small intestine of a mouse, in particular with the goal to compare gene expression of differentially treated mice or of mice with different genetic background.

3. MATERIALS AND METHODS

3.1 Materials / Chemicals

Materials / Chemicals	Company
ABC	Fluka
Acetic acid	Merk
Acetonitrile	Merk
a-CYANO	Bruker
Agarose	Sigma
Amastatin	Bachem AG
Angiotensin 1-5	Bachem AG
BCA-kit	Pierce
Bestatin	Bachem AG
beta Mercaptoethanol	Fluka
C57B/6J-Bom male mice	Taconic
Calcium chloride	Fluka
CHAPS	Sigma
Complete, inhibitor cocktail tablets	Roche
DTT	Merck
ECL-kit	GE healthcare
EDTA	Fluka
EtBr	Sigma
FA	Pierce
Hyperfilm-ECL	Amersham
Met-Bradykinin	Bachem AG
Methanol	Merk
MOPS buffer	Invitrogen
PMSF	Sigma
Ponceau S 1%	Sigma
PVDF	Perkin Elmer
PVDF membrane	Millipore
SDS	Fluka
Sodium chloride	Fluka
Superblock blocking buffer	Pierce
TFA	Pierce
Transfer buffer	Invitrogen
Tris	Sigma
Tween-20	Merk
Whatmann papper	VWR

3.2 Methods

3.2.1 BBM Preparation

Brush Border Membrane (BBM) vesicles were isolated by using a modified protocol from Kessler et al. using calcium chloride precipitation (79). The small intestine (rinsed with PBS) or the scrapped mucosa thereof from wild type C57B/6J-Bom male mice (6-8 weeks old) was suspended in Homogenization buffer (50 mM Mannitol, 2 mM Tris, pH 7.1) in presence of protease inhibitor cocktail tablets (Complete , Roche) and 1 mM PMSF (from a 100 mM stock solution in ethanol). Each gram of tissue was suspended in 15 ml homogenization buffer. The suspension was homogenized four times 20 seconds using a Polytron (Kinemetica GmbH) at maximum speed with intervals of 40 seconds in ice. The homogenized tissue was then centrifuged at 3,000 x g for 15 min at 4 $^{\circ}$C and the cell debris was discarded. The supernatant was re-centrifuged at 27,000 x g for 30 min at 4 $^{\circ}$C. The resulting pellet was weighted and re-suspended in Homogenization buffer (at the ratio of 1 gram of pellet in 15 ml Homogenization buffer) and $CaCl_2$ was added to a final concentration of 10 mM (from a stock solution of 1M $CaCl_2$ freshly prepared). After incubation on ice for 20 min, the 3,000 x g centrifugation step was repeated. The supernatant was then re-centrifuged at 27,000 x g for 30 min at 4 $^{\circ}$C. The resulting pellet, which contained the BBM vesicles, was first washed twice with 1 M KCl, then twice with 0.1 M Na_2CO_3, with 30 min centrifugation at 50,000 g at 4 $^{\circ}$C after each washing step. The purified BBM pellet was finally solubilised in storage buffer (100 mM Mannitol, 1 mM Heppes-Tris pH 7.5, 1% SDS and 1% CHAPS) in presence of protease inhibitor cocktail tablets Complete and 1 mM PMSF and stored at -80 $^{\circ}$C.

The targeted inhibition of the His-Zn-dependent metalloproteases was achieved by adding 1 mM Amastatin (80, 81), 1 mM Angiotensin 1-5, and 1 mM Met-Bradykinin in all buffers and washing solutions. The excess of Ca^{+2} after the calcium precipitation step was removed with 40 mM EDTA at the first wash of the BBM fraction with KCl. (§ 4.1.2.3 figure 4.9: flow diagram of BBM preparation).

3.2.2 Protein concentration estimation by the BCA method

The BCA Protein assay is a detergent-compatible assay for the colorimetric detection and quantitation of total protein using bicinchroninic acid (BCA) as a reporting agent. The method combines the well known reduction of Cu^{+2} to Cu^{+1} by protein in an alkaline medium with the highly sensitive and selective colorimetric detection of Cu^{+1} using a reagent containing bicinchrominic acid. The purple-colored reaction product of this assay is formed by the chelation of two molecules of BCA by one cuprous ion. The complex exhibits a strong

absorbance at 562 nm that is nearly linear over a broad working range (20 - 2,000 μg/ml). The macromolecular structure of the protein, the number of peptide bonds and the amino acids cysteine, cystine, tryptophan and tyrosine are responsible for the color formation with BCA. Protein concentrations are determined and reported with reference to standards of a common protein such as BSA.

In this work, a dilution series of known concentrations of BSA was prepared and assayed alongside the unknown samples before the concentration of each unknown was determined based on the standard curve. It was very important that the buffer used for the dilution series of the standards was identical with the buffer of the samples to be assayed. 25 μl of each standard or of unknown sample replicate was pipetted in a microplate well. 200 μl of working reagent (50 volumes of reagent A: 1% BCA-Na$_2$, 2 % Na$_2$CO$_3$·H$_2$O, 0.16% Na$_2$-tartrat, 0.4% NaOH, 0.95% NaHCO$_3$, pH 11.25, mixed with 1 volume of reagent B: 4% CuSO$_4$·5H$_2$O) was added in each well. The plate was mixed thoroughly on a plate shaker for 30 sec before being incubated for 30 min at 37 $^{\circ}$C. The plate was allowed to cool down to room temperature after which the absorbance at 562 nm was measured on a plate reader.

3.2.3 Protein Deglycosylation

There are two methods for protein deglycosylation: the chemical and the enzymatic procedure. Hydrazinolysis of glycoproteins (82), although capable of removing both N- and O-linked sugars, results in the complete hydrolysis of the protein and the modification of many amino acids. A milder chemical method, such as the trifluoromethanesulphonic acid method (TFMS) (83, 84), leads to incomplete sugar removal, partial protein hydrolysis and modifications of several amino acids. On the other hand, the enzymatic deglycosylation provides complete removal of the N-linked sugars from the protein with no measurable side-reaction. In this study, protein deglycosylation was performed enzymatically under denaturing conditions (heating with SDS and βME) to increase the yield of deglycosylation. The removal of the N-linked oligosaccharides was performed using N-Glycanase (PNGase F, Prozyme). The only amino acid modification that occurs in this reaction is the deamidation of the arsparagine residue to aspartic acid after the removal of the sugar from this residue. There is no enzyme comparable to N-Glycanase to remove intact O-linker sugars. Monosaccharides must be removed by a serie of exoglycosidases until only the Galβ (1-3)GalNAc core remains attached to the serine or threonine residue. O-Glycanase can then remove the core structure intact with no modification of the serine or threonine residues. Sialic acids attached to O-linked sugars can be removed by Sialidase A.

Samples were deglycosylated under denaturing conditions according to the protocol provided by the kit's manufacturer (Prozyme). 10 µl of 5x incubation buffer (0.25 M sodium phosphate, pH 7.0) and 2.5 µl of denaturation solution (2 % SDS, 1M βME) were added to a maximum of 100 µg of sample (dissolved in 30 µl). The sample was heated for 5 min at 100 $^{\circ}$C and, once cooled down to room temperature, 2.5 µl of detergent solution (15% NP-40) were added. Finally, 1 µl (approximately 5 mU) of each deglycosylation enzyme (N-Glycanase, Sialidase A, and O-Glycanase) was added to the solution and the sample was incubated for 3 h at 37 $^{\circ}$C. The protein deglycosylation process was monitored by 1D SDS PAGE.

3.2.4 1D SDS-PAGE electrophoresis

Sodium dodecylsulphate polyacrylamide gel electrophoresis (SDS-PAGE) is a method to resolve proteins according to their molecular size. The negatively-charged SDS molecules in the sample buffer bind to the heat-denatured proteins in a constant ratio (approx. 1.4 g SDS per g protein). As a result, SDS-loaded proteins migrate towards the positive pole in an electrical gradient, impeded by the polymerized and cross-linked polyacrylamide, at a mobility that is roughly proportional to their size.

The following protocols were used for Novex Bis-Tris precast SDS gels (Invitrogen).

3.2.4.1 Sample preparation and electrophoresis

Each sample was reduced and alkylated before being applied onto the SDS gel. Samples (approximately 15-20 µg total protein) were mixed with the appropriate amount of 4x NuPAGE LDS sample buffer (1x NuPAGE LDS buffer contains 141mM Tris, 2 % LDS, 10 % Glycerol, 0.51 mM EDTA 0.22 mM Coomasie Blue G250 and 0.175 mM Phenol Red pH: 8.5 from Invitrogen). 100 mM freshly prepared DTT (in water) was added and the samples were incubated for 15 min at 70 $^{\circ}$C. The samples were then allowed to cool down to room temperature and 300 mM Iodacetimide (aliquots of 1 M iodoacetamide in 300 mM Tris pH 8.8, 2.5% SDS, 10% Glycerol) was added. Each sample was incubated for 25 min at 45 $^{\circ}$C before being loaded onto a 10 % acrylamide Novex Bis-Tris SDS gel (Invitrogen) that had already been placed in the Mini cell module (Invitrogen). 10 µl of the prestained SDS molecular mass marker (See blue plus 2 from Invitrogen) was loaded in one well of the gel to calibrate the molecular mass of the separated proteins in each sample. The gel was run using the MOPS SDS running buffer system (20x MOPS SDS running buffer contains 1 M MOPS, 1 M Tris, 69.3 mM SDS and 20.5 mM EDTA, pH: 7.7 from Invitrogen) at 200 volts constant for approximately 1 hour typically until the front dye had reached the bottom of the gel.

3.2.4.2 Protein staining

Protein staining using colloidal Coomassie is based on the work of Neuhoff et al., (85). The method is based on the colloidal properties of Coomassie Blue dyes in aqueous or methanolic solutions in presence of inorganic acids and high salt conditions. The free dye in solution is greatly reduced due to the hydrophobic effect, resulting in low background and high affinity binding of the dye to the proteins fixed in the gel. A typical limit of detection of < 10 ng of BSA protein loaded onto a 1 mm 4-12 % acrylamide Tris Glycine gel can be obtained using this protein staining method.

After completion of the run, the gel was removed from the mini cell. The gel was first incubated in freshly prepared fixing solution (50% methanol, 10% acetic acid in deionized water) for 10 min to obtain a more efficient staining. The gel was then shaken for a maximum of 12 hours in the staining solution (20% Methanol, 20% Stainer A (containing ammonium sulfate and phosphoric acid) and 5% Stainer B (containing Coomassie G250) in deionized water), after which the gel was then destained in deionized water until a clear background was obtained.

3.2.5 Western Blotting

A Western blot (alternately, immunoblot) refers to a method to immunodetect a specific protein, usually in a complex mixture, on a membrane support after electrophoretic separation. This process is performed in three steps. First, gel electrophoresis is used to separate native or denatured proteins by the length of the polypeptide (denaturing conditions) or by the 3-D structure of the protein (native/ non-denaturing conditions). Proteins are then transferred from the gel to a membrane support (i.e., nitrocellulose or PVDF) via electrotransfer. Finally the blot is processed for the detection of specific proteins using antibodies, the product of which is the final Western immunoblot.

In this protocol, protein transfer from SDS gel to PVDF membrane was achieved using the Wet blotting technique. A PVDF membrane (cut to the size of the gel) was activated (immersed) in methanol for 30 seconds, briefly rinsed with water, and then equilibrated in the Transfer buffer (NuPAGE Transfer buffer 1x: 25 mM Bicine, 25 mM Bis-Tris and 1 mM EDTA, pH: 7.2 complemented with 20 % methanol and 0.005% SDS) along with the two filter papers. Samples of interest were loaded and separated onto a 10% SDS Bis-Tris gel (Invitrogen) as described in the previous chapter. The gel was soaked in deionised water immediately after electrophoresis (for approximately 10 min) before to rinse it shortly in Transfer buffer before use. The XCell II SureLock Mini-Cell module (Invitrogen) was used

for the transfer of the proteins from the gel to the PVDF membrane. The sandwich of the PVDF membrane and the gel was placed in the XCell module according to the manufacturer's instructions. The transfer took place at 38 volts constant for 90 minutes with a starting current of 100 mA with the module kept cold in an ice bath.

After transfer, the PVDF membrane was rinsed briefly with deionised water and was stained with 0.1% Ponceau S for evaluating the efficiency of the protein transfer. The membrane was then destained with deionised water and was washed several times with TTBS buffer (50 mM Tris pH 8.0, 80 mM NaCl, 2 mM $CaCl_2$, 0.1% Tween 20) until the stain was completely removed. Blocking was achieved by incubating the membrane in Superblock TBS blocking buffer (Pierce) and then in Blocking buffer (TTBS buffer plus 5% non-fat dried milk) for 30 min each. The membrane was then incubated overnight at 4 $°$C in blocking buffer with the primary antibody at an appropriate dilution. On the following day, the PVDF membrane was washed three times with TTBS buffer for 10 min each and 1 h in the blocking buffer. The membrane was then incubated for 90 min with the secondary antibody (HRP-conjugated) in the blocking buffer at room temperature. The membrane was washed subsequently three times with TTBS buffer for 10 minutes each. Immunodetection by chemiluminescence was performed by incubating the PVDF membrane in the ECL detection reagent (GE healthcare) for 1 min. Excess reagent was drained out and the membrane was wrapped up in a saran wrap. The blots were placed protein side-up in an X-ray film cassette and a sheet of chemiluminescence film (Hyperfilm ECL from Amersham Biosciences) was placed on the top of the membrane. The cassette was closed and the film was exposed for a given amount of time (typically, from 30 sec to 10 min) before being developed. This last step was repeated several times until an exposure time was considered as optimal from the signal point-of-view.

3.2.6 In-gel protein digestion

This protocol describes a method for the direct digestion of proteins in Coomassie- or colloidal blue-stained polyacrylamide gels. The absence of detergents and a minimal use of salt allow the direct analysis of the resulting digest by nanoLC-tandem mass spectrometry without an additional desalting step. In this study the protein digestion was performed by Trypsin, an enzyme that cleaves specifically after Lysine and Arginine except if the following amino acid is Proline.

The colloidal blue-stained gel was placed on a clean glass plate and the gel bands of interest were excised out of the gel using a clean scalpel. Each band was cut into the smallest pieces possible that were placed in an Eppendorf tube. The bands were destained first with 50%

ACN for 10 min and second with destaining solution (100 mM ABC, 30% ACN) for 10 min. The last step was repeated if necessary until the acrylamide pieces were completely colorless. The protein reduction/alkylation step was skipped as the sample was already reduced and alkylated prior to SDS PAGE. The bands were then dehydrated with 100% ACN for 15 min after which the bands were dried in a Speed Vac for 10 min without heating. The dried gel pieces were re-swelled in 50 µl tryptic digestion buffer (40 mM ABC containing 10 ng/µl Trypsin (Promega)) for 30 min. Additional buffer (5 mM ABC) was then added to the gel pieces if they were not completely covered with buffer. Digestion was carried out overnight at room temperature. On the following day, the supernatant was first collected and additional 5 mM ABC (50 µl) was added to the gel pieces for 15 min at 37 °C. After collection of the supernatant, the gel pieces were further extracted with 300 µl 100% ACN for 10 min at room temperature, after which the supernatant was again collected. The gel pieces were then extracted with 50 µl of 5% Acetic acid, 5% ACN at 37 °C for 15 min and the supernatant was collected. Finally, the gel pieces were extracted again with 300 µl 100% ACN at room temperature for 15 min. The above procedure has been adapted from several papers (86-88). The pooled supernatant was then dried in a speed Vac without heating and the peptide extract was kept at -80 °C until further use.

3.2.7 Mass spectrometry

3.2.7.1 Packing of NanoLC columns

An analytical fused silica emitter (75/360 µm i.d./o.d., tip diameter 8±1 µm, New Objectives) was packed with 12-15 cm of 3 µm Reprosil C_{18} A.Q. reverse phase material (Dr. Maisch). The resin was mixed with methanol (approximately 50 mg/ml) and the fused silica was packed at 170-200 bar using pressurized Nitrogen. The column was then rinsed with pure methanol at the same pressure and kept dried until use.

3.2.7.2 Method development for NanoLC ESI-MS/MS

Peptide samples were analyzed by nano liquid chromatography electrospray tandem mass spectrometry using an Ultimate 3000 nanoflow chromatographic system (Dionex) coupled to a LTQ-Orbitrap tandem mass spectrometer (Thermo Electron) equipped with a nanoelectrospray ion source (Proxeon Biosystems)

The peptide mixture of each sample was dissolved in 20 µl of buffer A (2% ACN, 0.5% acetic acid). 10 µl of the sample was transferred to a glass vial of 0.25 ml volume (sun – sri) of which 5 µl were injected into the system at a flow rate of 450 nl/min at 100 % buffer A for 12

min. After loading, the flow was decreased to 250 nl/min and peptides were eluted from the reverse phase column as follows: 12—14 min, 0—5% buffer B (80% ACN, 0.5% acetic acid); 14—30 min, 5—30% buffer B using the curve 4 (slightly concave) of the Chromoleon software (Dionex); 30—90 min, 30—55% Buffer B using the curve 6 (slightly convex) of the Chromeleon software. The column was then washed for 15 min with 100 % buffer B at 350 nl/min, after which it was re-equilibrated for 25 min in 100% buffer A at 350 nl/min.

Peptides were analyzed by tandem mass spectrometry using standard operating parameters as follows: the electrospray voltage was set to 2.2 kV and the ion transfer capillary temperature was at 170 $^\circ$C. Survey scans (scanning range m/z 400-1500) were recorded in the Orbitrap mass analyzer at a resolution of 30,000 with the lock mass option (20) enabled. Data-dependent MS/MS spectra of the five most abundant ions from the survey scan were recorded in the LTQ ion trap using a normalized collision energy of 32% for MS/MS (30 ms activation, q=0.25) and a selection threshold of 500. Target ions selected for MS/MS were dynamically excluded for 30 sec.

3.2.7.3 Data Processing Method and Protein Identification

The raw data files of each LC-MS/MS run were processed using the SEQUEST search algorithm 27 (SEQUEST version 27.0, revision 12, Thermo Electron). Searches were performed against the in house-generated MouseGP database (Version January 2007, genome version mm8/NCBI36, 60862 entries). The MOUSEGP database is generated by taking the most recent version of all the non-redundant sets of Swissprot and Trembl sequences from human, mouse, rat, other vertebrates, drosophila, *C. elegans* and Yeast and blast the sequences against the mouse chromosomes. The putative exons are then assembled into genes. The alignments between genomic chromosome sequence and proteins are refined with GeneWise (89, 90), a tool that finds splice sites and corrects frame shifts. Data were searched with a mass tolerance of +/-5 ppm for parent ions and +/-1.0 Da for fragment ions. Methionines (reduced/oxidized; +15.9949 Da) were considered as differential modifications while cysteines were considered as fully carbamidomethylated (+57.0199 Da). Only fully tryptic peptides with no more than one miscleavage were considered for data analysis. Peptides were considered as unambiguously identified if their XCorr scores (91) exceeded 1.5, 2.0 or 2.5 for singly, doubly and triply charged ions, respectively, and if the corresponding ΔCn scores (the normalized Sequest XCorr score difference between the first and the second best peptide match,(91)) were larger than 0.2. In this study, only proteins for which at least two different peptides were identified were considered as successfully identified. The False

Positive Discovery Rate at the peptide level was estimated by searching the raw data files against a shifted database. This database, which is generated from MOUSEGP, has an identical number of proteins and tryptic peptides by keeping fixed the position of the arginines and lysines while the position of all remaining amino acids were left-shifted by two positions. The raw files were searched by using the above searching criteria and the raw spectra were submitted to the scrambled database for scoring. These spectra were binned based on their score and the number of spectra in each bin was counted and stored in the database. An objective estimate of the false positive discovery rate at the peptide level is obtained by comparing the distribution and the counts of spectra for a given score for spectra that were scored against MOUSEGP and the scrambled database (52, 92, 93). Using the criteria mentioned above, the false positive discovery rate for 1+, 2+ and 3+ charged peptides was estimated to be 7.4%, 3.2%, and 0.6%, respectively.

Comparison of the LC-MS raw data files based on the total ion current of the peptides was performed by using the Genedata RefinerMS (version 4.5) software suite. Software-specific settings were as follows: baseline subtraction (20% quantile value, m/z Window 10 Da, RT window 0.1 min); chromatogram retention time (RT) alignment (prior internal peak identification on m/z window with 5 points, RT window 0.5 min, gap penalty 0.75, RT search interval 100 scans); peak identification as in RT alignment; peak shaping (multiplicity: 67%); isotopic clustering (minimal charge 2, maximal charge 4, correlation threshold 0.5, maximal missing peaks 1, mass tolerance 0.05, ionization: protonation; mass consistency: peptides). The resulting peak table associates the m/z and RT of a feature with its extracted ion count (XIC) and signal-to-noise ratio (S/N) measured in each sample. Features with a median S/N<5.0 were excluded from the data analysis as these outliers displayed broad ranges of nonlinear chromatographic shifts across samples**.**

3.2.7.4 Sequence and topology analysis

The proteins' cellular location and topology were analyzed with a variety of tools as follows.

Prediction for signal peptide was performed using the in-house "signal_anchor" software tool (http://bioinfo.bas.roche.com:8080/sawicgi/sawi.cgi?signal_anchor) using an algorithm inspired by the web-based SignalPep tool (94) but using a support vector machine prediction model using Roche internal data rather than a neural net prediction model.

Membrane protein prediction was performed using either the "ALOM" software tool (http://bioinfo.bas.roche.com:8080/sawicgi/sawi.cgi?alom) as described by Klein P et al. (95),

or the web-based "TMHMM 2.0" (http://www.cbs.dtu.dk/services/TMHMM-2.0/), as described by Krogh A. et al. (96).

The version of the GO annotations (as described in Gene Ontology Consortium (2000)) used in this work was from June 2008. In this work, the multiple GO annotations of a protein were reduced to what was believed to be the most relevant entry according to the following rules:

a) plasma membrane > ER, GOLGI, endosome > lysosme, microsome, peroxisome> mitochondria > cytoplasm, cytosol > nucleus > other

b) integral > anchored > peripheral, associated to > no mention of membrane interaction

c) any membrane interaction > proteasome, cytoskeleton, vesicle, etc. >extracellular

In doubt, the protein annotation of the SwissProt database was used to direct the selection to one or the other direction.

3.2.8 RNA extraction from the small Intestine

A typical mouse intestinal tissue (100 mg) contains about 150 µg total RNA. However, only 1.0–1.5% of it is accounted as mRNA. RNA molecules, compared to DNA, are relatively unstable. In order to ensure accurate gene expression analysis, it is important that the RNA analyzed truly represents the *in vivo* gene expression of the sample. This is complicated by the fact that changes can occur during handling of the sample and isolation of the RNA. Two major types of artifacts are known to occur. Gene down regulation and enzymatic degradation of RNA, due to ribonucleases (RNases), result in an artificial reduction of both nonspecific and specific mRNA species. At the same time, other genes can be induced during handling and processing of the sample. The combination of these two effects can result in a transcription profile that differs from the true *in vivo* gene expression pattern.

Efficient disruption and homogenization of the tissues under conditions that preserve the RNA stability is essential for the total RNA isolation procedure. In the current protocol, the freshly removed small intestine was flushed with PBS buffer (GIBCO) and RNA*late*r (QIAGEN) to better preserve the RNA from degradation. Approximately 100 mg of tissue was then embedded in 1 ml TRI Reagent (Sigma). As reported by Chomczynski et al. (97) the phenol and guanidine thiocyanate contained in the TriReagent are effective inhibitors of RNases. The tissue was homogenized with Lysing Fast RNA Matrix D tubes (Bio 101 systems) using a FastPrep Cell disrupter (Thermo Savant) at a speed of 6 m/sec for 20 sec. The cell disruption process is due to the collision of matrix and sample within the FastPrep® Instrument sample tube. The rate of collision and energy of impact (both of which determine the effectiveness of the disruption process) are a function of the FastPrep® Instrument speed

settings and specific gravity of the bead material used. The homogenate was then separated into an aqueous and organic phase by adding 0.1 volume of chlorophorm. Alternatively the step of the homogenization of the tissue was replaced by grounding the tissue under liquid Nitrogen. In this case the TRI reagent was added to the grounded tissue and immediately after was following the step of the phase separation by addition of chlorophorm. In both cases the preparation were incubated on ice for 5 min and centrifuged at 12,000 x g for 15 min at 4 $^{\circ}$C. The RNA was present in the aqueous phase, the DNA in the inter-phase, and the proteins in the organic phase. Next, the RNA was precipitated from the aqueous phase by adding 2 volumes of cold isopropanol. The preparation was incubated on ice for 15 min and centrifuged at 12,000 x g for 10 min at 4 $^{\circ}$C. The precipitate was washed with 75% ethanol and centrifuged at 7,000 x g for 5 min at 4 $^{\circ}$C. The extracted RNA was finally dissolved in RNase-free water and heated for 10 min at 65 $^{\circ}$C. RNA aliquots were then stored at -80 $^{\circ}$C until use.

3.2.9 RNA Electrophoresis

The overall quality of an RNA preparation may be assessed by electrophoresis using a denaturing agarose gel; this type of analysis provides simultaneously some information about RNA yield. A denaturing gel system is suggested, because most RNAs form extensive secondary structure via intramolecular base pairing, preventing them from migrating strictly according to their size. The quality of the extracted RNA was assessed using 1% agarose denaturating gel. 1 g of agarose was added to 72 ml of triple distilled water and heated until it was dissolved. The mixture was cooled to 60 $^{\circ}$C and 10 ml of 10x MOPS running buffer (10x MOPS buffer: 0.4 M MOPS, 0.1 M sodium acetate, 0.01 M EDTA, pH 7.0 adjusted with NaOH), 18 ml 37% formaldehyde (12.3 M) and 1 µl EtBr were added to the agarose solution. The gel was poured using a comb to form the wells for the samples. The gel was then assembled in a tank and 1x MOPS running buffer was added to cover the gel by a few millimeters.

Each RNA sample (1–3 µg) was mixed with 2 volumes of formaldehyde Load Dye (Ambion) and heated for 15 minutes at 70 $^{\circ}$C. The gel was loaded and electrophoresis was performed at 5 V/cm until bromophenol blue (the faster migrating dye) had migrated as far as 2/3 of the length of the gel.

The gel was visualized on a UV transilluminator (EtBr fluoresces under UV light when intercalated into DNA or RNA). Intact total RNA will show sharp 28S and 18S rRNA bands (eukaryotic samples) on a denaturing gel. In an ideal case, the 28S rRNA band should be

approximately twice as intense as the 18S rRNA band, a good indication that the RNA is intact. Partially degraded RNA will have a smeared appearance, will lack the sharp rRNA bands, or will not exhibit a 2:1 ratio. Completely degraded RNA will appear as a very low molecular weight smear. For this effect, a RNA size marker was loaded on the gel to allow the size determination of any bands or smears and to serve also as a positive control to ensure the gel was run properly. Alternatively, RNA quality was assessed using the Experion Systen (BIO-RAD) an automated Electrophoresis station. The analysis of the samples was performed according to the manufacturer's instructions.

4. RESULTS AND DISCUSSION

4.1 An improved protocol for the specific isolation of BBM from small intenstine

The first attempts to isolate BBM from the small intestine were performed following the protocol of Kessler et al. (79) developed initially for the kinetic measurements of sugar and mineral transporters in BBM. A piece of small intestine, or scrapped mucosa if a cleaner preparation was needed, was homogenized using a polytron and the cell debris was removed by a low-spin centrifugation step. The following step, a calcium chloride precipitation, was used to segregate the Brush Border Membrane (the apical side of the enterocyte's plasma membrane) from the Basolateral membrane (the lumen side of the enterocyte's plasma membrane). Due to the different polarity between the two sides of the membrane, the addition of $CaCl_2$ specifically precipitated the basolateral membrane, which could then be removed from the preparation by low-spin centrifugation, while the BBM membrane remained in solution (the exact mechanism by which the process is driven is not known). The enriched BBM fraction could then be collected by high-spin centrifugation and resuspended in an appropriate buffer for further analysis.

The degree of BBM enrichment from the small intestine was monitored by Western blot using a specific BBM marker (FATP-4) and a contaminant basolateral marker (Na+/K+ ATPase a1). The BBM fraction showed only a modest enrichment factor (approximately 10-fold compared to the crude lysate) and still contained a detectable contamination from the basolateral membrane (Fig. 4.4 and 4.5, see also below). Simultaneously, BBM preparations from three different mucosa intestinal sections were prepared to investigate the level of purification (at the protein level) obtained with the Kessler protocol. Survey of the literature indicated that the protein composition of the BBM changes along the intestine. Hence, a highly enriched preparation was expected to result in protein patterns that should differentiate each of the sections. Fig. 4.1 shows the 1D-SDS-PAGE analysis of the BBM preparations from proximal, central, and distal mucosa sections obtained from a pool of three mice. The obtained protein patterns were almost identical with each other, with no obvious difference allowing discrimination between the three sections, indicating that the BBM preparations contained a significant level of contaminants masking the BBM-specific proteins.
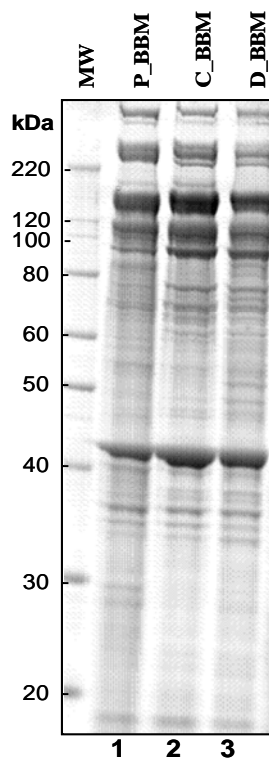
Figure 4.1: *10% 1D-NuPAGE Bis-Tris gel analysis of the Kessler BBM preparation from the three mucosa intestinal sections. 15 μg total protein was loaded per sample. P_BBM, C_BBM, D_BBM: proximal, central, and distal BBM preparation, respectively; MW: Molecular Weight marker.*

For that reason a significant amount of time was spent to improve the performance of the original protocol. At first, my efforts were concentrated on the precipitation step, using several $CaCl_2$ concentrations to obtain more specific precipitation conditions, and on the centrifugation conditions in an attempt to increase the purity of the BBM preparation. Results of this optimization process were mitigated at best and did not yield any noticeable change in the quality of the preparation (data not shown). As a next step, I chose to focus my attention on improving the isolation protocol by adding purification steps which could remove a large part of the cytosolic and membrane-associated contaminant proteins so that a higher enrichment of the BBM's hydrophobic proteins could be achieved. First, a low- and high-spin centrifugation cycle was performed immediately after tissue homogenization so that the bulk of the cytosolic proteins could be removed from the membrane fraction before the $CaCl_2$ precipitation. Further, two washing steps using high salt (1 M KCl) and high pH (100 mM Na_2CO_3) were added after collection of the BBM fraction to remove most of the membrane associated proteins. A summary of the original and the modified protocol is represented on the flow diagram shown in Fig. 4.2.

Intestinal tissue, mucosa scrapping

Step 1 — **Homogenization**
Homogenization Buffer:
       2mM Tris pH 7.1, 50 mM Mannitol

Homogenized tissue

Step 2 — **Centrifugation at 3000 x g** for 15 min, 4°C

Supernatant

Step 3 — **Centrifugation at 27.000 x g** for 30 min, 4°C

Pellet 1

Step 4 — **Pellet resuspension** in Homogenization Buffer
CaCl$_2$ Precipitation (10mM final concentration)
     20 min, on ice

Step 5 — **Centrifugation at 3000 x g** for 15 min, 4°C

Supernatant

Step 6 — **Centrifugation at 27.000 x g** for 30 min, 4°C

Pellet 2   BBM fraction

Steps 7-10 —
2 x washing with 1M KCl
2 x washing with 0.1M Na$_2$CO$_3$
Centrifugation at 50.000 x g for 30 min, 4°C
between each wash step

Washed BBM fraction in Storage buffer plus 1% SDS and 1% CHAPS

Figure 4.2: *Flow diagram of a BBM preparation from mucosa or from intestinal tissue. The original steps from the Kessler protocol are marked in black. The steps 2, 3 and 7-10, which were added to the original protocol, are highlighted in green.*

The improved BBM preparation protocol has been tested with intestinal tissue and mucosa scrapping, total tissue or in sections, with excellent results. The performance of the protocol and the level of BBM purification obtained from a total small intestine preparation are illustrated in Fig 4.3. Samples were collected along the BBM preparation and were analyzed by 1D-SDS-PAGE.
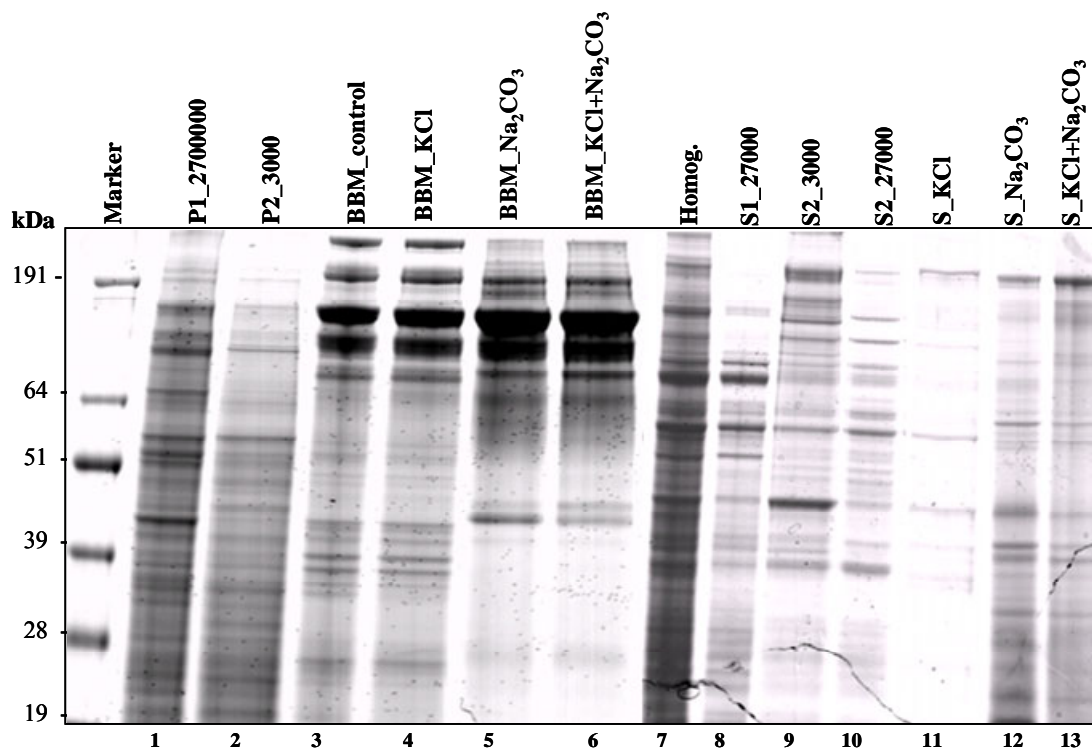
Figure 4.3: 1D-SDS-PAGE analysis of the fractions collected during the improved BBM isolation protocol. *A sample amount of 15 μg (according to the BCA protein assay) of total protein (except for the supernatants S1_27000 and S_KCl, where the maximum volume of sample was loaded) was loaded onto a 10 % 1D NuPAGE BisTris gel. Proteins were stained using colloidal Coomassie.*

It was clearly demonstrated, as shown in the 1D-SDS-PAGE analysis of the collected fraction, that the initial low/high spin centrifugation step removed a considerable number of contaminant proteins prior to the $CaCl_2$ precipitation (Fig. 4.3, lane 8). Similarly, the high salt/high pH wash steps (1 M KCl followed by 100 mM Na_2CO_3, steps 7-10 of the protocol) of the initial BBM fraction also removed a number of membrane-associated proteins (Fig. 4.3, lanes 11, 12). Interestingly, the two washes clearly targeted different sets of proteins, as shown in the 1D-SDS-PAGE analysis.

The purity and the enrichment of the isolated BBM from total intestinal tissue was assessed by Western Blotting using FATP-4, a transmembrane, BBM-specific protein (Fig. 4.4), and Na^+/K^+ ATPase a1, a transmembrane basolateral marker protein.
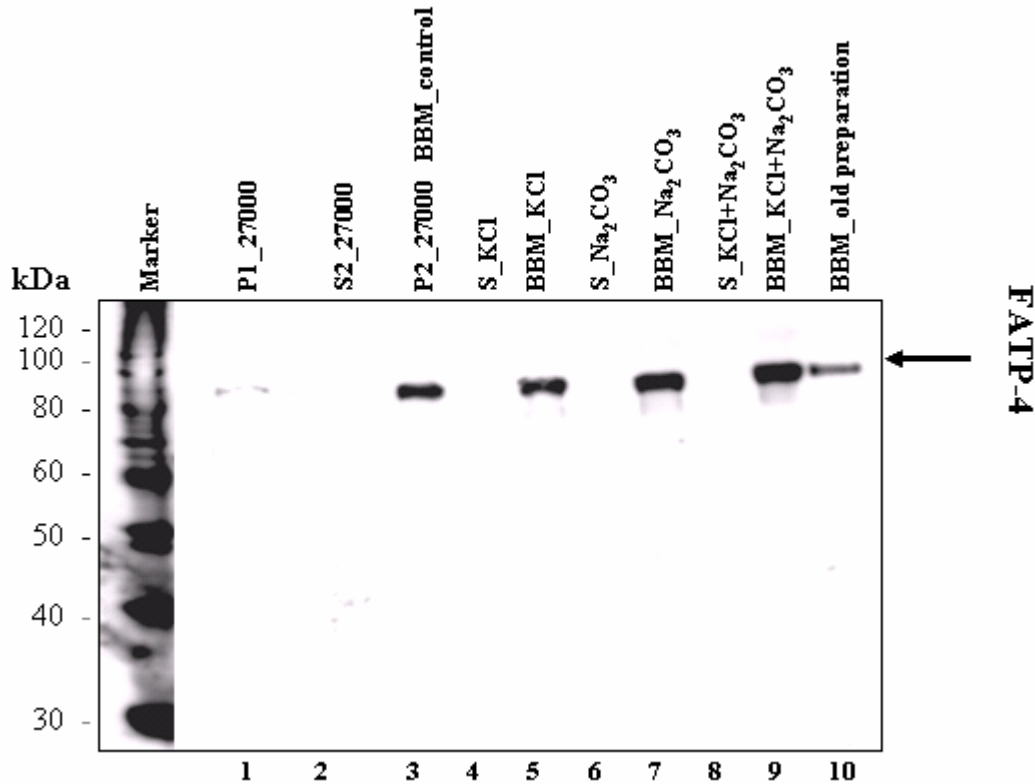
Figure 4.4: *FATP-4 Western Blot analysis of the fractions collected during the improved BBM isolation protocol. 10 μg total protein were loaded for each sample. 1st antibody 1:2000 goat polyclonal specific for FATP-4 (Cat. No: sc-5835 Santa Cruz Biotechnology), 2nd antibody 1:7500 Horseradis peroxidase goat anti rabbit IgG. Lane 9 corresponds to the BBM purified with the new improved BBM isolation protocol while lane 10 corresponds to the BBM purified according to the Kessler's protocol.*

The FATP-4 Western Blot analysis demonstrated convincingly that this BBM-specific protein was markedly more enriched in the BBM fraction collected with the improved protocol (approximately 30- to 50-fold compared to the crude lysate) than following the original Kessler's protocol. Comparison of the BBM fraction before the washes (lane 3) and the Kessler's BBM fraction (lane 10) illustrated that the first low/high spin centrifugation cycle was the key for improving the enrichment of the BBM fraction, most likely by removing the bulk of contaminant cytosolic proteins before the $CaCl_2$ precipitation step. The high salt/high pH wash steps also contributed to the enrichment of the FATP-4 BBM marker, although to a lesser extent, by the removal of additional cytosolic and membrane-associated proteins.

The quality of the improved BBM isolation protocol was also investigated with respect to the most likely BBM contaminant, the Basolateral membrane, using a Western Blot analysis against Na^+/K^+ ATPase a1, a specific marker of this compartment (Fig. 4.5).
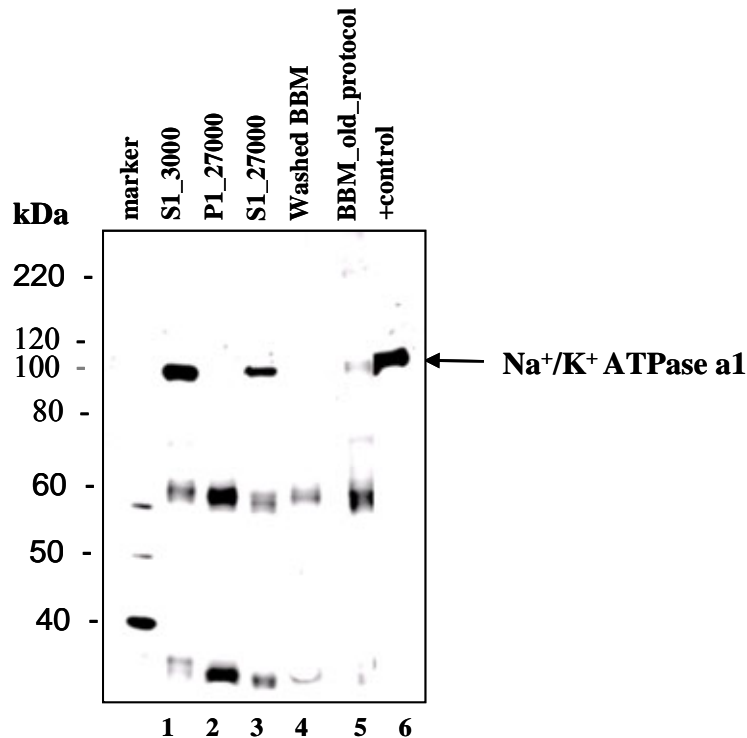
Figure 4.5: *Na⁺/K⁺ ATPase a1 Western Blot analysis of the fractions collected during the improved BBM isolation protocol. 10 μg total protein were loaded in each sample. 1ˢᵗ antibody 1:10000 mouse monoclonal specific for Na⁺/K⁺ ATPase a1 (Cat. No: 05-369, Upstate), 2nd antibody 1:7500 Horseradis peroxidase rabbit anti mouse IgG. Lane 4 corresponds to the BBM purified with the new improved BBM isolation protocol while lane 5 corresponds to the BBM purified according to the Kessler's protocol. Lane 6, positive control (rat liver extract)*

The Na⁺/K⁺ ATPase a1 western blot analysis of the BBM fractions confirmed the excellent quality of the improved BBM preparation protocol. The 100 kDa band was assigned to Na⁺/K⁺ ATPase a1 (according to the rat liver positive control) while the 60 kDa and 35 kDa signals most probably corresponded to antibody heavy and light chains (the secondary antibody was anti-mouse). There was no detectable basolateral membrane contamination (no signal for Na⁺/K⁺ ATPase a1 in lane 4) in the BBM fraction prepared according to the improved isolation protocol while the BBM fraction prepared according to Kessler's protocol showed a low level basolateral membrane contamination (weak signal of Na⁺/K⁺ ATPase a in lane 5).

In summary, an extended version of the original $CaCl_2$ precipitation protocol from Kessler was demonstrated to yield a highly enriched and purified BBM fraction according to the Western Blot analyses against FATP-4, a specific BBM protein marker, and against Na⁺/K⁺ ATPase a1, a known basolateral protein marker. In particular, the improved isolation protocol removed many cytosolic and membrane-associated proteins, thus facilitating a better enrichment of the hydrophobic BBM proteins, and yielded a BBM fraction with undetectable

level of basolateral membrane, the major contaminant of BBM in all previous isolation protocols.

4.1.1 1D-SDS-PAGE analysis of BBM fractions

Throughout this study, BBM have been isolated from mice total intestinal tissue, from sections of intestinal tissue, from total intestinal mucosa, and from sections of intestinal mucosa. A BBM preparation from intestinal mucosa, for example, has the advantage that the starting material is highly enriched in enterocytes (the source cell type for BBM). Hence, as contaminations from erythrocytes and from the intestinal epithelium are avoided, the purity of the BBM preparations automatically increases. Conversely, scrapping mucosa from the inside wall of the intestine is a laborious process that has to be performed immediately after sacrifice (see also following sections on protein degradation). In certain types of experiments, because of logistic reasons, only intestinal tissue might become available. In this section, I summarize some of the findings on BBM preparations from diverse source materials.

Fig. 4.6 shows a composite 1D-SDS-PAGE gel analysis of BBM preparations from total intestine, from total mucosa, and from mucosa sections that have been prepared following either the original Kessler protocol or the improved BBM isolation protocol (see previous section). The section A of the 1D-SDS-PAGE represents the BBM fractions isolated from the total small intestine. The protein pattern obtained from the BBM prepared using the improved protocol (section A, lane 2) shows clearly that a number of proteins have been removed compared to the BBM fraction prepared according to the Kessler's protocol (section A, lane 1). A significant increase in the number of proteins can be observed, especially in the lower mass range, when the BBM is prepared from mucosa tissue instead from the whole small intestine (see lanes 4 of sections B and C). However, the protein patterns of BBM prepared using mucosa sections following the improved isolation protocol (section C, lanes 2-4) show significant differences between the three different sections, in contrast to the almost identical protein patterns obtained from BBM isolated from mucosa sections that have been prepared according to Kessler's protocol (section B lanes 2-4).
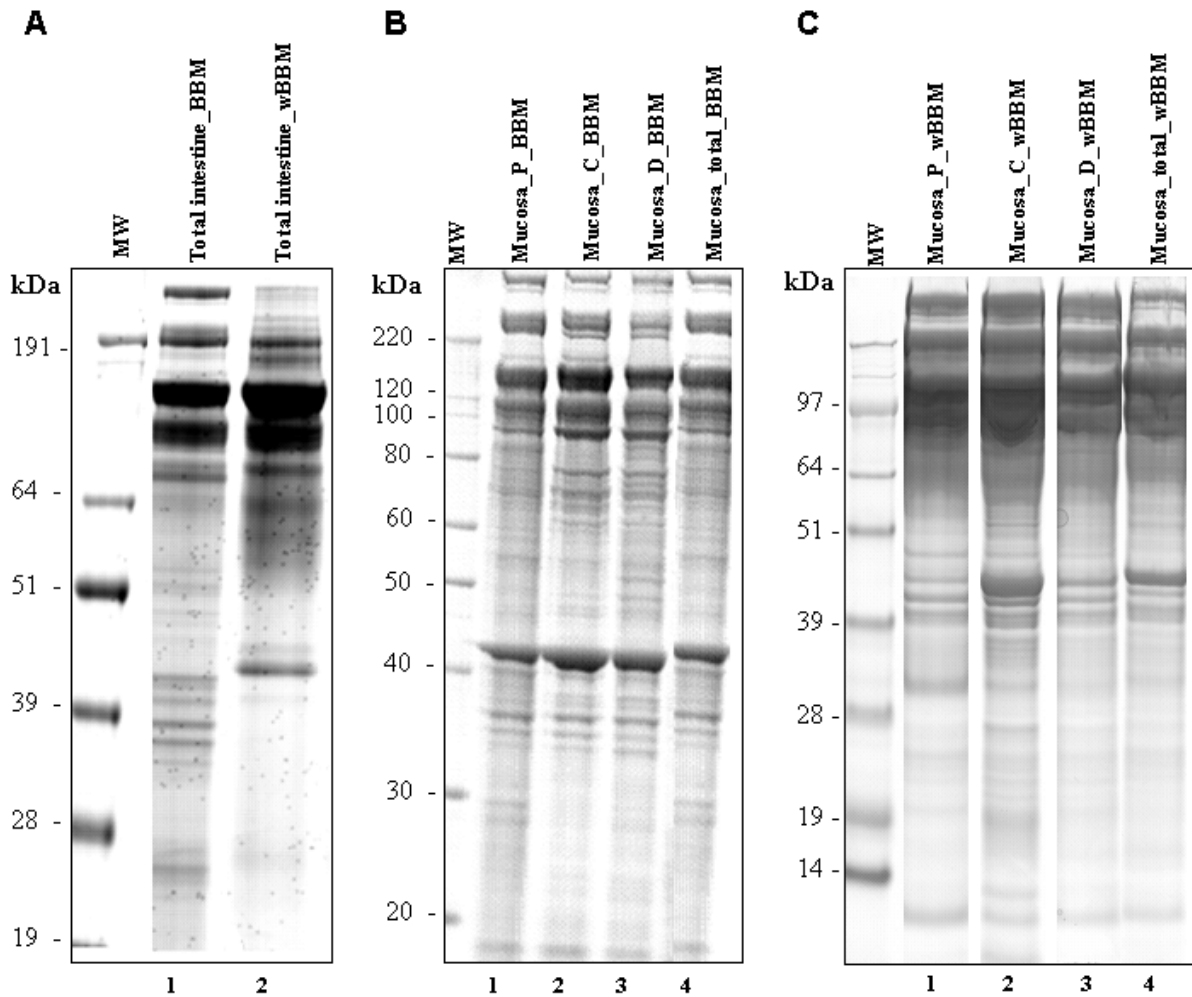
Figure 4.6: *1D SDS PAGE analysis of isolated BBM fractions. **Section** A: BBM fractions from total small intestine prepared either according to Kessler's protocol (lane 1) or using the improved preparation protocol (lane 2). 15 µg total protein were loaded per lane in a 10 % 1D NuPAGE Bis-Tris gel run in MOPS buffer system. **Section B:** BBM fractions from mucosa tissue in sections (lane 1-3) and total (lane 4), prepared according to Kessler's protocol. 15 µg total protein were loaded per lane in a 10 % 1D NuPAGE Bis-Tris gel run in MOPS buffer system. **Section C:** BBM fraction from mucosa tissue in sections (lane 1-3) and total (lane 4), prepared according to the improved protocol. 30 µg total proteins were loaded per lane in a 10 % 1D NuPAGE Bis-Tris gel run in MES buffer.*

In summary, the results of the 1D-SDS-PAGE analysis from BBM fractions prepared from different starting materials using two different protocols, in combination with the results obtained from the Western blots analysis (see section 4.1), were clearly indicative that the initial BBM characterization should be performed with mucosa as starting material following the improved isolation protocol to ensure maximum purity.

4.1.2 BBM preparation and protein degradation

4.1.2.1 Protein deglycosylation

A large number of proteins located in the BBM are highly glycosylated. Indeed, the 1D-SDS-PAGE analysis of a BBM fraction showed a large number of proteins tightly concentrated in the higher mass range of the gel. In order to better distribute those proteins in a wider mass range, and to achieve a better protein enzymatic digestion and peptide ionization in the mass spectrometer, an enzymatic deglycosylation process was performed under denaturing conditions (see section 3.2.3 for more experimental details). The BBM fraction was deglycosylated using N-Glycanase, Sialidase A, and O-Glycanase for 3 h at 37 $^{\circ}$C. The outcome of the deglycosylation process was monitored by 1D SDS PAGE electrophoresis (Fig. 4.7)
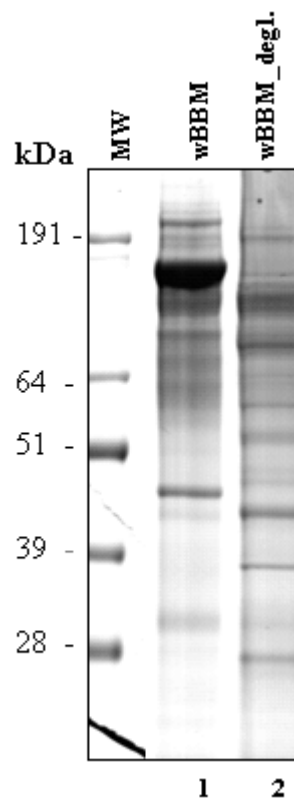
Figure 4.7: *1D-SDS-PAGE analysis of a BBM fraction before (lane 1) and after (lane 2) deglycosylation*. *15 µg total protein were loaded per lane in a 10 % 1D NuPAGE Bis-Tris gel run in MOPS buffer. MW molecular weight marker*

According to the 1D-SDS-PAGE analysis, the deglycosylation process was successful as the protein patterns of the native and deglycosylated BBM fractions were quite different from each other. In the deglycosylated sample, many protein bands have been shifted to lower molecular range and the proteins were distributed in a wider range of molecular masses.

4.1.2.2 Western Blot analysis of Aminopeptidase N

The quality and the extent of deglycosylation of the BBM protein fraction was examined in more detail through the Western blot analysis of Aminopeptidase N, an abundant single transmembrane helix BBM protein marker predicted to bear 13 N-linked glycosylation sites. The Western blot analysis of Aminopeptidase N before and after the deglycosylation procedure should substantiate the completeness of the process and confirm that the molecular mass of Aminopeptidase N has been reduced to its predicted molecular mass (approx. 110 kDa). For this procedure, a polyclonal antibody raised against the full amino acid sequence (69-966 amino acids) of mouse Aminopeptidase N was used. The results of the western blot analysis are represented in Fig. 4.8:
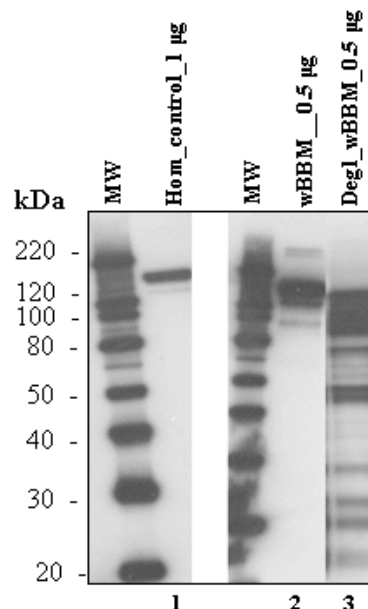
Figure 4.8: *Western blot analysis of Aminopeptidase N in a BBM fraction before and after deglycosylation. Lane 1: starting material (homogenized mucosa tissue); Lane 2: wBBM from total mucosa; Lane 3: deglycosylated BBM material from lane 2. Protein load is as indicated in the legend. 1ˢᵗ antibody 1:1000 goat polyclonal anti-Aminopeptidase N (Cat. No: AF2335, RnDSystems), 2nd antibody 1:7500 Horseradish peroxidase donkey anti goat IgG. MW, Molecular Weight marker*

The results of the Western Blot analysis (Fig. 4.8) confirmed that Aminopeptidase N was an abundant BBM protein generating a very strong signal from a single band at around 170 kDa using 1 µg of total protein load (Fig. 4.8, lane 1). The generated signal was even stronger when the western blot was performed using 0.5 µg total protein load from the isolated BBM membrane (Fig. 4.8, lane 2). In contrast, the Western blot results obtained from the deglycosylated BBM fraction were quite different (Fig. 4.8, lane 3): the signal representative for Aminopeptidase N was shifted completely to a lower molecular mass, confirming that the protein deglycosylation procedure was successful. At the same time, however, a number of

additional signals at lower molecular masses than expected were also detected in the deglycosylated BBM fraction. As the primary antibody was polyclonal and was raised against the whole Aminopeptidase N protein sequence, we came to the conclusion that these additional bands could only represent Aminopeptidase N fragments, indicating that degradation was occurring during the deglycosylation process.

Protein degradation is, by nature, one of the primary roles of the small intestine. It was believed, however, that the use of the commonly available Ser- and Cys-protease inhibitors cocktail during the BBM preparation combined with the storage of the BBM in high concentration of detergents (1% SDS and 1% CHAPS) should be sufficient to inhibit proteolytic degradation. Additionally, it was very surprising that these peptidases should remain active during the deglycosylation process as the procedure was performed under reducing, denaturing conditions and the sample was heated at 90 degrees before the addition of the deglycosidases.

The presence of proteolytic enzymes in the BBM was confirmed by systematic protein identification of the glycosylated BBM fraction, as it will be discussed in the section 4.2, showing that the most abundant BBM proteins were indeed peptidases. These results, combined with the observation that abundant proteins were always identified in numerous fractions, confirmed that protein degradation plays a central role in the BBM.

4.1.2.3 Inhibition of protein degradation

The role of the many peptidases that were identified in the BBM membrane is to degrade proteins and long peptides to smaller molecules that can be easily transported through the small intestine. Based on the identification data, a high percentage of the most abundant BBM proteins was accounted by peptidases belonging to the family of His-Zn-dependent metalloproteases, such as Aminopeptidase N, Aminopeptidase A, Ileal dipeptidyl peptidase (NAALADase), Neprilysin, or Meprin, (see section 4.2 table 4.3), which are not inhibited by common Serin, Cystein proteinase inhibitors. Chemically, these peptidases can be inhibited efficiently using trifluoromethanesulphonic acid (83, 98). However, this procedure must take place in anhydrous conditions and this was clearly not compatible with the isolation protocol.

The importance to inhibit these peptidases should not be underestimated: among others, proteolytic activity can lead to false results in a western blot data interpretation (epitope degradation) or to major difficulties when a label free protein quantitation is used for the comparison of the samples (signal dilution in several fractions, irreproducible isolation procedure). In order to reduce proteolytic activity, two peptidase partial inhibitors that were

described in the literature to be active against His-Zn metalloproteases, Amastatin and Bestatin (80), were evaluated. Using the same Western Blotting analysis as described in the previous section, Bestatin did not appear to have any effect while Amastatin was found to lessen but not suppress the proteolytic activity in the BBM fraction (results not shown).
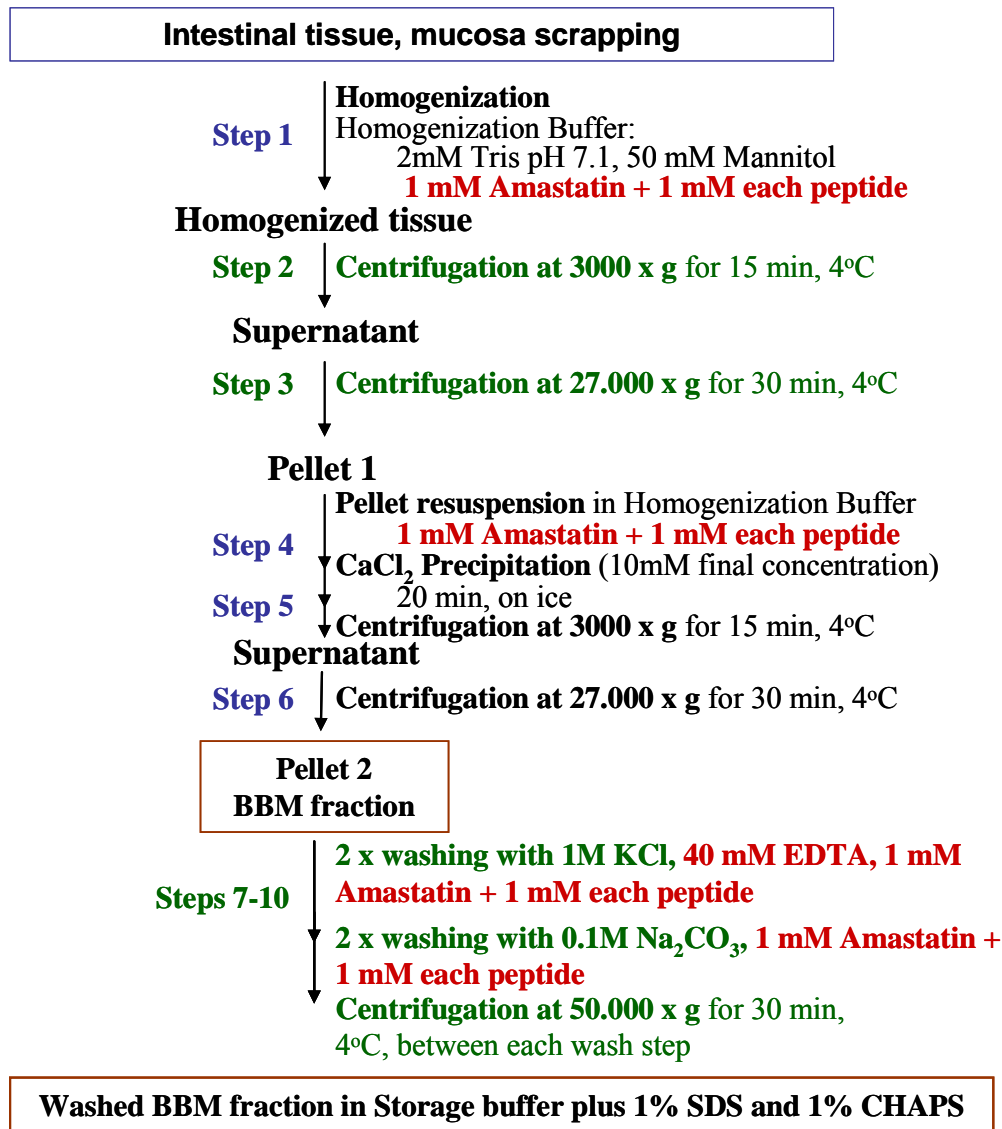
Figure 4.9: *Optimized flow diagram of a BBM preparation from mucosa or from intestinal tissue.* The original steps of the Kessler protocol are marked in black while the steps 2-3 and 7-10, labeled in green, were added to the protocol during this study (see Fig. 4.3). The steps where Amastatin, the peptide substrates, and EDTA were added to the protocol are labeled in red.

Superior results were obtained when Amastatin was used in combination with the addition of high concentration of peptide substrates, such as Angiotensin 1-5 and Met-Lys-Bradykinin, which could significantly compete out BBM proteins from degradation. Finally, according to the literature (99) divalent cations, such as Ca^{+2} and Mg^{+2}, activate His-Zn-dependent metalloproteases. It was then essential to remove the excess of Ca^{+2} after the $CaCl_2$ precipitation and 40 mM EDTA was added to the first wash of the BBM membrane for this purpose. The flow diagram of the optimized BBM preparation, including the above modifications, is represented in the figure 4.9.

In order to check the level of inhibition, a Western blot analysis of two BBM fractions (no inhibitor; in presence of the inhibitor, peptide substrates and EDTA) was performed to compare the levels of degradation of Aminopeptidase N (fig. 4.10):
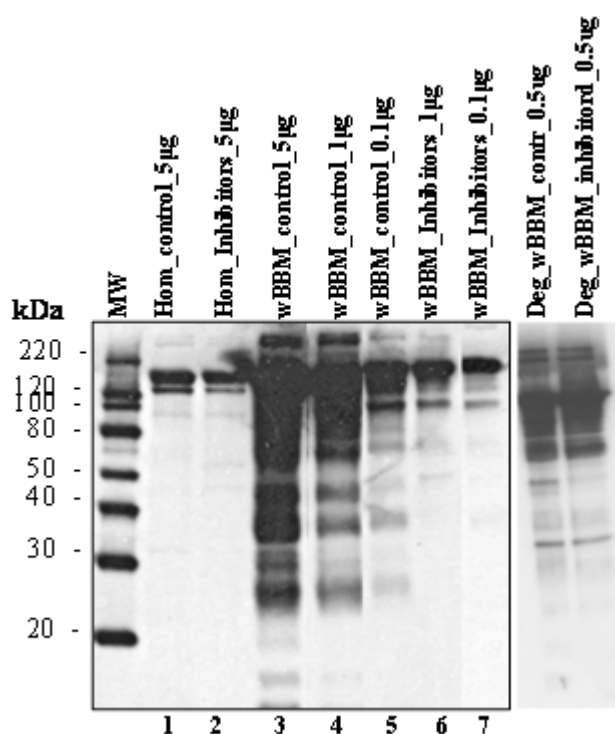
Fig. 4.10: *Western Blot analysis of Aminopeptidase N of a control BBM (lanes 1, 2 following protocol as in Fig. 4.3), of a BBM fraction isolated with inhibitor, peptide substrates and EDTA lanes 3-7, following protocol as in Fig. 4.9), and of a deglycosylated BBM fraction (lane 8, 9, following protocol as in Fig. 4.8). Total protein load of BBM were loaded as indicated in the legend. Probing and detection conditions were as indicated in Fig. 4.8. MW, molecular weight marker.*

The Western Blot analysis against Aminopeptidase N of a dilution serie of the two BBM fractions demonstrated that the combination of a partial inhibitor, Amastatin, of two peptide substrates in large amounts, and of EDTA to remove excess Ca^{2+} after $CaCl_2$ precipitation, resulted in an approximately 10-fold reduction of degradation (Fig. 4.10, compare lanes 3-5

85

versus lanes 6, 7). Interestingly, the deglycosylation procedure appeared to completely void the relative protection conferred by the addition of the inhibitor, peptides substrates and EDTA. Once deglycosylated, the two BBM samples were observed with the same (extensive) level of Aminopetidase N degradation (Fig. 4.10, lanes 8, 9). Since protein degradation couldn't be controlled in any way during the deglycosylation process, the glycosylated form of the BBM fraction was chosen for further investigation in the current study.

The BBM fractions, isolated according to the modified protocol in presence and absence of inhibitors were compared in 1D SDS PAGE representation (Figure 4.11).
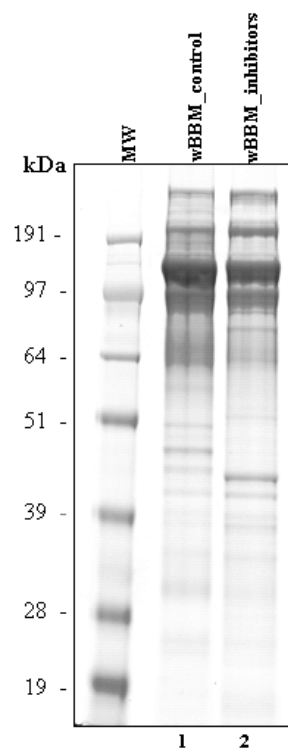
Figure 4.11: *1D SDS PAGE analysis of BBM fractions. Lane 1: wBBM_control, BBM isolated according to the modified protocol without addition of extra inhibitors (Fig. 4.3) Lane 2: wBBM_inhibitor, BBM isolated according to the modified protocol in presence of Amastatin, the peptide substrates and EDTA. 15 μg total protein were loaded per lane in a 10 % 1D NuPAGE Bis Tris gel run in a MOPS running buffer system. MW, molecular weight.*

The BBM fractions in presence and absence of the Zn-His metallopeptidase inhibitor, peptide substrates and EDTA, showed some differences in protein distribution on the 1D SDS PAGE analysis (figure 4.11). The BBM_control lane (lane 1) showed, in comparison to the BBM_inhibitors lane (lane 2), a higher diffusion of the proteins in the higher molecular masses and several prominent protein bands in the lower mass range, which is consistent with a higher degradation activity in the BBM fraction isolated without inhibitor. Characterization of the BBM fraction prepared in presence of Amastatin and the peptide substrates also

showed that protein identification was less redundant between SDS-PAGE bands (see also section 4.2) than with the BBM fraction isolated without inhibitor. Indirectly, this observation confirmed that protein degradation, the extent of which could be easily followed through the fragment distribution of abundant proteins in a SDS-PAGE analysis, was significantly reduced. As desirable as it would be, however, total inhibition of peptidase activity did not appear to be achievable without a major change of the BBM isolation protocol, which was outside of the scope of this study.

4.2 Protein identification of BBM mice intestinal mucosa

Proteins from the BBM mice intestinal mucosa were identified in triplicate using the following analytical strategy. A pool of intestinal mucosa (scrapped from 8 mice) was thawed, weighted, homogenized, and immediately divided in three equal parts. The BBM isolation procedure was performed in parallel for the three preparations. 30 µg of each purified BBM preparation were then analyzed on a single SDS-PAGE gel and each lane was then divided in 19 unequal bands to keep the protein amount in each band as similar as possible. The 57 resulting bands were then in-gel digested with trypsin and the extracted peptides were stored at -80 °C until further use. Each sample was then analyzed randomly (taking care not to analyze adjacent bands one after the other) by LC-ESI-MS/MS with one blank run between each sample to minimize carry over. The MS/MS analyses were then submitted to SEQUEST and the proteins identified in each of the 57 gel bands were stored in MSpresso to be then downloaded to Microsoft Excel.

A unique, non-redundant protein list for each BBM preparation was created as follows: for each band, redundant protein entries were first removed by collapsing a common group of proteins to one single entry, preferably with a SWISS-PROT entry, so to remove splice variants and multiple naming of the same protein from the list. Further, a protein list for each BBM preparation was created by combining the protein lists of the 19 bands, removing the redundant identification of a protein among the analyses but keeping the maximum number of different peptides found in a given band and the sum of peptide counts in the whole preparation. Finally, in the absence of a validated protein scoring to determine a false discovery rate, a protein was arbitrary considered to be unambiguously identified if two different peptides belonging to the same protein could be characterized in a given band. Similarly, a protein was considered to be constitutively part of the BBM preparation if it was identified unambiguously with at least two different peptides in one band in one given preparation and with at least one peptide in another BBM preparation. Fig. 4.12 provides a

short overview of the protein identification from the triplicate measurement of BBM preparation.
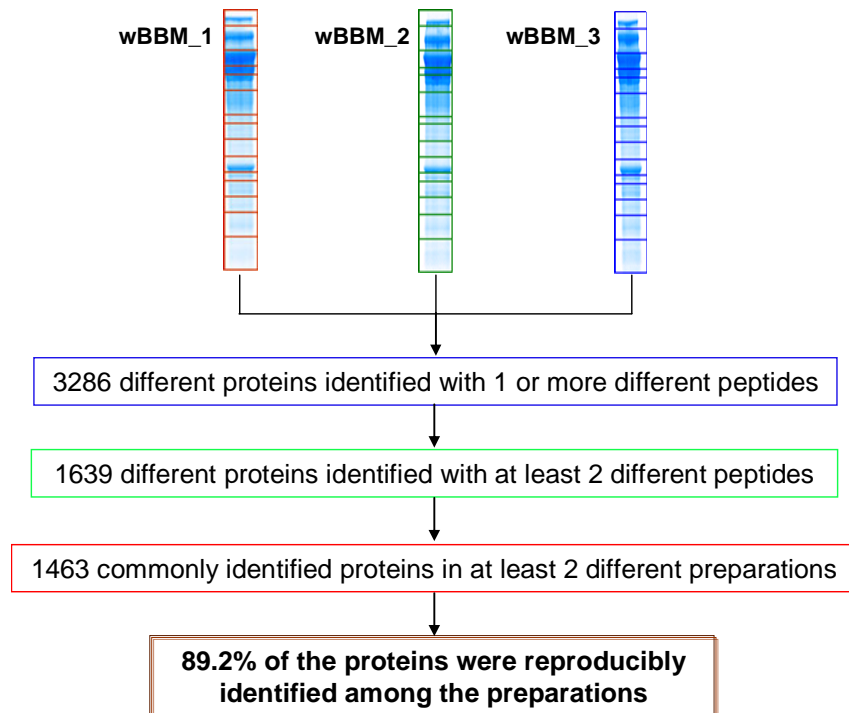
Figure 4.12: *Overview of the protein identification obtained from a triplicate BBM membrane preparation.*

The 1307 proteins for which a valid SWISS-PROT/TrEMBL entry could be found are listed in appendix A1 (see also below). Proteins were ranked-ordered using their average peptide counts as a very rough estimate for their abundance in the sample. Spectral (peptide) counting is by far not an ideal measure for quantification purpose but this strategy is useful evaluating the dynamic range of the proteins included in the analysis. As the abundance of a protein rises in the sample, the amount of peptides derived from this protein also increases in the peptide mixture submitted to the mass spectrometer. Abundant peptides elute in wider chromatographic peaks than scarce peptides, and will therefore be picked up more often for MS/MS analysis. This strategy, however, tends to bias large, cytosolic proteins, which will generate numerous peptides, versus small proteins, for which only a small number of peptides can be found, or membrane proteins, of which only domains outside of the membrane regions are accessible for analysis by mass spectrometry (see also below).

The first 50 rank-ordered proteins of the triplicate preparation are listed in Table 4.1. As expected, a large number of those abundant proteins are part of the cytoskeleton, an essential component of the membrane structure. The quality of the preparation can be measured by the presence of a number of multiply-transmembrane proteins, such as the *Abcb1a* or the high-

affinity sodium/glucose cotransporter 1 transporters, that have been identified for the first time in a proteomics study. Similarly, the Niemann-Pick C1-like protein 1, a predominant protein of the BBM and a major player in cholesterol absorption, was also found among the 50 most abundant proteins of the preparation. Note also the presence of 9 proteases and 4 glycosidases in this list, mirroring the major function of the small intestine in degrading complex molecules to facilitate their ingestion.

Interestingly, the rank-order of the identified membrane proteins did not seem to be biased by the number of transmembrane domains that these proteins were predicted to contain. Thus, the Niemann Pick C1-like protein, a predominant protein of the BBM bearing 13 transmembrane helices, was identified with 32 different peptides and an average sum of 142 peptide counts. In contrast, the fatty acid transporter protein 4 (FATP-4), a protein bearing only two transmembrane helices, was identified with 3 different peptides and average sum of 3 peptide counts. In this preparation highly enriched for membrane proteins, the number of different peptides and the average sum of peptide counts reflected the relative abundance of a given membrane protein in the preparation. Thus, as an example, the FATP-4 protein could not be identified anymore in the preparation when the protein loading on the SDS-PAGE was reduced to 2/3 of the original amount (results not shown).

Rank order	Gene Symbol	Gene Description	Max Diff peptide	Avg Pep count	# AA	Number of TM
1	Actb	actin, beta, cytoplasmic	29	1737	375	0
2	Vil1	villin 1	58	1566	827	0
3	2010204N08Rik	hypothetical protein LOC69983	sucrase-isomaltase (alpha-glucosidase)	93	1464	1818	1
4	Mgam	maltase-glucoamylase	81	1196	1857	1
5	Actg1	actin, gamma, cytoplasmic 1	29	983	375	0
6	Anpep	Aminopeptidase N (EC 3.4.11.2)	54	961	966	1
7	Myo1a	myosin IA	75	780	1043	0
8	Myo7b	Myosin-VIIb.	108	733	2113	0
9	Slc5a1	High affinity sodium/glucose cotransporter 1 (Solute carrier family 5 member 1).	23	587	665	14
10	Enpep	Glutamyl aminopeptidase (EC 3.4.11.7) (Aminopeptidase A)	53	431	945	1
11	Abcb1a	Multidrug resistance protein 3 (EC 3.6.3.44) (ATP-binding cassette sub-family B member 1A)	51	311	1276	11
12	Cltc	clathrin, heavy polypeptide (Hc)	73	303	1675	0
13	Acta1	actin, alpha 1, skeletal muscle	14	280	377	0
14	Actc1	actin, alpha, cardiac	14	280	377	0
15	Lct	lactase	47	251	1220	0
16	Actg2	actin, gamma 2, smooth muscle, enteric	14	250	376	0
17	Acta2	actin, alpha 2, smooth muscle, aorta	14	250	377	0
18	Atp1a1	Sodium/potassium-transporting ATPase subunit alpha-1 precursor (EC 3.6.3.9)	36	244	1023	10
19	Pcdh24	protocadherin 24	28	241	1308	1
20	Gna11	guanine nucleotide binding protein, alpha 11	28	236	359	0
21	Ubb	ubiquitin B	8	226	76	0
22	Muc13	Mucin-13 precursor (MUC-13) (Cell surface antigen 114/A10)	13	219	573	1
23	Mep1b	Meprin A subunit beta precursor (EC 3.4.24.18) (Endopeptidase-2).	23	204	704	1
24	Pdzk1	PDZ domain containing 1 (Na/Pi cotransporter C-terminal-associated protein)	38	179	519	0
25	2210407C18Rik	RIKEN cDNA 2210407C18 gene / EP1 protein (novel protein)	11	176	220	0
26	Mme	Neprilysin (EC 3.4.24.11) (Neutral endopeptidase 24.11)	39	175	750	1
27	Ace2	Angiotensin-converting enzyme 2 precursor (EC 3.4.17.-) (ACE-related carboxypeptidase)	34	174	805	1
28	Myo1d	myosin ID	54	172	1006	0
29	Npc1l1	Niemann-Pick C1-like protein 1 precursor.	32	142	1333	13
30	Papss2	3'-phosphoadenosine 5'-phosphosulfate synthase 2	33	140	366	0
31	Eps8l3	ESP8-like 3	28	140	600	0
32	Akp3	Intestinal alkaline phosphatase precursor (EC 3.1.3.1) (IAP).	28	137	559	0
33	Lima1	LIM domain and actin binding 1	18	136	753	0
34	Hsp90ab1	Heat shock protein HSP 90-beta (HSP 84)	33	136	724	0
35	Ezr	ezrin	37	131	586	0
36	Ggt1	Gamma-glutamyltranspeptidase 1 precursor (EC 2.3.2.2) (Gamma-glutamyltransferase 1)	14	127	568	1
37	AI427122	Plastin 1 (I isoform) homolog	34	122	630	0
38	Gnb1	guanine nucleotide binding protein (G protein), beta 1	15	116	130	0
39	Anxa2	annexin A2	29	115	339	0
40	Gnaq	guanine nucleotide binding protein, alpha q polypeptide	20	115	353	0
41	Tuba1b	tubulin, alpha 1B	16	115	451	0
42	Ubc	ubiquitin C	8	115	76	0
43	Dpep1	dipeptidase 1 (renal)	19	113	410	0
44	Enpp3	Ectonucleotide pyrophosphatase/phosphodiesterase family member 3 (E- NPP 3)	26	112	874	1
45	Dpp4	Dipeptidyl peptidase 4 (EC 3.4.14.5)	33	107	760	1
46	Gnb2	guanine nucleotide binding protein (G protein), beta 2	12	107	493	0
47	Tuba1c	Tubulin alpha-1C chain (Tubulin alpha-6 chain)	16	105	449	0
48	Hspa8	heat shock protein 8	33	104	646	0
49	Treh	Trehalase precursor (EC 3.2.1.28)	26	103	576	0
50	Naaladl1	N-acetylated-alpha-linked acidic dipeptidase-like protein (EC 3.4.17.21) (NAALADase L).	26	101	745	1

Table 4.1 The 50 most abundant proteins in the triplicate BBM preparation. *Transmembrane proteins are noted in yellow, proteases are marked in pink.*

90

Analysis of the nature and topology of the proteins identified in the triplicate BBM membrane preparation was performed as follows. Each of the 1463 protein sequences (which were originally derived from the mouseGP sequence database) was first matched to a SWISS-PROT/TrEMBL entry, resulting in 1307 valid entries (see appendix A1). Further, histone proteins, ribosomal proteins, as well as entries pointing to pseudogenes, were also removed from the pool of proteins to be analyzed to simplify the ensuing data analysis. (Table 4.2).

1463 mouseGP sequences	1206 SW/Tr sequence entries	further analyzed
	38 histone protein entries	Not analyzed
	54 ribosomal proteins	Not analyzed
	9 pseudogenes	Not analyzed
	156 unmatched gene products	Not found

Table 4.2: *Overview of the proteins identified in the triplicate BBM membrane preparation*
As there is no general method to comprehensively query the topology of a large number of protein species, the remaining 1206 entries were analyzed by partially overlapping strategies. First, proteins were categorized as "secreted", "membrane proteins" or "cytoplasmic" based on the predicted presence of a signal peptide targeting the protein for secretion (signal_anchor tool) or on the presence of transmembrane sections anchoring a protein in a biological membrane (ALOM tool). The categorization "membrane protein" was further investigated with the TMHMM software package, which also allows predicting the number of transmembrane sections in a protein. Finally, the most relevant GO "cellular location" annotation of a protein was compiled to assess the nature and the topology of the 1206 investigated proteins (see appendix A1).

In a general manner, the three approaches provided similar results, albeit with considerable variation in respect to sensitivity and specificity. Fig. 4.13 shows the performance of the TMHMM and the ALOM algorithms (which both will flag a membrane protein) versus a manual search in the GO annotation filtered for the concept "integral membrane proteins".
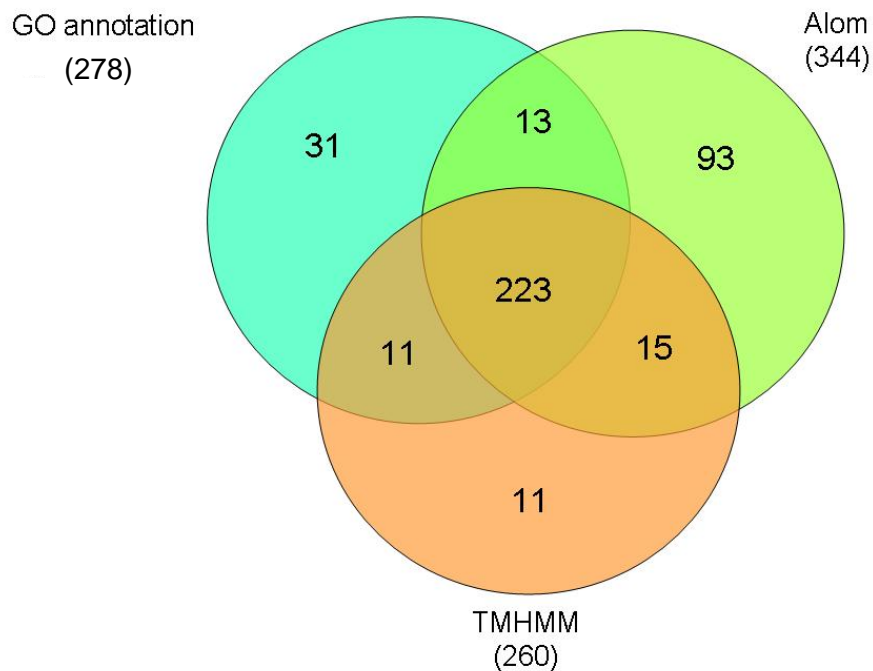
Figure 4.13: *Predictive assignment of membrane proteins by the Alom software package and the TMHMM software package compared to the GO annotation of the 1206 proteins included in the analysis.* *The Go assignment included the terms "apical plasma membrane", "basolateral plasma membrane" and "integral to membrane".*

The GO annotation method and the TMHMM algorithm roughly flagged the same proteins with more than 85% of the potential candidates commonly identified as a membrane protein. On the other hand, the Alom algorithm assigned about 20% more candidates as membrane proteins as the other strategies. This algorithm might be slightly more sensitive than TMHMM to pick membrane proteins, at the cost of a large increase in false assignment. However, compared to the GO annotation method, both algorithms appear to significantly underestimate the number of integral mitochondrial membrane protein (e.g. ATP synthase, H^+ transporting mitochondrial F1 complex, beta subunit, ATPA_MOUSE, or solute carrier family 25 (mitochondrial carrier ornithine transporter), member 15, ORNT1_MOUSE) and ion channels (e.g. voltage-dependent anion channel 2, VDAC2_MOUSE). Most of those proteins most probably belong to the structural class of transmembrane β-barrel proteins, which can not be easily predicted even using the most recent software packages for structure predictions (100).

The transmembrane segment distribution of the 260 membrane proteins predicted by the TMHMM algorithm is shown in Fig. 4.14. The distribution of membrane proteins with one or two predicted transmembrane segments represented more than 50% of all the membrane proteins, closely following a whole genome predictive analysis of the number of protein with

transmembrane helices in human (101). The number of remaining membrane proteins was rather equally distributed among all the other membrane protein species with the interesting exception of the 7-transmembranes protein family (dominated by the GCPRs), which was underrepresented in this analysis while it is overrepresented and should account to 10% of all the proteins with transmembrane helices in human (Liu et al., 2001). This observation might reflect the main functional aspect of the enterocyte BBM, which is the transport of nutrients from the intestinal lumen to the lymphatic system rather than intracellular signaling.

#proteins with TM segment
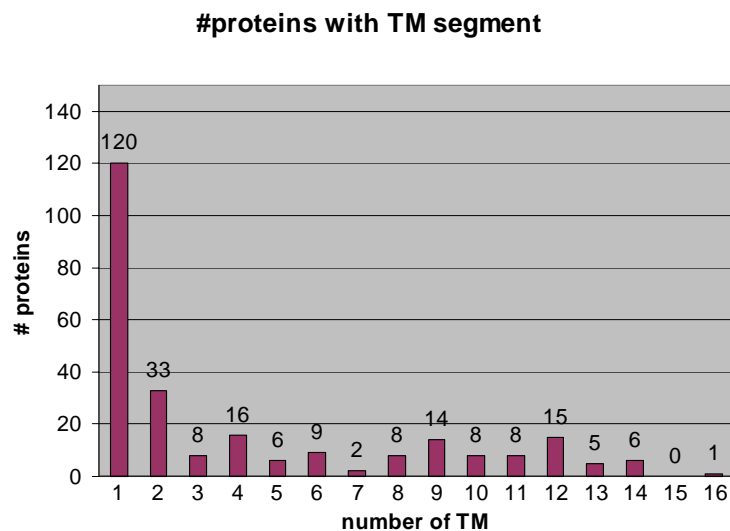
Figure 4.14: *Distribution of transmembrane segments in the 260 membrane proteins predicted by the TMHMM algorithm.*

At first sight, the prediction of "only" about 260 membrane proteins out of a total of 1206 proteins (slightly more than 20%) might look disappointing. However, a detailed analysis of the GO annotation associated with the data showed a much more differentiated picture of the protein population that was characterized. As a note of caution, however, it should be noted that the vast majority of the protein analyzed was associated with more than one GO annotation, depending of the function and localization of the protein in a cell. Also, many of those annotations were rather vague or even sometimes contradictory. As a result, the GO annotations of all the considered proteins were manually filtered so to keep what was believed to be the most relevant entry in the context of the small intestine and the enterocyte function (see appendix B for a complete listing of the GO annotation kept in this analysis). Finally, the approximately 90 categories remaining after the first data compilation had to be further collapsed in organelle-based groups to enable a more general overview of the proteins (Table 4.3)

"Collapsed" GO annotation	# proteins
associated to plasma membrane	52
anchored to plasma membrane	60
integral to plasma membrane	144
associated to vesicle/ER/Golgi/endosome/microsome	80
anchored to vesicle/ER/Golgi/endosome/microsome	29
integral to vesicle/ER/Golgi/endosome/microsome	104
ER/Golgi/endosome/microsome matrix	23
associated to mitochondrial membrane	19
anchored to mitochondrial membrane	2
integral to mitochondrial membrane	16
mitochondrial matrix	26
cytoskeleton	98
proteasome	29
cytoplasm	331
nucleus	31
integral to nuclear membrane	1
extracellular	35
No or ambiguous "GO" annotation	126

Table 4.3: *Overview of the collapsed GO annotation categories retained for analysis. "associated": described an interaction with a membrane (such as e.g. "peripheral proteins"); "anchored": annotated as being anchored to the membrane through lipids or other types of membrane anchors; "integral": annotated as having at least one transmembrane segment.*

As could have been expected from this type of membrane preparation, one of the most prominent groups of proteins in the BBM preparation comprised plasma membrane proteins, closely followed by a group of proteins interacting in large with the ER/Golgi/endosomal systems, which are known to be a major "contaminant" of a plasma membrane preparation. The vast majority of the proteins included in these two groups are channels, transporters, and membrane trafficking proteins most likely taking part in the import of nutrients in the enterocyte and the recycling of the membrane proteins back to the plasma membrane. A third important group in this preparation comprised cytoskletal proteins which have also been known to interact with the plasma membrane. In contrast, a comparatively small number of mitochondrial and nuclear proteins were characterized in this preparation, reflecting the high grade of purification achieved by this protocol. Only about a quarter of the proteins characterized in this preparation were annotated as "cytoplasmic", representing many

metabolic pathways, but also proteins known to interact with membrane-bound proteins such as chaperonins or heat shock proteins. Finally, while the high number of ribosomal proteins might be reasoned by the tight interaction of the ER network with the plasma membrane, the presence of a large number of histone proteins in this preparation can not directly be explained, as they should locate exclusively in the nucleus.

4.3 Examples of protein localization

Some BBM proteins have been described in the literature to be localized in specific regions of the small intestine, so that the question was raised whether such specificity could be demonstrated using the BBM preparation described above. For this effect, BBMs from mice intestinal mucosa were prepared taking care to keep separate the proximal, central and distal segments of the small intestine. In this experiment, the three segments had equal length (the small intestine was cut in thee equal parts) although, according to the anatomical characteristics of the small intestine, the duodenum (the proximal part) is only a short section after the Pylorus, while the jejunum (the central part) accounts for 90% of the total length of the small intestine and the ileum (the distal part) is only a short terminus section. The BBM of each section were loaded in 1D SDS PAGE (see figure 4.6, panel C). Each gel lane was cut in 21 unequal gel bands and the samples were in-gel digested with trypsin. The extracted peptides of each band were then analyzed by LC-MS/MS and the raw data were processed according to the criteria previously described. The proteins identified for each section and for the total mucosa are listed in Appendix A2 along with their distribution along the sections (based on the maximum of different peptides and the total peptide counts for each protein). This experiment clearly demonstrates that some proteins were identified with more peptides from BBM isolated in section rather than from whole mucosa. In particular, there were several proteins that clearly located exclusively in only one of the three sections, such as the SR-BI receptor. The identification of this receptor was rejected in the BBM prepared from whole intestinal mucosa because it was identified with only one peptide. The protein identification carried out in each section separately confirmed the existence of the SR-BI receptor and its specific localization in the proximal part (see table 4.4 below). This finding was in agreement with a Western blot analysis against SR-BI, where the scavenger receptor was present almost exclusively in the duodenum segment (Fig. 4.15).
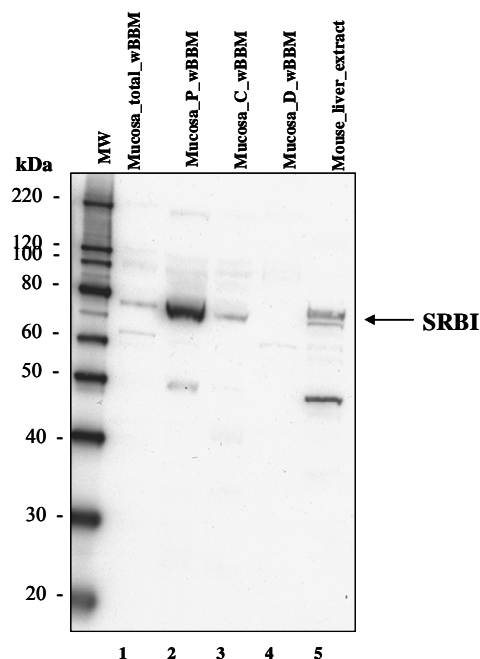
Figure 4.15: *Western blot analysis of SR-BI in BBM prepared either from full intestinal mucosa or from sections. SR-BI immunoblot in BBM isolated from total mouse intestinal mucosa (lane 1) or from each section separately (lane 2-4). Mouse liver extract was used as a positive control (lane5). 30 μg of total protein was loaded in each case. 1st antibody 1:2500 rabbit polyclonal anti-SR-BI (Cat. No:NB 400-101, NOVUS Biologicals), 2nd antibody 1:7500 Horseradish peroxidase goat anti rabbit IgG.*

In the sectional approach of protein identification in the BBM membrane, several proteins were found equally distributed along the small intestine (see examples on table 4.4). Some other proteins, such as FATP-4, were identified with more peptides in the BBM prepared from total mucosa than in sections. This might be due to more efficient washes (higher volume to mass ratio) of the BBM with high salt and high pH solutions. In this case, not only BBM contaminants were removed more effectively but also membrane-associated proteins and even integral proteins with one or two transmembrane helices were washed out.

GeneID	Data	Proximal	Central	Distal	Total	Protein Name
237636	Max_Dif_pept	36	31	33	29	Niemann-Pick C1-like protein 1 precursor
	Sum_pept_count	200	179	215	148	
20778	Max_Dif_pept	3			1	SR-BI
	Sum_pept_count	3			1	
26458	Max_Dif_pept	16	3	1	10	Long-chain-fatty-acid--CoA ligase
	Sum_pept_count	32	3	1	11	
20494	Max_Dif_pept		2	6	5	IBAT
	Sum_pept_count		3	35	15	
67470	Max_Dif_pept	18	16	14	12	Sterolin-2
	Sum_pept_count	99	62	69	50	
64452	Max_Dif_pept	9	11	9	10	Low affinity sodium-glucose cotransporter 3
	Sum_pept_count	42	41	32	32	
59020	Max_Dif_pept	29	38	28	42	Na(+)/H(+) exchanger regulatory factor 3
	Sum_pept_count	205	356	218	297	
26569	Max_Dif_pept	1		1	9	FATP-4
	Sum_pept_count	1		2	10	

Table 4.4: *Examples of protein distribution along the BBM membranes of the small intestine. Comparison of BBM analysis in sections vs. total BBM.*

According to the data presented on the table 4.4, it is most probable to assign the SR-BI receptor and the long-chain-fatty acid-CoA ligase to be exclusively located in the proximal part of the small intestine and the IBAT transporter to the distal part.

4.4 Cholesterol absorption

4.4.1 Identified proteins related to Cholesterol absorption

One of the stated goals of this study was to investigate whether a proteomics approach could confirm the presence of proteins known to be involved in cholesterol absorption in the BBM. Table 4.5 lists the proteins (among the 1300 proteins identified in the total BBM intestinal mucosa preparation) that have been described in the literature, mainly based on gene data or immunoassays, to play a role in cholesterol absorption in the enterocyte.

Rank	Gene Symbol	Gene Description / Protein Header	Avg Pep count	Max diff peptides
6	Anpep	Alanyl (membrane) aminopeptidase	961	54
11	Abcb1a	Multidrug resistance protein 3 (MDR1A)	311	51
29	Npc1l1	Niemann-Pick C1-like protein 1 precursor	142	32
39	Anxa2	Annexin A2	115	29
65	Lgals4	Galectin-4 (Lactose-binding lectin 4)	80	11
66	Abcc2	Canalicular multispecific organic anion transporter 1	78	27
78	Abcg2	ATP-binding cassette sub-family G member 2 (CD338 antigen)	63	20
102	Abcg8	ATP-binding cassette sub-family G member 8 (Sterolin-2)	50	14
108	Atp8b1	Potential phospholipid-transporting ATPase	48	24
190	Mttp	Microsomal triglyceride transfer protein (MTP)	27	19
192	Abcg5	ATP-binding cassette sub-family G member 5 (Sterolin-1)	26	14
335	Cd36	Platelet glycoprotein 4 (CD36 antigen)	13	9
378	Apoa4	Apolipoprotein A-IV	11	8
389	Slc10a2	IBAT (Apical sodium- dependent bile acid transporter)	11	5
417	Anxa7	Annexin A7	10	7
479	Pdia3	Protein disulfide isomerase associated 3	8	5
501	Slc27a2	Fatty acid transport protein 2 (FATP-2)	8	8
545	Slc16a1	Monocarboxylate transporter 1 (MCT 1)	8	3
559	Apoa1	Apolipoprotein A-I	8	8
562	Npc1	Niemann-Pick type C1 protein	5	2
586	Lrp1	Low density lipoprotein receptor-related protein 1	8	8
614	Fabp6	Fatty acid binding protein 6, ileal (gastrotropin)	5	3
653	Abcb11	Bile salt export pump (Sister of P-glycoprotein)	5	2
826	Cubn	Cubilin (intrinsic factor-cobalamin receptor)	4	4
919	Cav1	Caveolin-1	3	3
947	Slc27a4	Long-chain fatty acid transport protein 4 (FATP-4)	3	3
1001	Scp2	Sterol carrier protein 2, liver	2	2
1055	Apoe	Apolipoprotein E	3	3
1056	Sec14L2	Sec14-like	3	3
1452	Scarb1	Scavenger receptor class B member 1 (SR-BI)	2	1

Table 4.5: *Potential protein candidates for intestinal cholesterol transport. The Rank order is a very rough approximation for protein abundance, based on the average peptide counts combined with the maximum different peptides for a given protein, not taking into account its size (large proteins generate more peptides than small ones) or its hydrophobicity (hydrophilic proteins generate more peptides than hydrophobic ones).*

The Niemann-Pick C1-like 1 protein (NPC1L1) is a representative example of an abundant protein of the BBM preparation (table 4.5: rank order 29). This 13-transmembrane segments protein has been described to play a critical role in cholesterol absorption and it is believed to be the target of Ezetimibe, a cholesterol absorption inhibitor that blocks the transport of cholesterol and phytosterols across the BBM of enterocytes (13). Other studies have attempted to show that Aminopeptidase N (table 4.1: rank order 6) might represent the targeted protein of Ezetimibe (102). Further, malfunction or mutation affecting the two half-size, 6-transmembrane segments ABC transporters ABCG5 and ABCG8 (table 4.5: rank order 192 and 102, respectively) can lead to a rare autosomal recessive disorder called sitosterolemia or phytosterolemia and is caused by hyperabsorption and impaired biliary secretion of cholesterol and plant sterols. The complex formed by Caveolin-1 and Annexin-2 (table 4.5: rank order 919 and 39) has been suggested as key element for the cholesterol trafficking from the BBM to the endoplasmic reticulum (103). Finally, a pioneering work of Hauser et al. (104) showed that cholesterol uptake is reduced if SR-BI, a 4-transmembrane segments receptor, is blocked by anti-SR-BI antibodies or by competitive ligands, such as

apolipoprotein A-I. Similarly, SR-BI overexpression was found to increase the intestinal cholesterol absorption (105). The role of many of the proteins mentioned above and of several other proteins, such as CD36, Galectin-4, and ABCB1, is the subject of hot debates. Several studies have supported their involvement in cholesterol absorption but their precise mechanisms of action have remained unclear so far (see for example a review from Levy E. et al. (106)).

It is worth mentioning here the presence of three lipoproteins, ApoA-I, ApoA-IV and ApoE, within the identified proteins of the BBM preparation. Their confident identification in this membrane preparation was rather unexpected as lipoproteins are by nature small, soluble proteins secreted by the liver and, therefore, they are not considered as constituents of the BBM. Lipoproteins were expected to be washed out from the BBM preparation during the several high salt (1M KCl) and high pH (100 mM Na_2CO_3) washes and to be identified in rather large amounts in these two fractions. Surprisingly, they were found in the BBM preparation, and not in the wash fractions (table 4.6).

Gene Name	wBBM	wNa2CO3	wKCl
ApoA-I	8	1	0
ApoA-IV	8	2	0
ApoE	3	0	0

Table 4.6: *Maximum of different peptides obtained for the ApoA-I, ApoA-IV and ApoE proteins. Identification of the three apolipoproteins in the BBM fraction, the KCl and Na_2CO_3 wash fractions.*

The finding that the three lipoproteins (ApoA-I, ApoA-IV and ApoE) were strongly enriched in the BBM fraction postulates a strong interaction between these lipoproteins and a BBM receptor or transporter. Indeed, two lipoproteins receptors, LRP-1 and Cubilin, were identified among the BBM-identified proteins; LRP-1 was shown to directly interact with ApoE (107) while Cubilin (by itself not a transmembrane receptor, but rather a co-transporter most probably located on the luminal side of the enterocyte plasma membrane) has been documented to interact with ApoA-I (108). Interestingly, in the kidney proximal tubule, cubilin interacts strongly with another receptor, megalin (also known as LRP-2) which is postulated to mediate internalization of cubilin and its ligands. Megalin, however, could not be identified in our study and was shown by mRNA study to be only present in the distal part of the intestine (109). Finally, as noted above, the SR-BI receptor has also been shown to interact with the apolipoprotein ApoA-I (110).

In addition to the apical surface localization of the above described proteins (table 4.5), intestinal cholesterol transporters have also been detected in intracellular compartments. For example, SR-BI has been described to be mainly localized in the microvillar membrane of enterocytes in the fasting state, but was endocytosed during absorption of dietary fat (111). The role of NCP1L1 in cholesterol trafficking has also been documented: while it is mainly observed in endocytic recycling compartments as long as internal cholesterol pools are abundant, it is rapidly translocated to the plasma membrane in situation of cholesterol depletion, which in turn was associated with a large increase in cellular cholesterol uptake (112).

These findings suggest that intestinal cholesterol transporters may also act intracellularly, mediating the movement of cholesterol from BBM to various organelles. In particular, they may assist in the shuttling of cholesterol from BBM to the endoplasmic reticulum where the incoming cholesterol represents the major source of substrate for ACAT and restrains HMG-CoA reductase activity (113). The localization of several transporters and other cholesterol absorption related proteins to several organelles shows that the process of cholesterol absorption is probably much more complicated than is known at this point of time. More studies will clearly be required to help understand the molecular mechanism of intestinal transporter-mediated cholesterol trafficking and regulation.

4.4.2 Comparison of protein expression in the BBM of wild type mice and ApoE knockout mice

One of the most widely used mouse models to study dislipidemia is the apolipoprotein E–deficient mice (ApoE-/- mice), in which targeted deletion of the *apoE* gene leads to severe hypercholesterolemia and spontaneous atherosclerosis. ApoE is synthesized in the liver and in macrophages and has a number of important anti-atherogenic functions. As a constituent of plasma lipoproteins, it serves as a ligand for the cell-surface lipoprotein receptors such as LDL-receptor (LDLr) and LDLr-related proteins (LRPs), thereby promoting the uptake of atherogenic particles from the circulation. Consequently, homozygous deletion of the apoE gene in mice results in a pronounced increase in the plasma levels of LDL and VLDL attributable to the failure of LDLr- and LRP mediated clearance of these lipoproteins (114).

The finding that ApoE strongly interacts with the BBM membrane (for details see § 4.4.1), added to the fact that many proteins involved in cholesterol absorption could also be identified using our protocol, encouraged us to investigate in more details the protein expression differences that could be observed in the BBM membrane of wild type mice versus ApoE knockout mice using the same proteomics approach as described above. Possible

differences in protein abundance could contribute to shed a better understanding in the mechanisms of high cholesterol absorption observed for the ApoE knockout mice.

The isolated BBM fractions of a pool of four male ApoE-deficient mice (B6.129P2-Apoetm1Unc/Crl) and four non-transgenic male mice of the same genetic background and age, were loaded in triplicate onto a 1D NuPAGE gel (see figure 4.16).
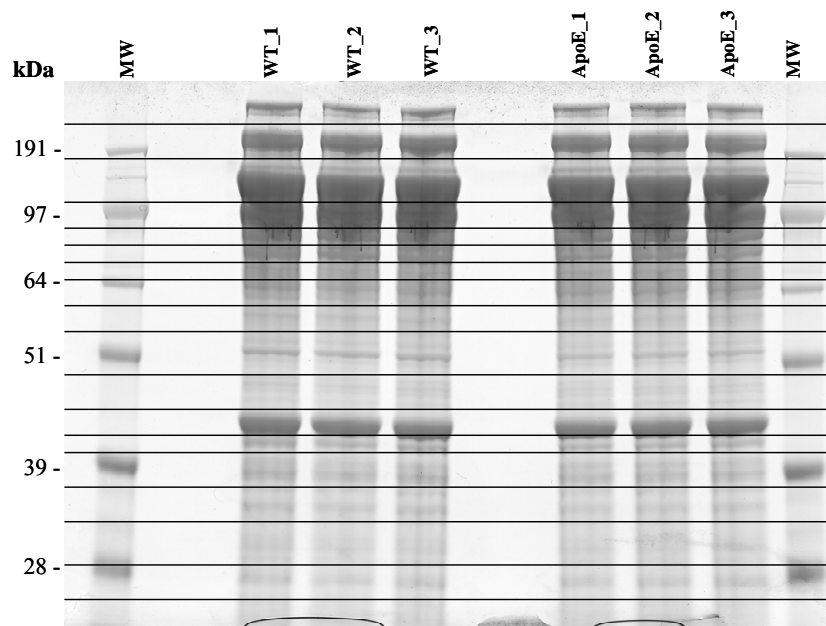
Figure 4.16*: 1D SDS-PAGE gel representation of the BBM fractions of wild type and ApoE-KO mouse. Each fraction (30 μg total proteins per lane) was loaded in a 10 % 1D NuPAGE Bis-Tris gel run in MES buffer.*

The gel bands of the fractions were cut and in-gel digested with trypsin. The extracted peptides of each band were then analyzed by LC-MS/MS and the proteins of each band were identified according to the criteria that have been described at the § 3.2.7.2-3. Identical bands of the compared BBM fractions were analyzed sequentially with washing intervals, to achieve the best possible technical reproducibility.

It was reinsuring that, in large, the identified BBM proteins in this experiment were identical to the BBM proteins that were identified in the previous study (see § 4.4.1). Most interestingly, several differences in protein abundance (based in peptide counts and number of different peptides) were clearly observed between wild type mice and ApoE KO mice. Since the study did not follow any formal quantification workflow, and only technical replicates were performed, we concentrated our interest only in proteins with major difference in abundance between the two types of mice, summarized in Table 4.7.

Gene ID	Protein Header	Data	ApoE	WT
11522	Alcohol dehydrogenase 1 (EC 1.1.1.1) (Alcohol dehydrogenase A subunit)	Max of DifferentPeps	11	6
		Sum of PepCount	15	9
11749	Annexin A6 (Annexin VI) (Lipocortin VI) (P68) (P70) (Protein III) (Chromobindin-	Max of DifferentPeps		9
		Sum of PepCount		9
11806	Apolipoprotein A-I precursor (Apo-AI) (ApoA-I)	Max of DifferentPeps		3
		Sum of PepCount		4
12333	Calpain-1 catalytic subunit (EC 3.4.22.52) (Calpain-1 large subunit) (Calcium-a	Max of DifferentPeps		3
		Sum of PepCount		3
12808	Protein cordon-bleu	Max of DifferentPeps	12	5
		Sum of PepCount	36	17
12846	Catechol O-methyltransferase (EC 2.1.1.6)	Max of DifferentPeps	1	4
		Sum of PepCount	1	4
14828	78 kDa glucose-regulated protein precursor (GRP 78) (Heat shock 70 kDa prot	Max of DifferentPeps	1	4
		Sum of PepCount	1	6
17831	Muc2 protein	Max of DifferentPeps		2
		Sum of PepCount		3
20494	Ileal sodium/bile acid cotransporter (Ileal Na(+)/bile acid cotransporter)	Max of DifferentPeps		4
		Sum of PepCount		6
66898	Brain-specific angiogenesis inhibitor 1-associated protein 2-like protein 1 (BAI1	Max of DifferentPeps	18	10
		Sum of PepCount	63	33
207495	Brain-specific angiogenesis inhibitor 1-associated protein 2-like protein 2 (BAI1	Max of DifferentPeps	4	2
		Sum of PepCount	10	6
237636	Niemann-Pick C1-like protein 1 precursor	Max of DifferentPeps	29	28
		Sum of PepCount	93	104
268663	Adult male intestinal mucosa cDNA, RIKEN full-length enriched library, clone:(	Max of DifferentPeps	14	22
		Sum of PepCount	152	154
11808	Apolipoprotein A-IV precursor (Apo-AIV) (ApoA-IV)	Max of DifferentPeps	5	4
		Sum of PepCount	7	4
27409	ATP-binding cassette sub-family G member 5 (Sterolin-1)	Max of DifferentPeps	9	11
		Sum of PepCount	33	36
67470	ATP-binding cassette sub-family G member 8 (Sterolin-2)	Max of DifferentPeps	11	13
		Sum of PepCount	55	47

Table 4.7: *Examples of protein abundance differences between ApoE KO and wild type mice. Proteins labeled in yellow have been described to participate in cholesterol absorption while those showing a major change in abundance are highlighted in red.*

According to the data presented in Table 4.7, the Ileal Sodium/bile acid cotransporter (IBAT) and apolipoprotein A-I were confidently identified in the wild type mice but were completely absent in the ApoE knockout mice. This finding was confirmed in all the technical triplicates.

These findings were further validated using Western Blot analysis, and the result for IBAT is shown in Fig. 4.17. Unfortunately, the Western Blot analysis for Apo-AI was inconclusive (complete absence of signal).
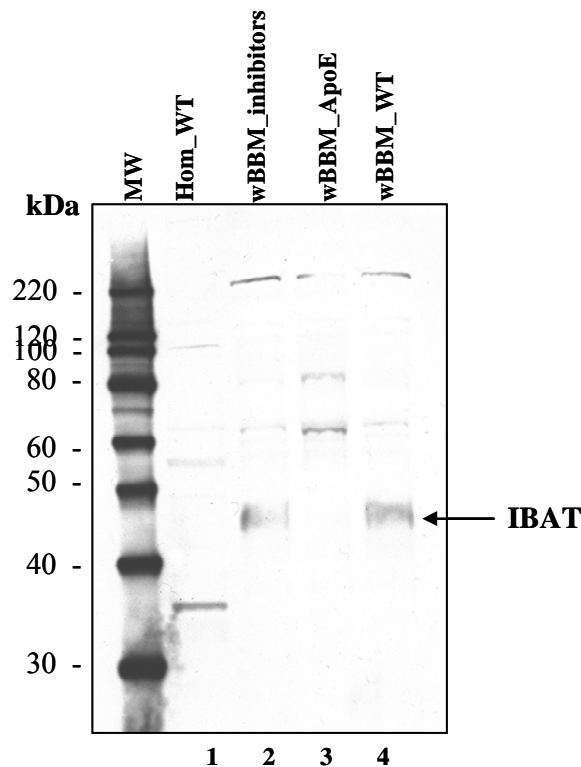
Figure 4.17: *Western blot analysis of IBAT.* Lane 1*, Homogenized intestinal mucosa tissue from WT mice with the same genetic background as the ApoE KO.* Lane 2, *BBM isolated from the total mouse intestinal mucosa of control mice C57B/6J.* Lane 3, *ApoE knockout mice and* (lane 4) *WT mice with the same genetic background as the ApoE-KO mice. 1st antibody 1:500 goat polyclonal anti-IBAT (Cat. No: sc-27493), 2nd antibody 1:7500 Horseradis peroxidase donkey anti goat IgG. MW, molecular weight marker.*

Due to its function, IBAT's down-regulation was expected to have an impact on the bile acid metabolism in the ApoE-knockout mice. For this purpose, a separate in-house study investigated the bile acid pool size and composition in apoE-deficient mice compared to that in wild type mice. Results of this study showed that the liver and intestinal bile acid pool sizes were significantly increased in the apoE-deficient mice compare to the wild type mice (Evelyne Chaput, unpublished data; see Fig. 4.18). Thus, the increased bile acid production in the liver is consistent with disrupted bile acid re-uptake from the intestinal lumen, as expected from reduced IBAT protein expression, and increase the secretion of bile acids in the pancreatic fluids. The increased amount of bile acids in the intestinal lumen might explain why cholesterol absorption is higher in apoE KO mice, since the cholesterol molecule is better solubilized and more easily absorbed. This scenario is well supported by our data, but additional studies need to confirm these findings.
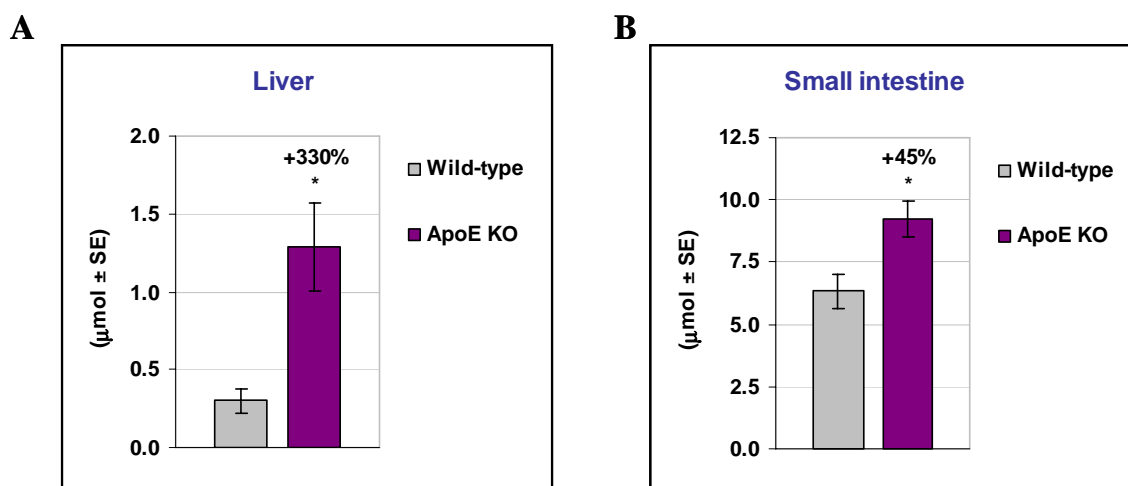
Figure 4.18: *Graphic representation of the total bile acids in liver (panel A) and small intestine (panel B) in ApoE knockout mice and wild type mice. The total bile acids are a sum of all the individual bile acids determined by GC-MS (Adapted from Evelyne Chaput, unpublished data).*

It is worth mentioning that conflicting data have been reported for the ApoE KO mice in the literature. For example, in disagreement with our observations, Hakansson et al. (115) reported that the bile acid pool size of apoE KO does not differ from that of wild-type mice. In addition, the authors report an increased IBAT activity based on the intestinal absorption of tauro-23-[75Se] selena-25-homocholic acid. In a contradictory report, and in agreement with our data, Woollett et al. (116) claim that plasma cholesterol and cholesterol absorption is higher in the ApoE KO mice than in the wild type. Interestingly, all these studies were based on gene data and on pharmacokinetic studies while our findings are based on protein-level datasets, which might be closer to the biological realm of the small intestine thann the above mentioned studies. Another interesting aspect is that most studies focus on the cholesterol absorption and the function of the intestine as a consequence of the cholesterol metabolism in the liver. Here, our data suggest that at least a part of the regulation of the cholesterol absorption happens in the small intestine independently of the liver metabolism. As a matter of fact, it could well be that the critical control elements of cholesterol metabolism are trigged at the small intestine level and that the liver is responding to this trigger rather than to lead it.

4.5 Assessing the reproducibility of the improved BBM preparation

In section 4.2, the measurement of a triplicate BBM preparation at the protein identification level (using rather stringent criteria) resulted in almost 9 proteins out of 10 being identified in at least two of the three replicates, a very encouraging result as to obtain BBM preparations in a reproducible and quantitative fashion. In the following paragraphs, we studied in a

systematic manner the technical steps that might affect most strongly the reproducibility and the stability of the BBM preparation, from the initial lysis step until the protein identification by LC-MS/MS. The inherently variable proteolytic activity of the numerous peptidases located in the BBM membrane was an additional reason to verify the reproducibility of the protein identification. Overall, the technical evaluation of a biological preparation represents an important first step to know how reproducibly identical samples can be prepared. It is also a necessary condition to evaluate with some probability to which extent the protein expression difference between sample and biological groups has to vary to become significant.

4.5.1 BBM preparation procedure

The BBM material required for the reproducibility experiment was generated by performing three technical replicates of the BBM preparation starting from the same intestinal mucosa pool of eight mice used in the protein identification experiment described in section 4.2. As mentioned above, the optimized BBM preparation protocol included the addition of Amastatin and the peptide substrates to all steps to reduce the proteolytic degradation to a minimum (see Fig. 4.9). The three BBM preparation technical replicates were analyzed by 1D SDS PAGE to ensure that the protein distribution between the three replicates were identical (Fig. 4.20).
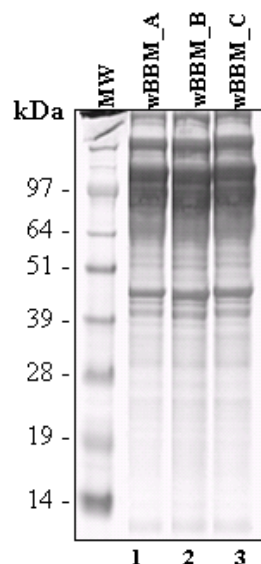
Figure 4.20: *1D SDS PAGE analysis of the three BBM replicates wBBM_A, wBBM_B and wBBM_C (lane 1-3). 30 µg of protein were loaded per lane in a 10 % 1D NuPAGE Bis Tris gel run in a MES running buffer system.*

4.5.2 Comparing the variability of the technical steps: design of experiment

A systematic approach to study the variation that each technical step added to the total variability is shown schematically in Fig. 4.21. The variability of the preparation was examined using these following parameters:

a) Preparation variation (is the preparation reproducible?)

b) SDS-PAGE variation and band excision (what happens if banding patterns are not absolutely reproducible?)

c) LC-MS/MS variation (how reproducible is a LC-MS/MS analysis?)

All these experiments were evaluated on the basis of three representative SDS-PAGE bands, Band 2, Band 9, and Band 11. Band 2 represented a very abundant, very well defined high molecular mass band (expectation: excellent reproducibility) while Band 9 represented a diffuse, very faint band (expectation: not so good reproducibility). Band 11 also displayed a well defined band but differed from Band 2 by the fact that the proteins present in the mass range of the gel might also encompass proteolytic products (expectation: average reproducibility).

The analysis of the three technical replicates as displayed in Fig. 4.21, Panel A, with the gel bands excised serially and analyzed in a random order, using several batches of LC buffers and LC columns (as it was done for the identification experiment described in section 4.2), was representative of the effects observed if most of the sources of technical variations were taken into account in the study and provided therefore an estimate for the maximum technical variability that could be expected if biological or technical replicates could not be analyzed concomitantly. The analysis of the second set of three technical replicates of the BBM preparation loaded in the same gel (Fig. 4.21, Panel B), with the gel bands excised simultaneously (horizontally) and analyzed using the same batch of LC buffers or LC column provided the level of variability related mainly to the BBM preparation since all the following steps remained constant (negligible gel excision, column or buffer variability). The influence of the gel excision on the analysis reproducibility was investigated in more details using a pooled BBM preparation (generated by mixing an equal amount of total protein from each BBM preparation) loaded three times on one SDS-PAGE gel compared to the loading of the same pooled BBM preparation on three different gels (Fig. 4.21, panel C). Finally, replicate injections of the protein digests of band 2 and 9 were representative for the variability coming from the LC/MS system since LC column buffers remained the same for the whole analysis. The variability that was caused by the column or buffer change was estimated by the LC-MS analysis of a 50 fmol standard peptide mixture that was regularly analyzed between samples.
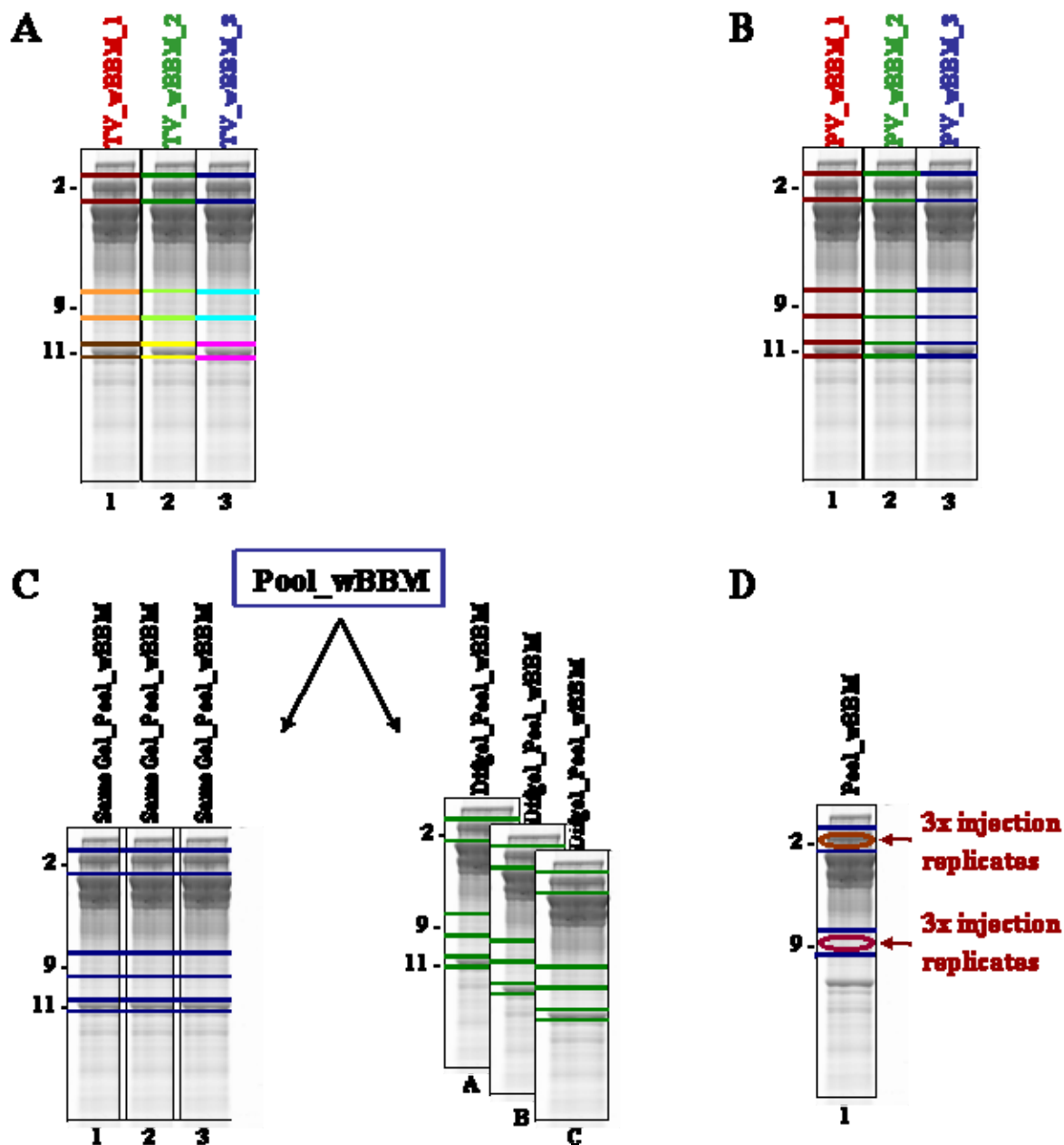
Figure 4.21: *Schematic representation of the experimental design to estimate the variability of each technical step. Three representative bands were chosen for all analyses: 2, 9 and 11. All the samples of this experiment (except the ones from panel A) were measured using the same batch of LC column and LC buffers.* ***Panel A: Total variation (TV).*** *1D SDS PAGE representation of the three technical replicates of the BBM membrane preparation. The samples were loaded in adjacent lanes but the bands were cut serially. This panel is identical to the samples that were analyzed for the BBM protein identification in section 4.2.* ***Panel B: Preparation variation (PV).*** *1D SDS PAGE representation of the three technical replicates of the BBM membrane preparation. The samples were loaded in adjacent lanes and the bands were cut horizontally.* ***Panel C****: A pool of the three technical replicates of BBM was generated (equal amount of protein from each fraction). 30 µg total proteins from the pooled BBM were loaded three times in the same gel in adjacent lanes (lanes 1-3) and once in three different gels (lanes A-C).* ***Panel D:*** *The bands 2 and 9 of the pooled BBM (see also Panel C, lane 1) were injected three times into the LC-MS system. See text for more details.*

4.5.3 Estimation of experimental reproducibility

The reproducibility of the technical steps and of the BBM preparation was estimated based on two different analyses. In the first approach, samples comparison was based on the identification rate in the samples considered. The number of commonly identified proteins in each representative band (2, 9 and 11) within the three replicates of each technical step was illustrated in the form of Venn diagrams. In the second approach, sample comparison was based on the total ion current of each sample's ion chromatogram. Datasets were processed with the Gene data software using the processing filters that have been described in the § 3.2.7.3.

4.5.3.1 Estimation of experimental reproducibility based on Protein identification

The study of the variability of each technical step based on protein identifications was performed using a set of rules that was different from the identification study. In particular, for the purpose of this investigation, proteins with one identified peptide were also considered as successfully identified. Sample comparison was performed in the form of Venn diagrams as there was no simple method to measure similarity between three samples.

4.5.3.1.1 Venn diagrams representation

Venn diagrams are a schematic representation to represent all the possible mathematical or logical relationship between groups. Each group is usually represented in the form of a circle, and overlapping area (intersections) stands for the similarity between groups. In this study, triplicate measurements of the three bands 2, 9 or 11, were compared to each other along the different technical steps. The Venn diagrams of all identified proteins in band 2 categorized by technical step are shown in Fig. 4.22. Each Venn circle represents an LC-MS analysis and the number of identified proteins for the band 2 in every of those analysis. The number outside the circle indicates the total number of proteins that have been identified with at least one peptide in this sample. The intersection between two circles shows the commonly identified proteins between the two compared samples, while the intersection of the three triplicates shows the number of commonly identified proteins in all the three compared samples. The triplicate injections of the band 2 from a BBM pool preparation (fig. 4.22, Panel A) was expected to exhibit the least variability of all the technical steps considered because it reflects the variability of the measurements in the mass spectrometer alone, keeping all other technical steps (nano LC column and buffers) constant. Comparison of the results shown in Panels B and C provides a measure about the variability that is due to the manner SDS-gel

Estimation of variation for the band 2

A   **Injection Replicate**

129   140
15   0   36
13   101   3
32
149

Pool_wBBM_1_2_a
Pool_wBBM_1_2_b
Pool_wBBM_1_2_c

B   **Three runs on the same gel**

143   145
27   6   38
15   95   6
15
131

SameGel_pool_wBBM_1_2
SameGel_pool_wBBM_2_2
SameGel_pool_wBBM_3_2

C   **Three runs on a different gel**

145   155
32   8   35
12   93   19
29
153

Difgel_pool_wBBM_A_2
Difgel_pool_wBBM_B_2
Difgel_pool_wBBM_C_2

D   **Preparation variation**

133   150
15   17   27
3   98   8
79
188

PV_wBBM_1_2
PV_wBBM_2_2
PV_wBBM_3_2

E   **Total analysis variation**

139   136
40   10   32
12   77   17
46
152

TV_wBBM_1_2
TV_wBBM_2_2
TV_wBBM_3_2

Figure 4.22: *Venn diagram representation of a triplicate analysis of the band 2 considering the variability of different technical steps. **Panel A:** Comparison of a triplicate injection of Band 2 obtained from a BBM pool preparation. **Panel B:** Triplicate comparison of Band 2 obtained from a BBM pool preparation loaded in adjacent lanes of the same gel. **Panel C:** Triplicate comparison of Band 2 obtained from a BBM pool preparation loaded in three different gels. **Panel D**: Triplicate comparison of Band 2 obtained from the three BBM preparation. All LC-MS technical steps were kept constant. **Panel E:** Triplicate comparison of Band 2 obtained from the three BBM preparations taking into account all technical variability (e.g. including LC-MS) into account.*

bands were excised, while panel D's Venn diagram is showing the variability between the three BBM preparations since the samples were injected sequentially into the LC-MS system and all LC-MS technical steps were kept constant. The Venn diagram in Panel E reflects the variability observed for the proteins identified in Band 2 for the three BBM preparations when the samples were analyzed in a random fashion that is, taking into account all the technical variability of all the steps. Figure 4.23 and 4.24 show the Venn diagrams for the corresponding analysis of band 9 and 11.

Figure 4.23: *Venn diagram representation of a triplicate analysis of the band 9 considering the variability of different technical steps. Panel A: Comparison of a triplicate injection of Band 9 obtained from a BBM pool preparation. Panel B: Triplicate comparison of Band 9 obtained from a BBM pool preparation loaded in adjacent lanes of the same gel. Panel C: Triplicate comparison of Band 9 obtained from a BBM pool preparation loaded in three different gels. Panel D: Triplicate comparison of Band 9 obtained from the three BBM preparation. All LC-MS technical steps were kept constant. Panel E: Triplicate comparison of Band 9 obtained from the three BBM preparations taking into account all technical variability (e.g. including LC-MS) into account.*

Estimation of variation for the band 11

A

Three runs on the same gel

SameGel_pool_wBBM_1_11
SameGel_pool_wBBM_2_11
SameGel_pool_wBBM_3_11

B

Three runs on a different gel

Difgel_pool_wBBM_A_11
Difgel_pool_wBBM_B_11
Difgel_pool_wBBM_C_11

C

Preparation variation

PV_wBBM_1_11
PV_wBBM_2_11
PV_wBBM_3_11

D
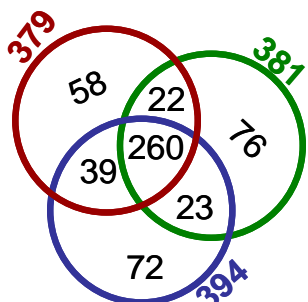
Total analysis variation

TV_wBBM_1_11
TV_wBBM_2_11
TV_wBBM_3_11
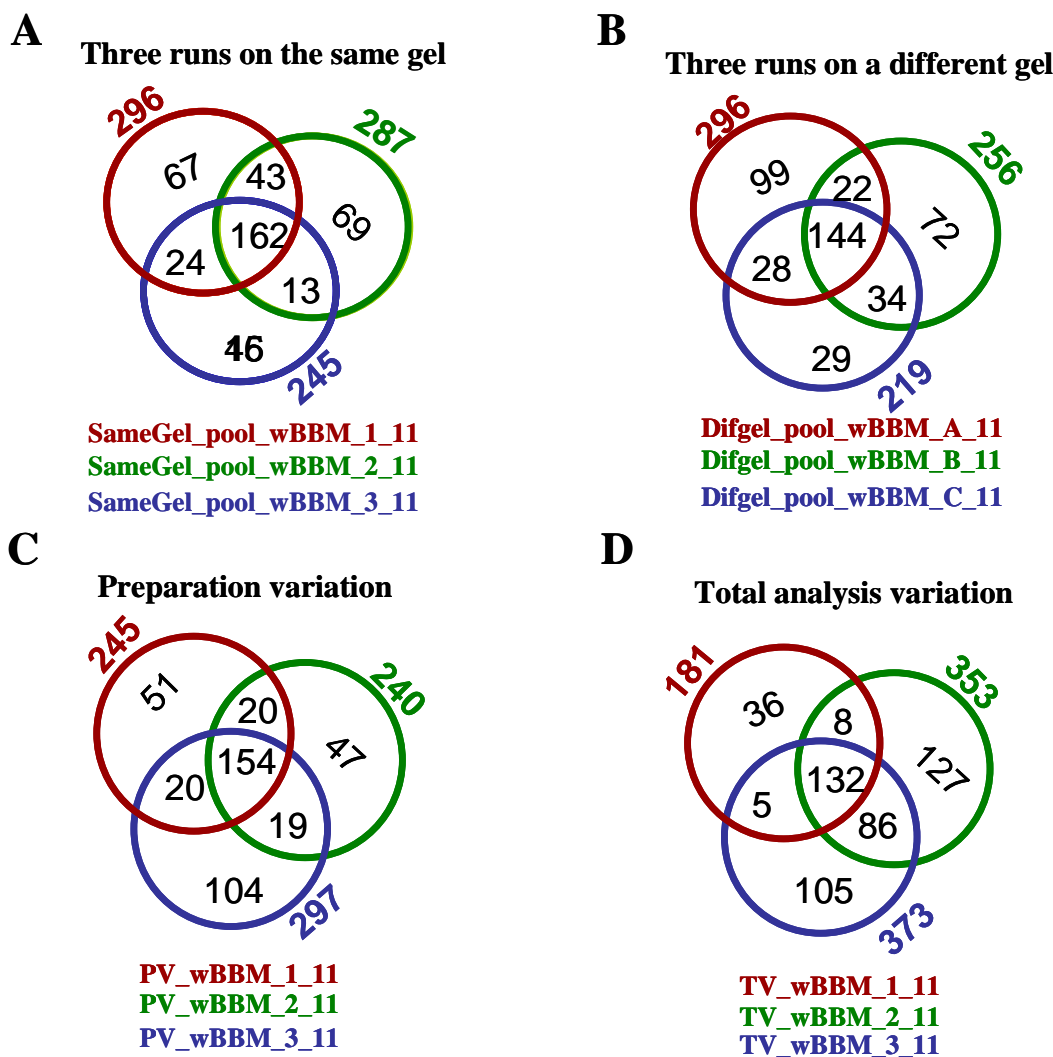
Figure 4.24: *Venn diagram representation of a triplicate analysis of the band 11 considering the variability of different technical steps. **Panel A:** Triplicate comparison of Band 11 obtained from a BBM pool preparation loaded in adjacent lanes of the same gel. **Panel B:** Triplicate comparison of Band 11 obtained from a BBM pool preparation loaded in three different gels. **Panel C**: Triplicate comparison of Band 11 obtained from the three BBM preparation. All LC-MS technical steps were kept constant. **Panel D:** Triplicate comparison of Band 11 obtained from the three BBM preparations taking into account all technical variability (e.g. including LC-MS) into account.*

The number of commonly identified proteins in bands 2, 9 or 11 across the different technical steps provides a first overview about the variability of each technical step. A graphic representation of the commonly identified proteins in two or three replicates, for each band and for each technical step is shown in Fig. 4.25.

Based on these results, it is apparent (and reinsuring) that the injection replicates show the highest reproducibility in respect to the number of commonly identified proteins, that is, the set-up and the operation of the LC-MS system was appropriate and reproducible for the number of proteins and the dynamic range expected for a 1D-SDS-PAGE band. Interestingly, the manner by which gel bands were cut (horizontally or one lane after the other) added some variability but the overall effect at the identification level was surprisingly moderate, indicating that, if proper care is taken, analysis run on several 1D-SDS gels are comparable. Further, the overall BBM preparation protocol (including a precipitation stage, several low/high spin centrifugation steps and many washes) appeared to be very stable and reproducible since the variability on the number of commonly identified proteins was in the same range as what was obtained with the BBM pool preparation loaded on the different gel variation and quite close to injection replicates. Finally, the comparison of panel E with panels C and D allow to draw some conclusions about the importance of a proper design of experiment to restrict unwanted variability. The major difference between those two conditions was that the LC-MS analysis performed for the panels C and D were purposely done using the same column and same buffer batch within the same sample measurement, whereas the LC-MS analysis performed for the panel E was done in a random order with sometimes column and buffer changes between the samples to be compared. Thus, the relative total variability increase observed for band 2 (see figure 4.22, panel E) was less than observed within band 9 (see figure 4.23, panel E) or 11 (see figure 4.24, panel D) because the three samples of band 2 were randomly analyzed however using the same column (additional variability might have been possibly due to the column history). The total variability for the band 9 was still moderate since there was only column change between the analyzed samples while the variability for band 11 was much higher, probably due to change of column and buffer between injections of the samples. This particular topic will be investigated more extensively in the next section. Interestingly, while band 2 was much more intense than bands 11 and 9 (the band 9 was very faint), the number of proteins identified in band 9 was double compared to band 2. This might be due in part to ion suppression due to very high abundant proteins in the band 2 (The instrument's duty cycle was limiting) and/or it might point to the presence of proteolytic fragments in this lower molecular mass band.
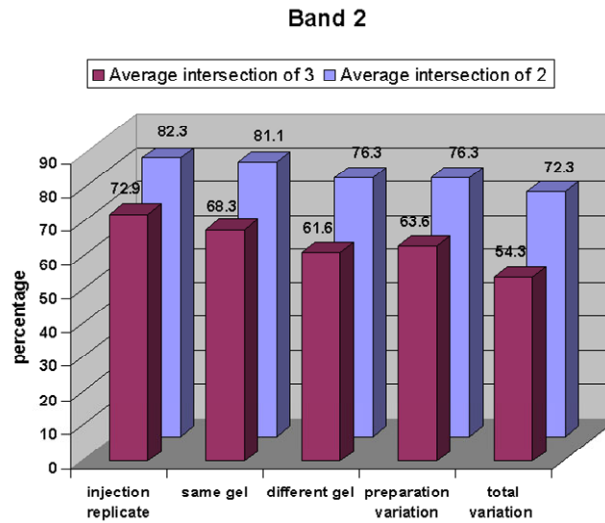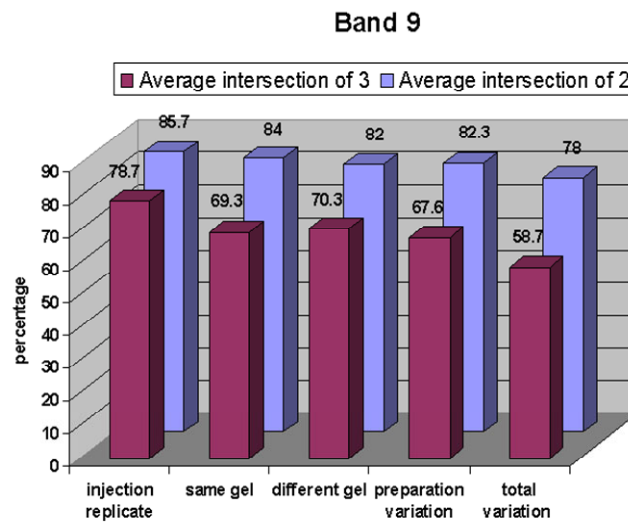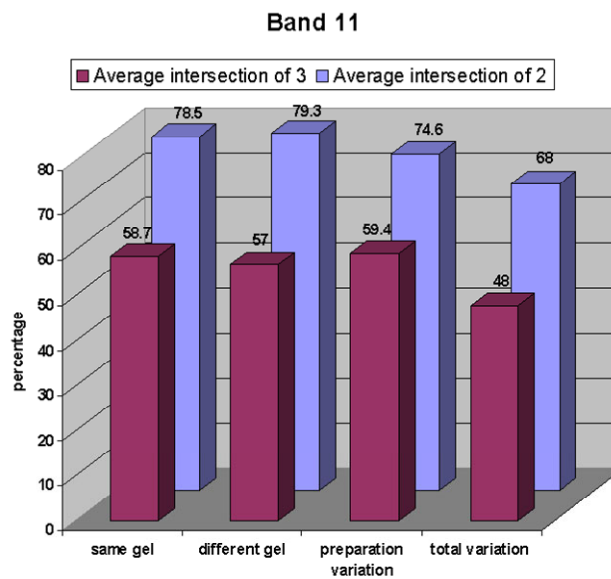
Figure 4.25: *Graphic representation of Average intersections of 2 and 3 samples in each technical step for band 2 (Panel A), band 9 (Panel B) and band 11 (Panel C).*

4.5.3.2 Estimation of experimental reproducibility based on LC-MS signals

Comparison of samples based on protein identification is a very well known and accepted practice in a proteomics experiment. Its main drawback lies in the fact that protein identification is based on properly annotated and identified MS/MS data corresponding to at most 10-20% of the overall mass spectrometric information due to the duty cycle and the sensitivity of the instrument. In other words, 80% of the generated information cannot be directly used, because either the MS/MS spectrum could not be successfully assigned to a peptide sequence or because the parent mass observed in the full MS spectrum was not selected for tandem MS analysis.

Recently, several groups (including ours) have started to consider the precursor ion intensity in the full spectrum as a more comprehensive descriptor for sample comparison. In this spirit, the data analysis performed at the peptide identification level was extended to include the MS precursor level. The variability of the BBM sample preparation and the influence of various technical parameters on the comparability of the samples were assessed to study how these can be directly derived at the MS level. Furthermore, the question on how much the information at the MS and the MS/MS levels corresponded was also explored.

4.5.3.2.1 Comparative analysis of a standard peptide mixture

In a first approach, a simple standard peptide mixture of 6 digested proteins Cytochrome C, Lysozyme, Alcohol dehydrogenase, Bovine serum albumin, Serotransferrin and ß-Galactosidase, (the so-called LCP Dionex peptide mixture) was used to tune the data processing scheme and to become acquainted with the data behavior in ideal settings to understand how to judge (dis)similarity in this context. The simplistic nature of this sample makes it significantly easier to monitor the dilution effect and column or buffer variability. In a second step, the key learning from this idealized study was applied to a more complicated dataset, such as the bands 2, 9 and 11 that have been discussed in the preceding section.

The raw data of each LC-MS run was processed according to the parameters described in § 3.2.7.3. Filtered signals that were present in 2/3 of the runs analyzed were accepted as common signals. Each of those MS signals was associated with a run-specific RT and accurate mass to allow the pairwise comparison of the processed MS signals between two samples in the form of scatter plots. The linearity and data distribution at the diagonal (that is, the intensity deviation of the common signals from an ideal linear relationship) was expressed using the Spearman correlation. Spearman's rank correlation coefficient is a non-parametric measure of correlation – that is, it assesses how well an arbitrary monotonic function could

describe the relationship between two variables, without making any assumptions about the frequency distribution of the variables. In the case of perfect reproducibility, we expect all the data points to be located in a diagonal, which gives a correlation of 1.0.

Fig. 4.26 shows the scatter plots representative for the comparison of three injection triplicates of a 50 fmol standard peptide mixture.

Injection replicates of a Standard Peptide Mixture
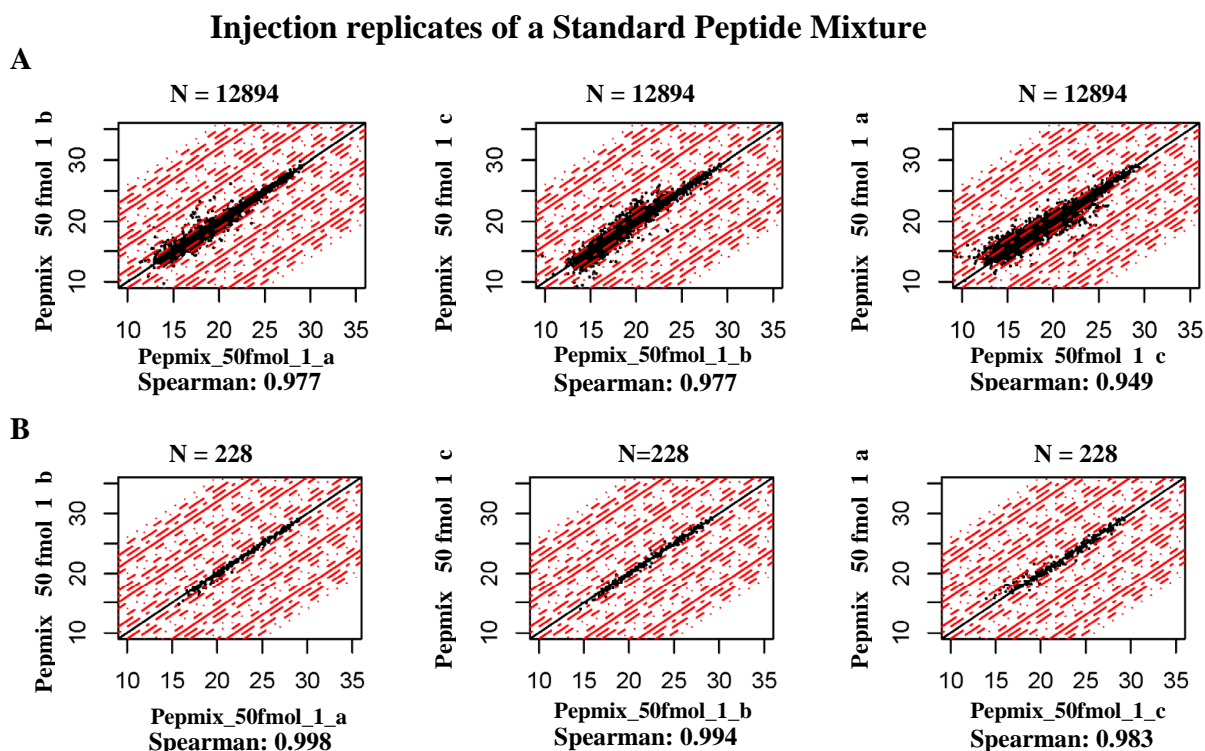
Figure 4.26: Scatter plot representations of three injection replicates of a 50 fmol Standard Peptide Mixture. Panel A: Scatter plots of all common MS signals found in the three replicates. Panel B: Scatter plots of common MS signals which have been successfully assigned to a MS/MS identification. N states for the number of common signals. The Spearman value reflects the similarity of the signal intensities between the compared samples (ideal case Spearman correlation=1). Each red dot line parallel to the diagonal represents a two fold difference, i.e. across three lines, the fold change is 2*2*2=8. Accordingly, the axes represent an arbitrary mass spectrometric intensity (in counts) in log2 units.

The scatter plots for each compared LC-MS(/MS) pair, based on the total MS signals (Panel A) or the successful MS/MS measurements (Panel B), highly correlated with each other. The clustering of the common features along the diagonal indicates that the precursor MS ions have the same intensity in the two considered samples, a feature to be expected in this experiment where replicates injections of the same sample were analyzed. Further, the Spearman correlation values for both types of analyses were almost identical, strengthening our conviction that a comparison of samples based on the precursor mass intensity should provide comparable results as using the protein identification descriptor.

Scatter plots as shown in Fig. 4.26 could also be used to visualize dilution effects. For example, as shown in Fig. 4.27, the precursor mass signal intensity of a 50 fmol standard peptide mixture is compared to the precursor mass signal intensity of a 125 fmol injection of the same peptide mixture.
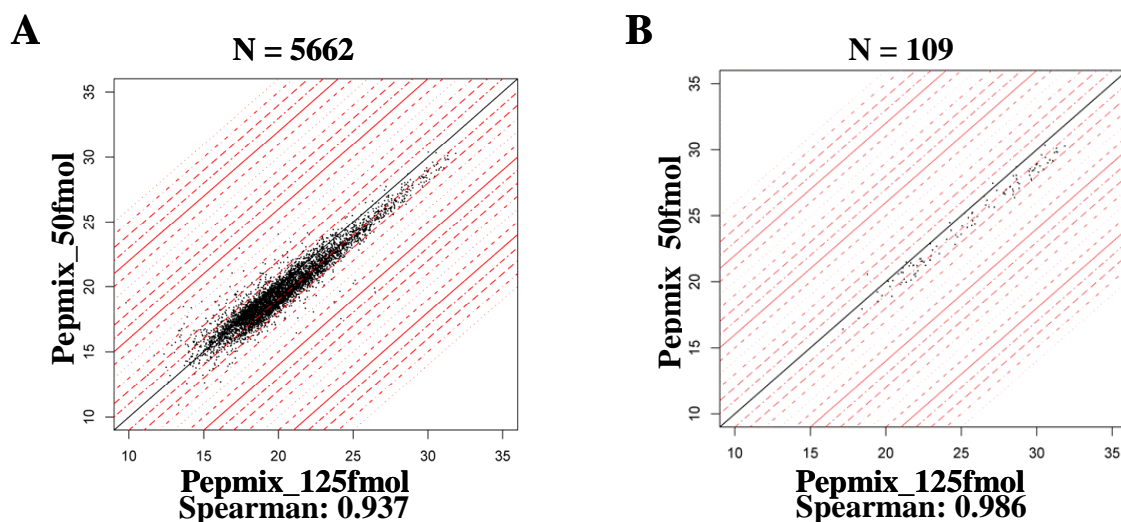
Figure 4.27: *Scatter plot representations of a comparison between a 125 fmol and a 50 fmol injection of a Standard Peptide Mixture. Panel A: Scatter plots of all the common MS signals between the samples. Panel B: Scatter plots of the MS Signals that correspond to successful MS/MS measurements. N is the number of common signals. The Spearman value reflects the similarity of the signal intensities between the compared samples (ideal case Spearman correlation=1). Each red dot line parallel to the diagonal represensts a two fold difference, i.e. across three lines, the fold change is 2*2*2=8. Accordingly, the axes represent an arbitrary mass spectrometric intensity (in counts) in log2 units.*

The 109 common precursor mass intensities based on MS/MS identification (Fig. 4.27, panel B) all clearly lied on a diagonal with an offset of approximately 2.5 fold difference from unity, as was expected from the experiment design. A similar pattern could also be observed for the 5662 common precursor masses found in both samples (Fig. 4.27, panel A) with the exception that two ions populations were clearly detected. The first population, comprising the most abundant ions of the analysis, were lying on a diagonal with an offset of approximately 2.5 fold difference similarly to the ions that were identified through a MS/MS identification. This first population was assumed to represent actual peptides that were differentially detected in the analysis. The second ion population, which included mostly low abundant signal, clustered along the diagonal, and could represent the consistent chemical noise that was co-analyzed with the samples.

More subtle effects, such as column ageing, could also be monitored using the same strategy. The effect of column history on MS signal intensity was put in evidence by following the

pattern of a 50 fmol standard peptide mixture analyzed in a new column and then again using the same column and LC buffers after several samples injections (Fig. 4.28).
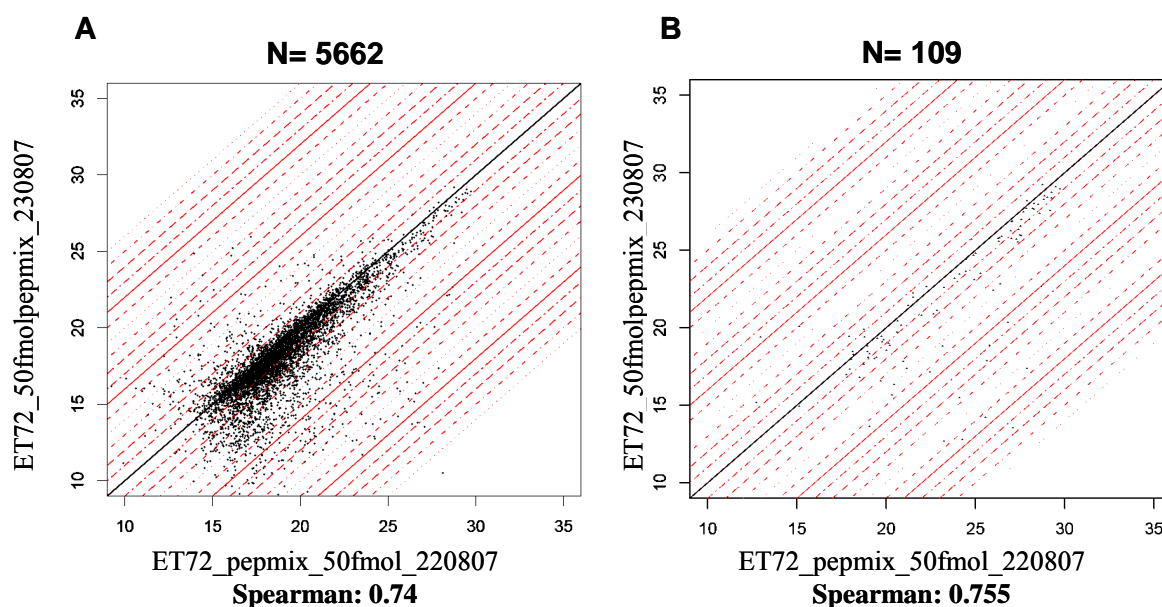
Figure 4.28: *Scatter plot representations of two 50 fmol injections of a Standard Peptide Mixture, injected onto the same column with the interval of several samples. Panel A: Scatter plots of all the common MS signals between the samples. **Panel B:** Scatter plots of the MS Signals that correspond to successful MS/MS measurements. N is the number of common signals. The Spearman value reflects the similarity of the signal intensities between the compared samples (ideal case Spearman correlation=1). Each red dot line parallel to the diagonal represents a two fold difference, i.e. across three lines, the fold change is 2*2*2=8. Accordingly, the axes represent an arbitrary mass spectrometric intensity (in counts) in log2 units.*

Not surprisingly, the common signals of the two identical peptide standards did not show the same intensity. MS signals were stronger when the fresh standard peptide mixture (ET_pepmix_50fmol_220807) was analyzed with a new column (for more details see appendix B: table of process variation time record). A Spearman correlation value of 0.74 for the total MS signals, or 0.755 for the MS signals with successful MS/MS identification, were significantly lower that for an ideal case, and demonstrated that column history might contribute to the experimental variability of comparable samples.

The impact of changing column and/or buffers during samples measurement was also investigated. The comparison of standard peptide mixtures measured using two identical columns (identical dimensions, lot number and sample history) showed that common MS signal intensities deviated somehow from the diagonal, also reflected in the lower Spearman correlation value (Fig. 4.29), but without dramatic changes in the data behavior.
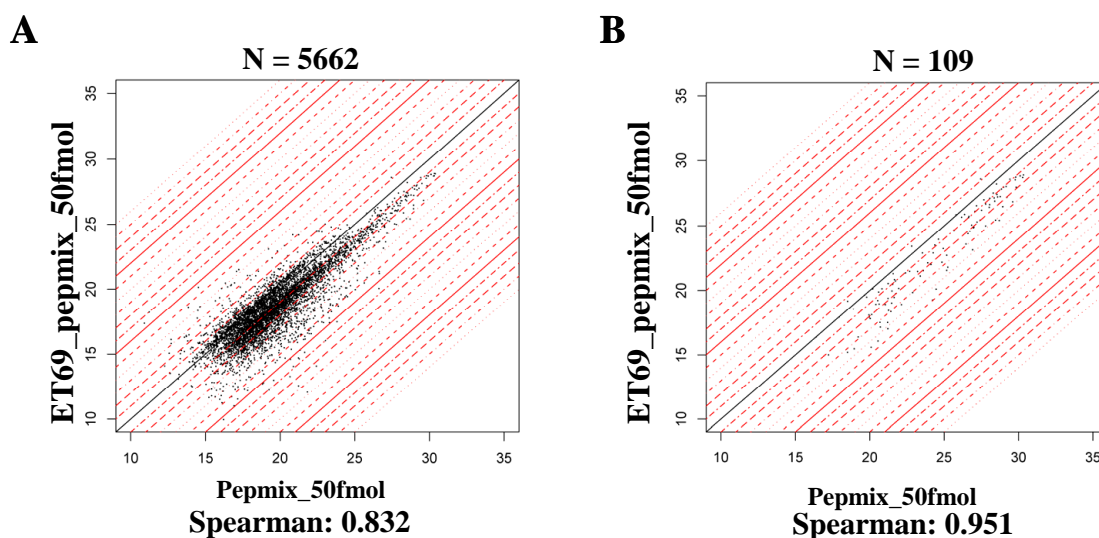
Figure 4.29: *Scatter plot representations of two 50 fmol injection of a Standard Peptide Mixture, injected in two identical columns with the same sample history. **Panel A:** Scatter plots of all the common MS signals between the samples. **Panel B:** Scatter plots of the MS signals that correspond to successful MS/MS measurements. N is the number of common signals. The Spearman value reflects the similarity of the signal intensities between the compared samples (ideal case Spearman correlation=1). Each red dot line parallel to the diagonal represents a two fold difference, i.e. across three lines, the fold change is 2*2*2=8. Accordingly, the axes represent an arbitrary mass spectrometric intensity (in counts) in log2 units.*

A similar data distribution was observed when the standard peptide mixture was analyzed on the same column but using different LC solvent batches (results not shown). However, these two effects acted synergistically when the standard peptide mixture was analyzed using different columns and different LC buffer batches (Fig. 4.30), as reflected by the lower Spearman correlation values. Interestingly, the column/buffer effects were consistently more pronounced at the global precursor MS level than at the common signals that were linked to a successful MS/MS measurement. However, the limited number of measurements that were performed using the standard peptide mixture did not allow to differentiate whether this difference in distribution was mostly due to a massive change of background ions distribution compared to the peptide signals, or whether the difference was due to minor mismatching of the precursor masses during sample analysis, leading to increased noise in the corresponding scatter plots.

A

N= 5662

Pepmix 50 fmol

ET72_50fmol_pepmix_230807
Spearman: 0.653

B

N= 109

Pepmix 50 fmol

ET72_50fmol_pepmix_230807
Spearman: 0.806
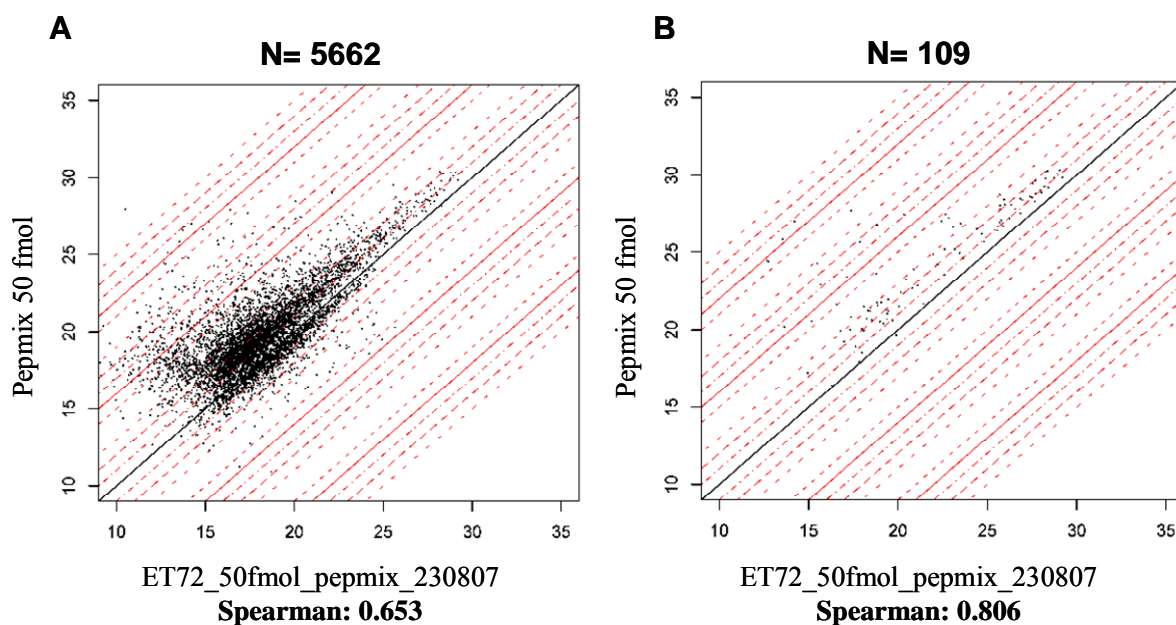
Figure 4.30: Scatter plot representations of two injections of a 50 fmol Standard Peptide Mixture onto two different columns and using different buffers batches. Panel A: Scatter plots of all the common MS signals between the samples. Panel B: Scatter plots of the MS signals that correspond to successful MS/MS measurements. N is the number of common signals. The Spearman value reflects the similarity of the signal intensities between the compared samples (ideal case Spearman correlation=1). Each red dot line parallel to the diagonal represents a two fold difference, i.e. across three lines, the fold change is 2*2*2=8. Accordingly, the axes represent an arbitrary mass spectrometric intensity (in counts) in log2 units.

In summary, the analysis of a simple standard peptide mixture using idealized differential conditions of the LC-MS system confirmed that the precursor ion intensity embodies an appropriate descriptor to evaluate the similarity of a sample to another and to pinpoint to common experimental deviations, such as column and LC buffer changes, or dilution effects. In particular, the behavior of all the ions considered in a pair of samples was very comparable to the ion population that was characterized by tandem mass spectrometry to represent the mass spectrometric signals of peptides commonly shared by the two samples considered.

The quality of this similarity was expressed using the Spearman correlation value, a non-parametric function to evaluate the degree of correlation between two parented ion intensity population to follow an arbitrary monotonic function, here a simple linear function of slope x=1. It is of interest that the Spearman correlation value was calculated here taking into account all the ions considered. However, the graphic representation of the ion distribution in the form of scatter plot clearly separated two populations of ions. The first group mostly clustered at the diagonal independently of the samples being compared and tended to encompass the lower intensity ions. The second group, which included mostly the higher intensity ions, also clustered along a diagonal but with a distinct offset from the first group

depending on the type of samples being compared. It was our belief that these two groups represented the solvent and LC contaminants ions and the sample peptide ions, respectively, which should be considered separately in an ulterior version of this similarity measure.

4.5.3.2.2 Comparative analysis of the BBM SDS gel bands 2, 9 and 11.

The quality and the reproducibility of the BBM preparation was assessed at the precursor ion level on the basis of the analysis of the selected gel bands 2, 9 and 11. The goal of this analysis was two-fold: firstly, to explore whether a differential analysis based on precursor mass intensities would align with the findings derived from the comparison of the same samples based on protein identification; and secondly, to evaluate to which extent the (ir)reproducibility of the LC-MS system could be observed in a more complex protein mixture such as used for the BBM identification study.

As already mentioned in section 4.5.2, all comparisons were performed using three representative SDS-PAGE bands, Bands 2, 9, and 11. Band 2 represented a very abundant, very well defined high molecular mass band (expectation: excellent reproducibility) while Band 9 represented a diffuse, very faint band (expectation: not so good reproducibility). Band 11 also displayed a well defined band but differed from Band 2 by the fact that the proteins present in the mass range of the gel might also encompass proteolytic products (expectation: average reproducibility). For simplicity reason, only representative scatter plots of chosen sample comparisons will be shown. The complete set of scatter plots and Spearman correlation values for all comparisons can be found in appendix B3.

Fig. 4.31 shows the comparison scatter plots for the three injection replicates of the band 2. The three replicates were very reproducible according to the scatter plot representation of all common MS signals and of the MS signals that correspond to successful MS/MS measurement. The Spearman correlation values for all comparisons were above 0.98 underlining the very good linearity of the signals at the diagonal. Similarly to what was observed with the simple Dionex protein mixture, the MS signals which were successfully assigned a peptide sequence through MS/MS analysis were in large of high signal intensity (panel B). However, this ion population represented less than 10% of the overall common MS signals (compare N between panel A and B).
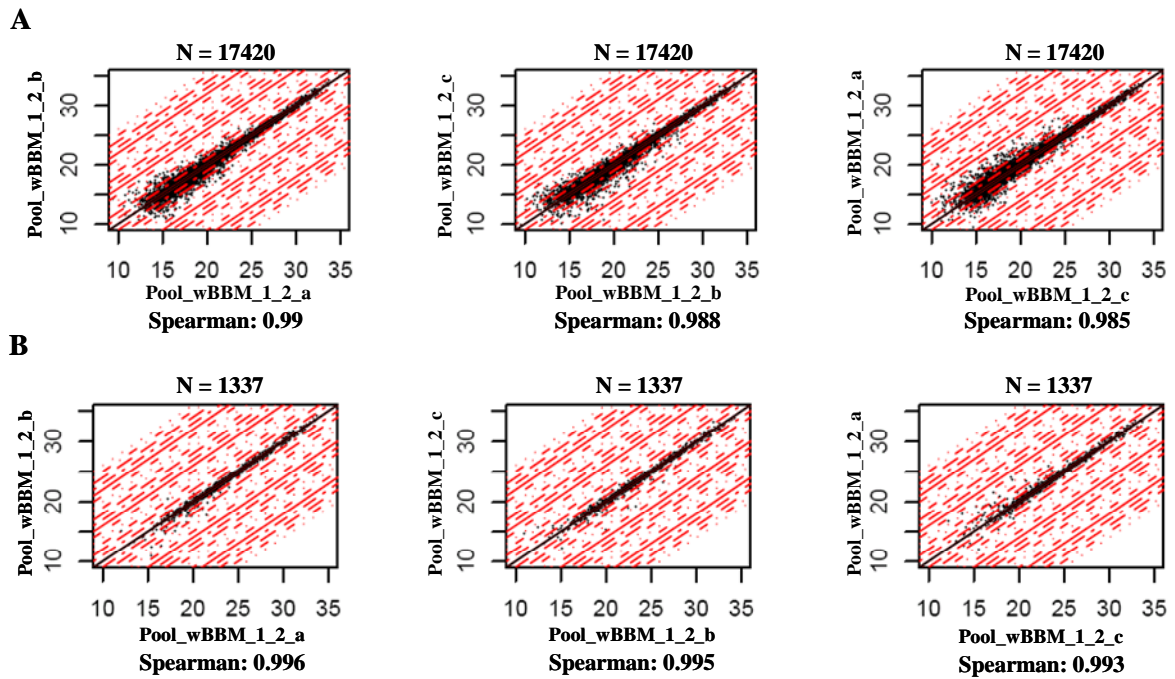
Figure 4.31: *Scatter plot representations of the injection replicates of the band 2. Panel A: Scatter plots of all the common MS signals between the replicates. Panel B: Scatter plots of the MS signals that correspond to successful MS/MS measurements. N is the number of common signals. The Spearman value reflects the similarity of the signal intensities between the compared samples (ideal case Spearman correlation=1). Each red dot line parallel to the diagonal represents a two fold difference, i.e. across three lines, the fold change is 2*2*2=8. Accordingly, the axes represent an arbitrary mass spectrometric intensity (in counts) in log2 units. All sample denominations are as described in Fig. 4.21.*

Further, the comparison scatter plots for a triplicate analysis of the band 2 excised from three adjacent SDS-PAGE lanes is shown in Fig. 4.32. The added variability of this technical step compared to the triplicate injection replicate was rather moderate as judged by the uniform and only slightly lower Spearman correlation values compared to the injection replicates. The main feature differentiating those two analyses were a broadening of the ion distribution along the diagonal, especially noticeable at the lower ion intensity scale, and a ion correlation matrix that did not include the least intense ion, probably because of lack of reproducibility (compare Fig. 4.31 and Fig. 4.32 panels A and panels B side by side). Interestingly, within the triplicate analysis included in Fig. 4.32, the samples pool_wBBM_2_2 and pool_wBBM_3_2 appeared to correlate slightly better to each other than with the sample pool_wBBM_2_1, a finding that was not observed from the Venn diagrams analysis of the same samples (compare with Fig. 4.22, panel B).

Same gel variation for band 2
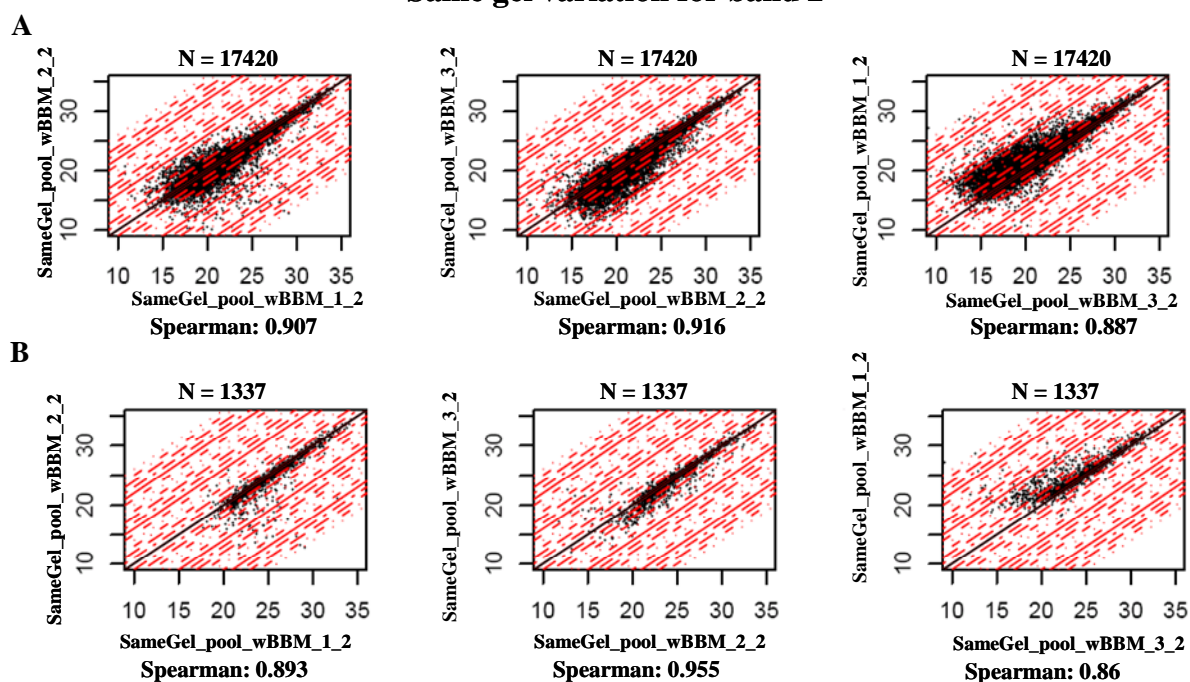
Figure 4.32: Scatter plot representations of the variability for the band 2, cut horizontally from adjacent identical lanes. Panel A: Scatter plots of all the common MS signals between the replicates. Panel B: Scatter plots of the MS signals that correspond to successful MS/MS measurements. N is the number of common signals. The Spearman value reflects the similarity of the signal intensities between the compared samples (ideal case Spearman correlation=1). Each red dot line parallel to the diagonal represents a two fold difference, i.e. across three lines, the fold change is 2*2*2=8. Accordingly, the axes represent an arbitrary mass spectrometric intensity (in counts) in log2 units. All sample denominations are as described in Fig. 4.21.

Similarly, the comparison scatter plots for a triplicate analysis of the band 2 excised from a lane run in three separate SDS-PAGE gels is shown in Fig. 4.33. As expected from the previous analysis, the added variability of this technical step was very comparable to the variability obtained from the triplicate bands excised from a single SDS-PAGE gel. The measurement reproducibility depended mostly on how accurately the bands were defined and excised from the SDS-PAGE gel, the type (wide/narrow; defined/diffuse, etc.) of bands that was analyzed, and how identically the gels to be compared had run. Interestingly, in this triplicate analysis shown in Fig. 4.33, the samples pool_wBBM_A2 and pool_wBBM_B2 appeared to correlate slightly better to each other than with the samples pool_wBBM_C2, a finding that was not observed from the Venn diagrams analysis of the same samples (compare with Fig. 4.22, panel C). Taken together, the results obtained in Fig. 4.32 and 4.33 suggest that the SDS-PAGE step appeared non-critical for the reproducibility of an experiment as long as the parameters for protein separation and band excision were kept tightly controlled.

Different gel variation for band 2
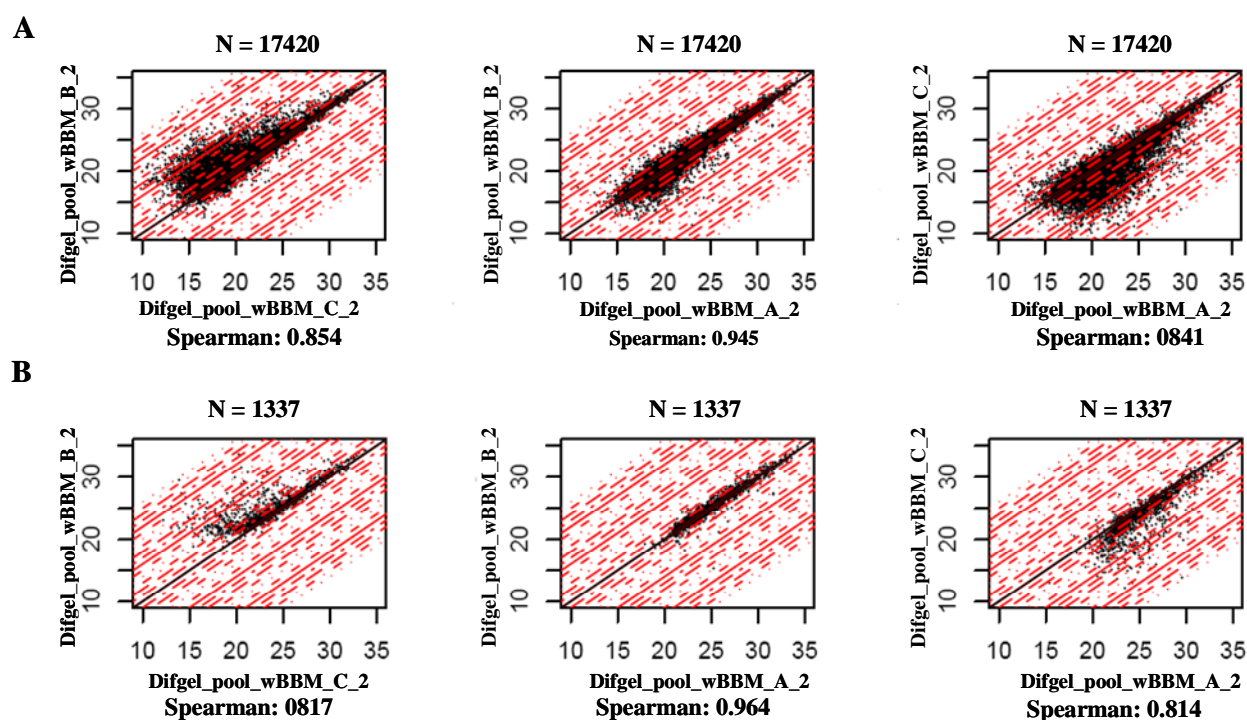
Figure 4.33: *Scatter plot representations of the band 2 from identical BBM samples, loaded in three different gels. Panel A: Scatter plots of all the common MS signals between the replicates. Panel B: Scatter plots of the MS signals that correspond to successful MS/MS measurements. N is the number of common signals. The Spearman value reflects the similarity of the signal intensities between the compared samples (ideal case Spearman correlation=1). Each red dot line parallel to the diagonal represents a two fold difference, i.e. across three lines, the fold change is 2*2*2=8. Accordingly, the axes represent an arbitrary mass spectrometric intensity (in counts) in log2 units. All sample denominations are as described in Fig. 4.21.*

The comparison scatter plots for the LC-MS signals of the band 2 originating from three different technical BBM preparations is shown in Fig. 4.34. Since all technical steps (samples run on the same SDS-PAGE gel, band excision, digestion, column and buffer batch, LC-MS conditions) were maintained constant, the observed variability for the common signals should mainly reflect the reproducibility of the BBM preparation. The common MS signals (Panel A) and the common MS signals that were assigned to a successful MS/MS identification (Panel B) were clustered at the diagonal indicating a high level of similarity between the three BBM preparations. This observation was supported by a high Spearman correlation value (above 0.94) for all the comparisons. This high degree of reproducibility was also noted for the preparation variation analysis of bands 9 and 11 (results not shown, for more details see appendix B) and correlated with the findings obtained from the protein identification level (see Fig. 4.22, panel E, Fig. 4.23, panel E and Fig. 4.24, panel D). These results demonstrate

that the BBM preparation protocol could be performed in a very stable and reproducible manner that added only a minimal variability compared to other technical steps.

Preparation variation for band 2
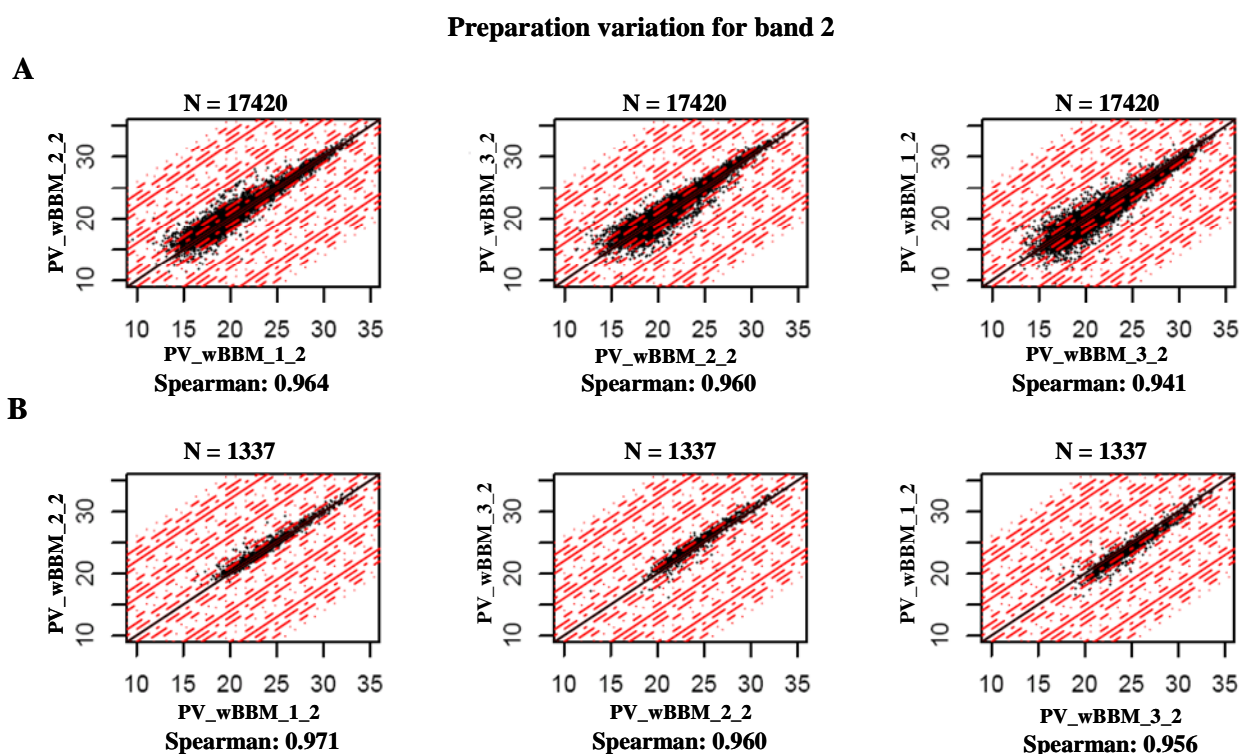
Figure 4.34: *Scatter plot representations of the preparation variability for the band 2 from three BBM technical replicate preparations. Panel A: Scatter plots of all the common MS signals between the replicates. Panel B: Scatter plots of the MS signals that correspond to successful MS/MS measurements. N is the number of common signals. The Spearman value reflects the similarity of the signal intensities between the compared samples (ideal case Spearman correlation=1). Each red dot line parallel to the diagonal represents a two fold difference, i.e. across three lines, the fold change is 2*2*2=8. Accordingly, the axes represent an arbitrary mass spectrometric intensity (in counts) in log2 units. All sample denominations are as described in Fig. 4.21.*

Finally, the comparison scatter plots for the LC-MS signals of the band 2, 9 and 11 originating from three different technical BBM preparations is shown in Fig. 4.35, 4.36 and 4.37, respectively. In this last set of analyses, however, the technical steps were not as tightly controlled and the samples were analyzed in random orders.

According to the scatter plot representation and the Spearman correlation values (Fig. 4.35), the variability observed for the band 2 was considerably lower than for the bands 9 and 11. In this experiment, variability must originate from the technical steps since the three BBM preparation were found to be highly reproducible in the previous section. In our experience, most

of the technical variability was due to different column history, column and LC buffer changes
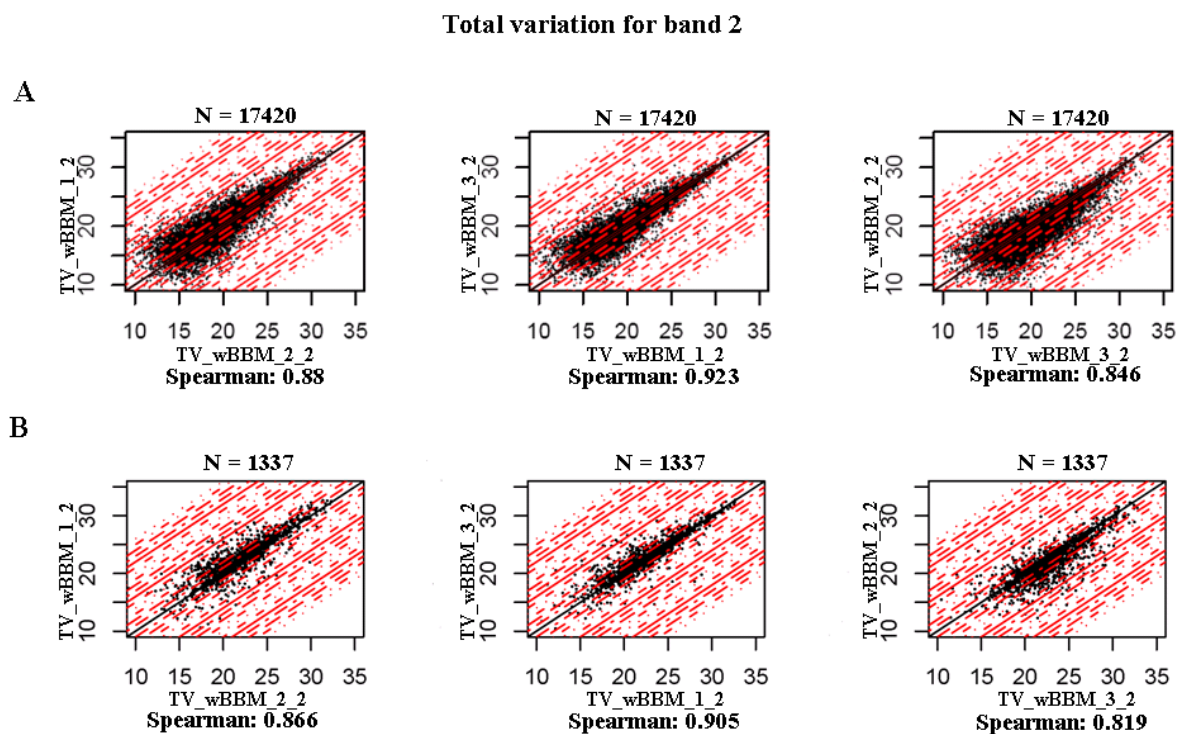
Figure 4.35: *Scatter plot representations of the total variability for the band 2 from three BBM technical replicate preparations, randomly analyzed. Panel A: Scatter plots of all the common MS signals between the replicates. Panel B: Scatter plots of the MS signals that correspond to successful MS/MS measurements. N is the number of common signals. The Spearman value reflects the similarity of the signal intensities between the compared samples (ideal case Spearman correlation=1). Each red dot line parallel to the diagonal represents a two fold difference, i.e. across three lines, the fold change is 2*2*2=8. Accordingly, the axes represent an arbitrary mass spectrometric intensity (in counts) in log2 units. All sample denominations are as described in Fig. 4.21.*

Indeed, the triplicates of the band 2 were randomly analyzed along the whole experiment but the column and LC buffer remained the same for the three analyses. The Spearman correlation value was slightly lower than was found for the preparation variation alone but remained in the same range. The slightly elevated variability could be due to the manner the gel bands were excised from the gel (one after the other versus horizontal parallel cutting) and to the LC column history. Thus, according to the Spearman correlation value, the samples TV_wBBM_3_2 and TV_wBBM_2_2 were significantly better correlated than with TV_wBBM_1_2 probably, because these two first samples were analyzed very closely to each other at the LC-MS level (only 2 samples separated the two band 2 replicates).

The variability observed for band 9 was significantly higher than that found for band 2 and was reflected in the scatter plots and the Spearman correlation values (Fig. 4.36). In our experience, this large increase in variability was due to column and buffer changes during the

analysis of the band 9 triplicates. Thus, the correlation of the replicates TV_wBBB_1_9 and TV_wBBM_3_9 was much higher than with TV_wBBM_2_9. TV_wBBB_1_9 and TV_wBBM_3_9 were analyzed on two different columns but using the same LC buffer batch, while TV_wBBM_2_9 was analyzed using a different column and another LC buffer batch. More details about the analysis sequence of all the samples included in this experiment can be found in the process variation measurement times in the appendix B.

Total variation for band 9
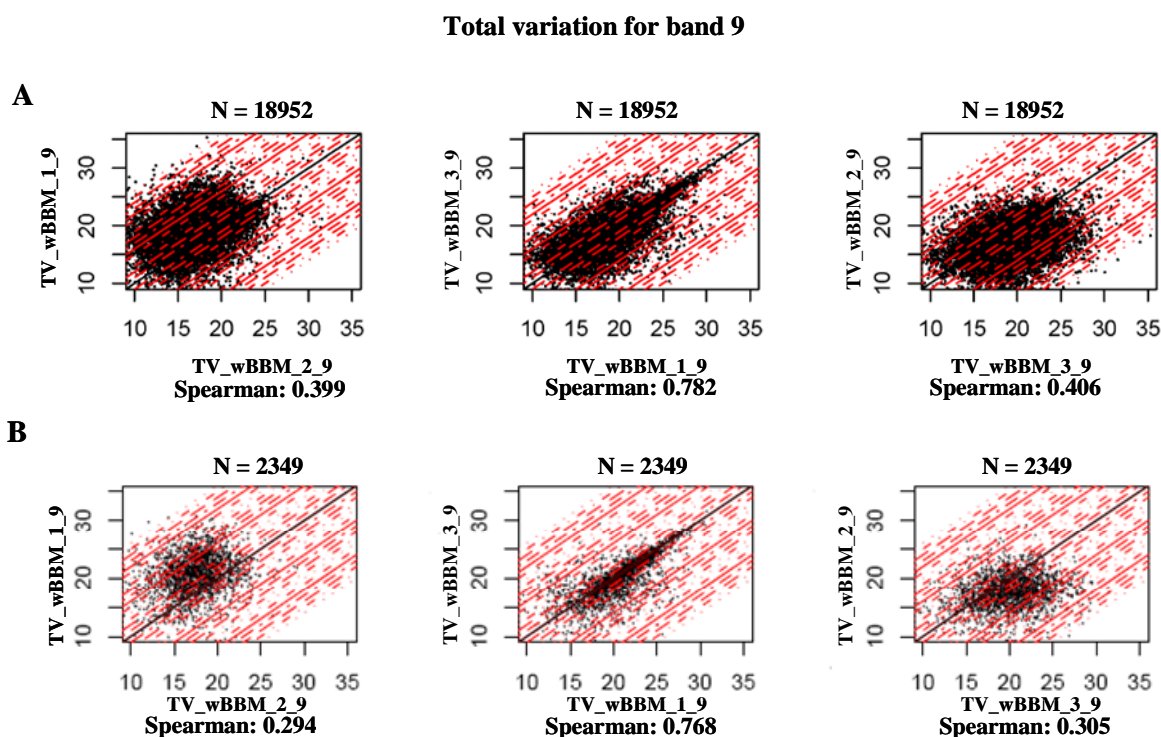
Figure 4.36: *Scatter plot representations of the total variability for the band 9 from three BBM technical replicate preparations, randomly analyzed. Panel A: Scatter plots of all the common MS signals between the replicates. Panel B: Scatter plots of the MS signals that correspond to successful MS/MS measurements. N is the number of common signals. The Spearman value reflects the similarity of the signal intensities between the compared samples (ideal case Spearman correlation=1). Each red dot line parallel to the diagonal represents a two fold difference, i.e. across three lines, the fold change is 2*2*2=8. Accordingly, the axes represent an arbitrary mass spectrometric intensity (in counts) in log2 units. All sample denominations are as described in Fig. 4.21.*

Column and buffer changes affected the analysis of band 11 in a similar fashion as seen in band 9 (see Fig. 4.37). Thus, the two replicates TV_wBBM_2_11 and TV_wBBM_3_11 clustered significantly better to each other than with the replicate TV_wBBM_1_11. Indeed, the replicates TV_wBBM_2_11 and TV_wBBM_3_11 were measured using different columns but the same LC solvent batch, while TV_wBBM_1_11 was analyzed using a different column and another LC buffer batch.

Interestingly, for all bands considered (but mostly for bands 9 and 11), the significantly lower comparability of the total variability analysis compared to the preparation variability analysis

was only partially observed at the protein identification level (see Fig. 4. 22, panel D and E; Fig. 4.23, panels D and E; and Fig. 4.24, panels C and D). The ion correlation matrix which
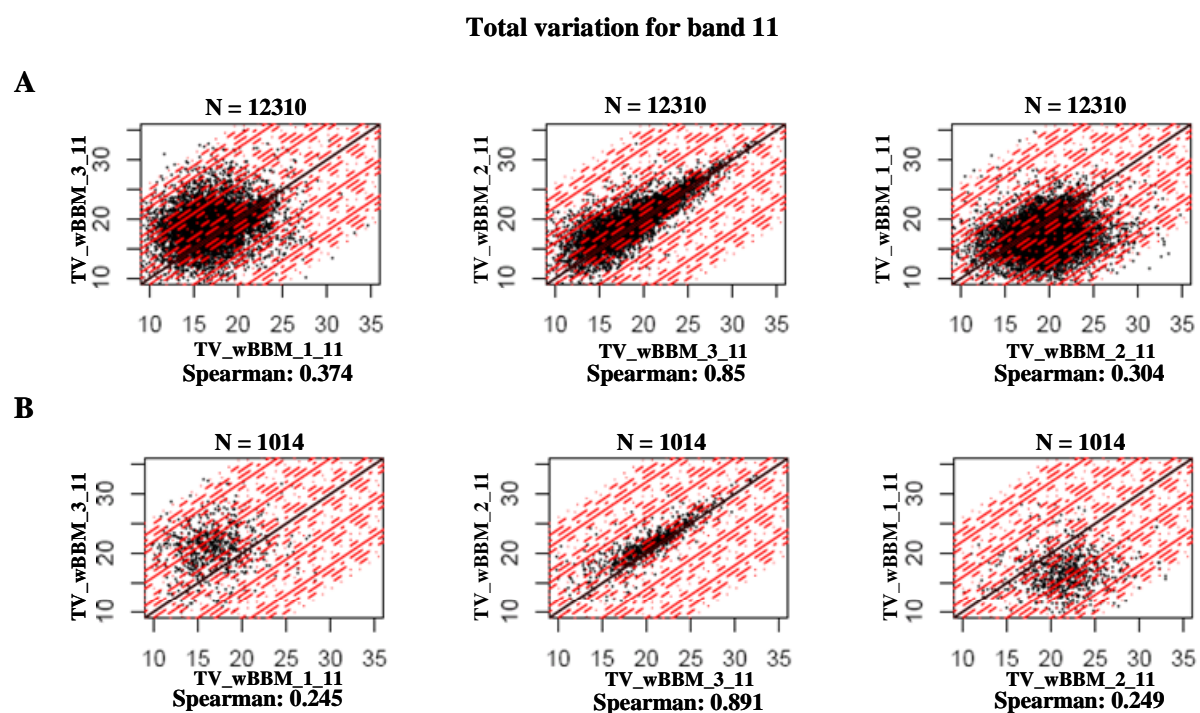
Figure 4.37: *Scatter plot representations of the total variability for the band 11 from three BBM technical replicate preparations, randomly analyzed. **Panel A:** Scatter plots of all the common MS signals between the replicates. **Panel B:** Scatter plots of the MS signals that correspond to successful MS/MS measurements. N is the number of common signals. The Spearman value reflects the similarity of the signal intensities between the compared samples (ideal case Spearman correlation=1). Each red dot line parallel to the diagonal represents a two fold difference, i.e. across three lines, the fold change is 2*2*2=8. Accordingly, the axes represent an arbitrary mass spectrometric intensity (in counts) in log2 units. All sample denominations are as described in Fig. 4.21.*

was used to create the scatter plots depends heavily on a reproducible RT and m/z parameters linked with the measured ion signal intensity, whose variability was minimized by measuring the samples to compare immediately one after another. Moreover, the correlation matrix took into account all MS signals that were commonly assigned to the samples, with a majority of them being rather of low intensity and subject to higher variability. In contrast, sample comparison by protein identification was based on a MS signal that was successfully assigned to an amino acid sequence through a MS/MS identification. Since this ion population comprised mostly well behaved, reproducible high intensity MS signal, the correlation matrix illustrated through Venn diagrams was much more resistant to intensity variability than for the common MS signals.

4.5.3.2.3 LC-MS reproducibility

The total variability analysis for the bands 9 and 11 showed that a major part of the observed technical variability was caused by the LC buffer followed by the column changes. Buffer changes, in particular, had a major impact on the peptides' retention times as shown in the figure 4.38 and 4.39.
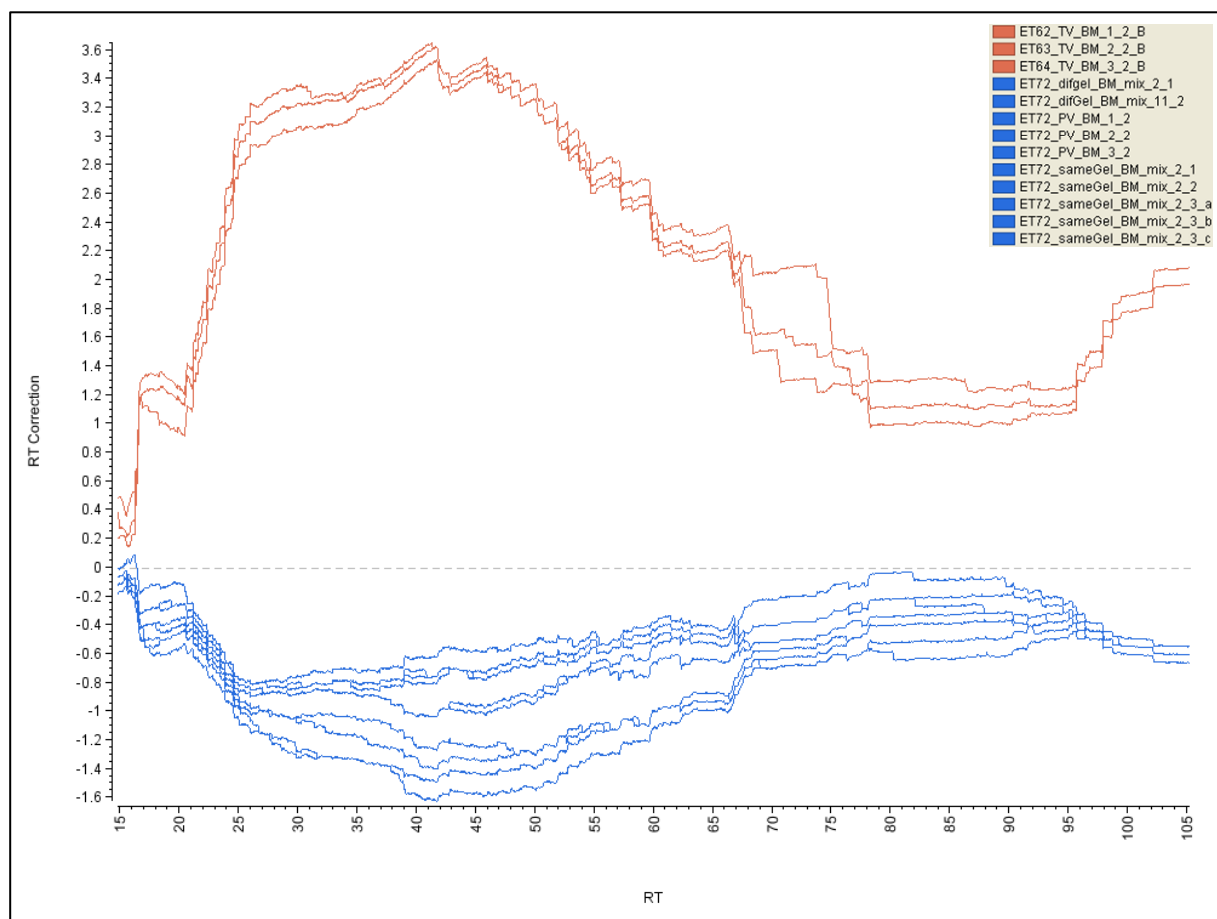
Figure 4.38: *Genedata Refiner MS signals representation of the retention time (RT) alignments of all the technical replicates related to band 2. The RT alignments of the total variation (TV) samples are marked in red while all other technical replicate samples are marked in blue. The two sample groups cluster according to the LC buffer and column batch they had been analyzed with, see text for additional details. The name conversion for the data files in the inset can be found in appendix B .*

The retention time alignment for the band 2 and 11 showed the obvious impact of LC buffer and column change on the MS signal alignment. In the case of band 2 (see figure 4.38) the RT alignment of the three "total variation" sample replicates (red traces) aligned very well with each other because these samples were analyzed using the same column and buffer batch. A similar good alignment was observed for all remaining LC-MS measurements (blue traces) of the band 2 as they were also measured using the same column and buffer batch. In this particular example, the rather different behaviors of both sample sets was due to the use of

different column and LC buffer batches, resulting in a less satisfying alignment between the common MS signals.

The picture was somehow different for the RT alignment of the band 11. The MS signals of the samples TV_wBBM_2_11 and TV_wBBM_3_11 showed very good RT alignments (Figure 4.39, red traces) to each other, although they were analyzed with two different columns. On the other hand, the RT coordinates from the third triplicate TV_wBBM_1_11 appeared to significantly deviate from the two other MS signals (see figure 4.39, green traces) due to a buffer change between the analysis of this sample and the other two replicates. Finally, all remaining LC-MS measurements (blue traces) of the band 11 were measured using the same column and buffer batch, thus showing a consistent RT alignment with each other.
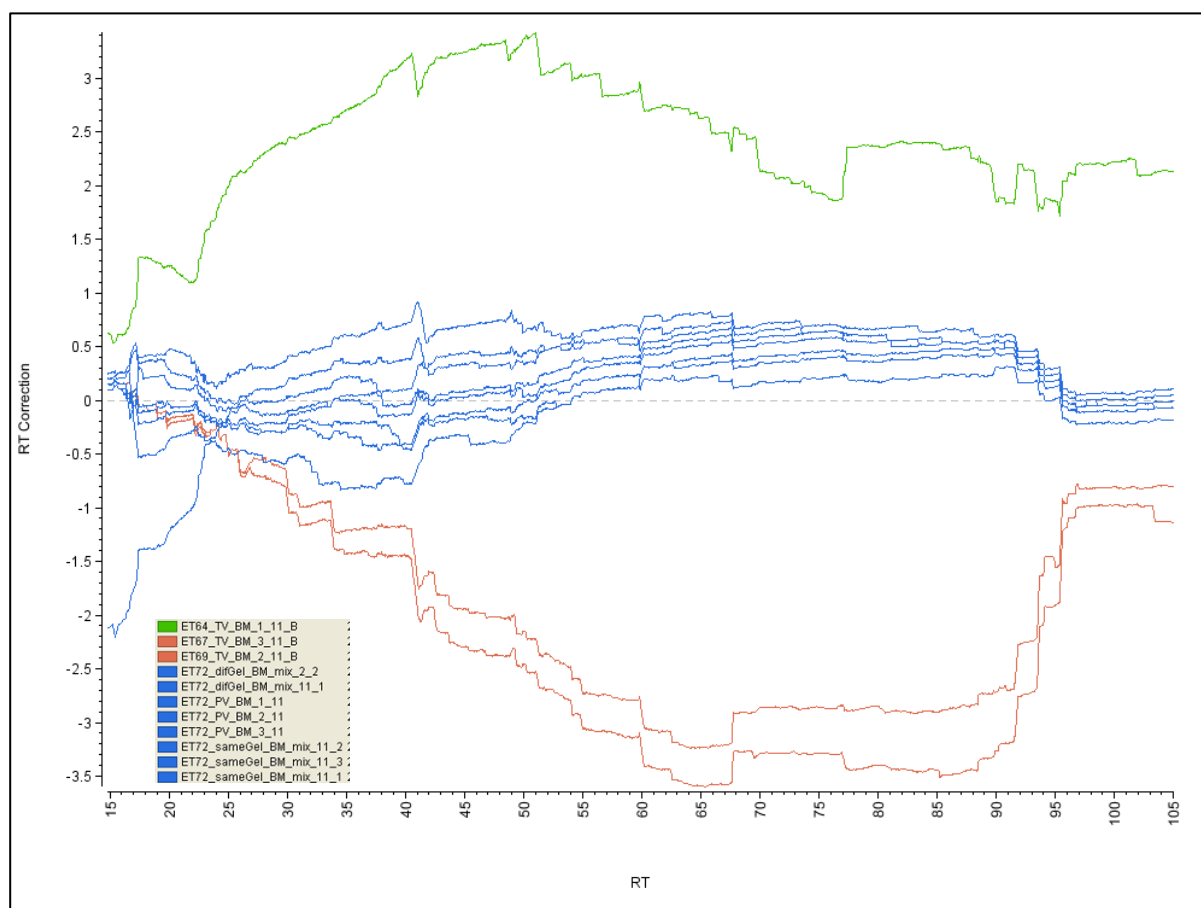
Figure 4.39: *Genedata Refiner MS signals representation of the retention times (RT) alignments of all the technical replicates related to band 11.* *The RT alignments of the total variation (TV) samples are marked in red and green while all other technical replicate samples are marked in blue. The three sample groups cluster according to the LC buffer and column batch they had been analyzed with, see text for additional details. The name conversion for the data files in the inset can be found in appendix B.*

4.5.3.3 Findings and discussion

Many researchers in the field have claimed that the LC-MS/MS approach lacks the technical reproducibility to reliably identify proteins in a proteomics experiment. In this study, we aimed to examine this common belief by setting up an experiment in which we evaluated the reproducibility of the LC-MS measurements by counting how many proteins could reliably be identified in the analysis of triplicate technical BBM preparations.

The triplicate LC-MS/MS measurements of the 19 gel bands corresponded to each BBM preparation (see § 4.2) resulted in almost 90% commonly identified proteins in at least two of three preparations with at least two different peptides in one of the reprlicates. This degree of reproducibility was much higher than what was optimistically expected in our group and better than what has been openly claimed so far by other proteomics groups.

This unexpected outcome led me to investigate in more details whether this very high level of identification reproducibility achieved in this study was not based on an over-optimistic interpretation of the data. For this reason I decided to study in a systematic manner the contribution in variability that selected technical steps might add in the overall analytical process. For this purpose, the gel bands 2, 9 and 11, and a standard peptide mixture were used as references to set up an experiment to monitor the variability of the LC-MS(/MS) system (mass spectrometer performance, column history and column or/and buffer change), the gel band excision process, and the variations due to the BBM isolation protocol.

Data analysis and interpretation proved to be a very hard and time consuming process, because there was no unique value to comprehensively report on the simultaneous comparison of three (or more) samples. One of the main reasons to adopt the Venn diagram representation was that the commonly identified proteins within the triplicate analysis could be easily visualized. Simultaneously, we sought to take advantage of the high mass accuracy provided by the Orbitrap mass spectrometer and the reproducibility of the liquid chromatographic system to evaluate, whether such a data analysis could also be performed using the precursor mass intensity.

Taken as a whole, the systematic evaluation of the technical variability added by selected technical steps to the overall analytical process unambiguously confirmed the high reproducibility achieved by the LC-MS/MS process at the protein identification level. As expected, sample comparison based on injection replicates showed the best reproducibility, stressing the stability and the robustness of the LC-MS/MS system. The strategy used to excise bands from SDS-PAGE gels (horizontally in adjacent lanes of the same gel, or one lane

cut after the other on different gels) was identified to potentially represent a major cause of variability if not appropriately controlled. An irreproducible gel excision pattern might cause a considerable variability in the protein content of a given gel band leading to inconsistent identifications when compared to the equivalent gel band of another sample.

The BBM isolation protocol itself was a potential source of variability due to its relative length and complexity and due to the inclusion of several steps that might have been difficult to carry out quantitatively, such as the $CaCl_2$ precipitation step. Also, the remaining proteolytic activity of the abundant BBM proteases, if unchecked, could also contribute to sample degradation and add extensive variability to the protein identification process. The preparation variation as mirrored in the comparison of the BBM technical triplicate analysis of bands 2, 9 and 11 showed a very high level of reproducibility in the number of commonly identified proteins (almost at the level of injection replicates) as long as all other technical steps (SDS-PAGE gel, band excision strategy, LC-MS/MS conditions) were kept constant. The variability in the number of commonly identified proteins was significantly higher when the samples were analyzed without a design of experiment. Thus, the number of proteins commonly identified in the three technical replicates of band 2 was rather similar in the preparation variation study (all technical steps were controlled) and in the total variation study (analysis without design of experiment) while the number of commonly identified proteins in the technical replicates of band 9 and 11 was significantly lower in the total variation study compared to the preparation variation study. One of the main reasons for this differences is believed to be due to column and buffer changes during the analysis of the bands 9 and 11 samples, while the analysis of the band 2 samples were performed using the same column and buffer batch.

The degree of variability added by each technical step and by the BBM preparation protocol was also investigated at the full MS signal level. The comparative analysis of the precursor ion signals might provide additional information on sample similarity, since data comparison at the protein identification level (using successful MS/MS signals) makes use of only 10% of the total available signals. On the other hand, the sample similarity information that was derived from the MS/MS-based comparison must also be detected at the precursor ion signal intensity level. Overall, the scatter plot representations of the total MS signals and of the subset of MS signals that were identified through a successful MS/MS analysis correlated well with each other across all compared samples. Moreover, the derived Spearman correlation value, which indicated the degree of similarity between two samples, showed the same trends as what was observed using the Venn diagram representations. Thus, the scatter

plots representations and the associated Spearman correlation values also confirmed the very high degree of reproducibility observed between sample injection replicates, as previously shown at the protein identification level. Likewise, the gel band excision scheme was also identified as a potential source of variability, if gel bands could not precisely be cut from the gel, as for example in the case of band 11. Finally, the preparation variation study, as reflected in the scatter plots and the Spearman correlation values for all the three bands, showed also a high degree of reproducibility, strong evidence that the BBM preparation protocol could be performed robustly and quantitatively, if all technical steps were controlled. An unexpected finding of this systematic study was the large impact of LC buffer change on the analytical variability. This effect was best illustrated in the total variation study of band 11, where in one of the three triplicates clearly deviated from the two other samples. The MS signal alignment of those three samples showed clearly the major shift of the peptides' retention time caused by the LC buffer change. In contrast, column change did not appear to contribute significantly to the overall technical variability, as minor column backpressure heterogeneities were compensated by the HPLC system's active flow splitter system (the LC system measures the actual flow going through the column, meaning that the LC column pressure is adjusted to keep the linear velocity constant through the column).

In summary, it is apparent that sample comparison based on protein identification or on precursor ion signals were inherently similar. However, it is obvious that the optimized use of the mass spectrometric information will heavily depend on the generation of reproducible experimental data following a standard operating procedure. As a general rule, samples that should be compared to each other should be analyzed with identical column and buffer batches. In cases where this is not feasible, it is very important that buffers and columns are reproducibly and accurately prepared. The repeated analysis of a standard complex peptide mixture along the experiment and after any changes of buffer or/and column allow to monitor the variation of the peptides' retention time and to judge the "good health" of the chromatographic system. However, even in an optimal experimental setting, comparison of serially acquired LC-MS(/MS) data will require the development of bioinformatics tools attuned to this specific data type. As highlighted in this study, one of the most immediate tasks to perform in a differential study is to measure the level of variability present in a given analysis, which in turn defines the criteria of acceptance for a reproducibly observed signal. The assessment on how to define such a signal, including the noise and variability introduced by the various technical steps of the analytical protocol, has been one of the most challenging

problem in data analysis of proteomics experiments, but also central to determine to which degree a signal difference can be confidently interpreted as a real change.

This study constitutes the first attempt of our group to compare biological samples using the precursor ion intensity information. In due course, this first step was to be followed by a full fledge proteomics study in which a much more comprehensive biological experiment could have been investigated, such as a control mouse vs statins vs ezetimibe treatment, knowing in advance the level of technical reproducibility that could be expected. Beside the obvious biological interest for such a study (the impact of those drugs in the BBM has not been investigated in detail), a long term goal of this experiment could have been to evaluate the feasibility of a label-free quantification scheme based on the common precursor ion signals. In this approach, every peptide signal within the sensitivity range of the MS analyzer can be extracted and incorporated into the quantification process independently of a MS/MS acquisition. In a first step, data acquisition is performed using a high resolution/high mass accuracy mass spectrometer using a stable chromatographic system to generate the most stable <m/z;RT;ion intensity> possible applying the analytical principles that this initial study helped to uncover. In a second step, the identity of the differentially regulated proteins is achieved using a targeted tandem mass spectrometry analysis approach using for example inclusion lists in selected LC-MS runs (78, 117, 118).

4.6 Peptide identification by LC-MS/MS

The reliable identification of membrane proteins in complex biological mixtures has always been a grey area in the proteomics field. As exemplified in this study, the contamination of even small amounts of cytosolic proteins in membrane preparations might significantly impede the identification of membrane proteins, especially if they are present in low abundance. The difficulties to obtain a high coverage of membrane proteins in biological preparations are several-fold: first, the biochemical enrichment of membrane proteins typically requires an extensive knowledge of the cell compartment to isolate, so to improve the yield in membrane proteins. However, optimized isolation protocols are not easily adjustable for other biochemical applications, so that most preparations do not achieve a level of purity sufficient for optimal detection of membrane proteins. Second, membrane proteins are by nature averse to most protein chemical purification methods (with maybe the exception of SDS-PAGE) and special care has to be taken to prevent their loss during the isolation protocol. Finally, membrane proteins are chemically composed of very hydrophobic stretches that, upon proteolytic cleavage, might not generate peptides detectable by mass spectrometry.

In general, it is not possible to specifically attribute the reasons for the limited success of a proteomics study to identify membrane proteins. However, since the BBM isolation protocol described in this study generated a preparation of highly enriched membrane proteins in complex mixture, there was a golden opportunity to investigate the nature, and more specifically, the length and the hydrophobicity of membrane proteins' peptides that were identified using a standard RPLC-ESI-MS/MS setup (with a LTQ mass spectrometer as mass analyzer).

4.6.1 Characteristics of identified peptides

For the purpose of this investigation, only the 663 identified tryptic peptides that were derived from 35 confirmed transmembrane BBM proteins were considered for subsequent analysis. First, the GRand AVerage of hydropathicitY (GRAVY) was manually calculated for each peptide using the ProtParam tool available at Expasy web site (http://www.expasy.org/tools/protparam.htmlsequences). The GRAVY value is calculated as the sum of hydrophathy values of all the constituting amino acids divided by the number of residues in the sequence(119) and reflects somehow the hydrophobicity of a protein or a peptide. Fig. 4.40 shows a two-dimensional graphic representation linking the GRAVY scores and the amino acid lengths of the 663 identified tryptic peptides, assuming that the GRAVY score calculation can also be applied to peptides. Tryptic peptides containing methionine oxidation were excluded while cysteines were assumed to be in reduced form.
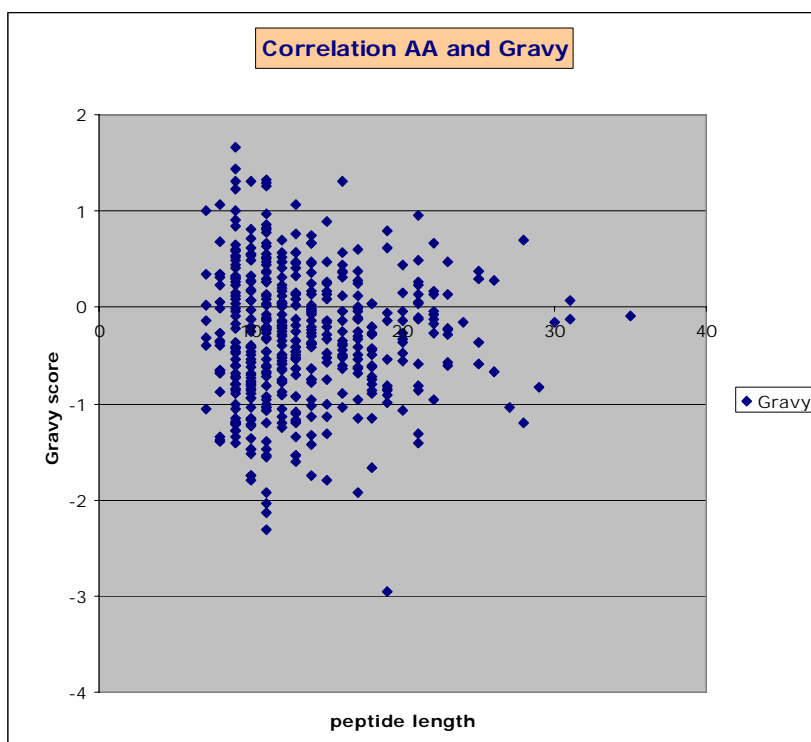
Figure 4.40: *Two-dimensional representation linking the GRAVY scores with the peptide amino acid length for the 663 identified tryptic peptides.* *Tryptic peptides containing methionine oxidation were excluded while cysteines were assumed to be in reduced form.*

For convenience, peptides with GRAVY values above 0.5 will be referred to as hydrophobic, those with GRAVY values below -0.5 as hydrophilic, while the remainder will be considered to be of medium character. Peptides distribution in the GRAVY scores was remarkably symmetrical with a slight tendency to negative values. Most of the identified peptides were of medium character and were less than twenty amino acids of length. There was a small number of hydrophobic peptides with GRAVY score values above 1, but those very hydrophobic peptides were typically short, i.e. between 10 and 15 amino acids long.

4.6.2 Comparison of identified peptides with predicted tryptic transmembrane peptides

In a next step, the position of the 663 identified tryptic peptides were determined in the 35 protein sequences in the dataset and compared to the location of transmembrane segments, as predicted by the TMHMM software. Fig. 4.41 shows two examples for representative integral BBM membrane proteins. The first example, the Nieman Pick C1-like protein 1, was one of the most abundant membrane proteins in the BBM preparation. This protein was predicted to possess 13 transmembrane helices and was identified with 16 different peptides. The second example, the FATP-4 protein, was a protein of rather low abundance in the preparation. It was predicted to possess 2 transmembrane helices and was identified with 2 different peptides.

A

Niemann-Pick C1-like protein 1 precursor

B

Fatty acid transport protein 4 (FATP-4)

Figure 4.41: *Amino acid sequence of two representative integral transmembrane BBM proteins. Panel A: Niemann-Pick C1-like protein 1, and Panel B: Fatty acid transporter protein 4. Lysines (K) and arginines (R) are highlighted in green and blue, respectively. Identified tryptic peptides are marked in bold red while the predicted location of transmembrane helices is underlined and marked in bold.*

Fig. 4.41 clearly shows that none of the identified tryptic peptides were located in the transmembrane helices, but rather within the loops (NPC1L1) or in the large cytoplasmic region of the protein (FATP-4).

Further, the hydropathicity of the predicted 228 transmembrane tryptic peptides belonging to the 35 integral membrane BBM proteins was calculated using the ProtParam tool. Fig. 4.42 shows a histogram comparing the amino acid lengths (panel A) and GRAVY scores (panel B) for the 663 identified tryptic peptides versus the 228 predicted tryptic transmembrane peptides. Not surprisingly, the vast majority of the identified tryptic peptides were much shorter and less hydrophobic than the predicted tryptic transmembrane peptides. However, according to these data, the two groups of peptides somehow overlap indicating that a small number of predicted transmembrane peptides should have been identified using the generic LC-MS conditions that were used in this experiment.
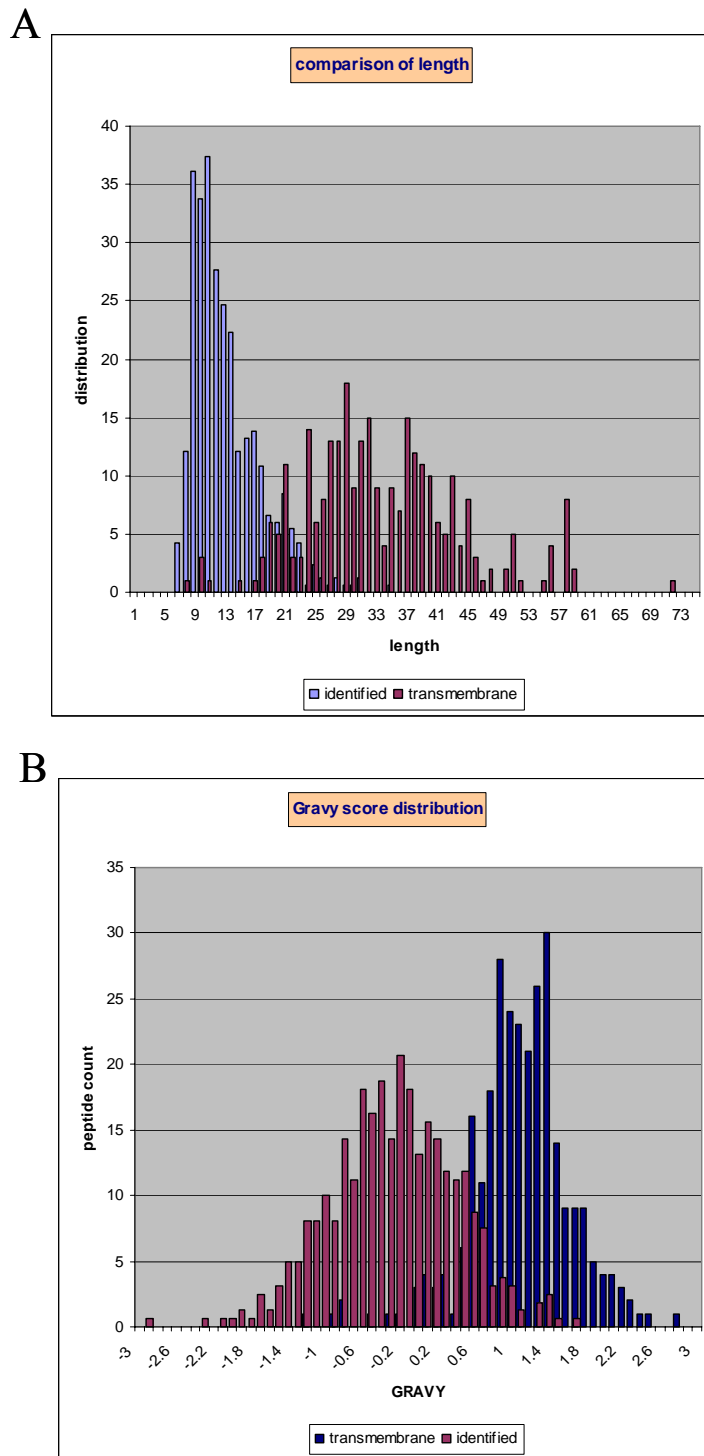
Figure 4.42: *Histogram of the amino acid length (Panel A) and hydropathicity (Panel B) for the 663 identified tryptic peptides versus the 288 predicted tryptic transmembrane peptides derived from the 35 identified transmembrane proteins.* The identified tryptic peptides are marked in red while the predicted transmembrane tryptic peptides are marked in blue. The total number of peptide counts was normalized.

Fig. 4.43 shows a two-dimensional view of the two histograms of Fig. 4.42 linking the GRAVY scores with the amino acid lengths for the identified tryptic peptides and the predicted tryptic transmembrane peptides.
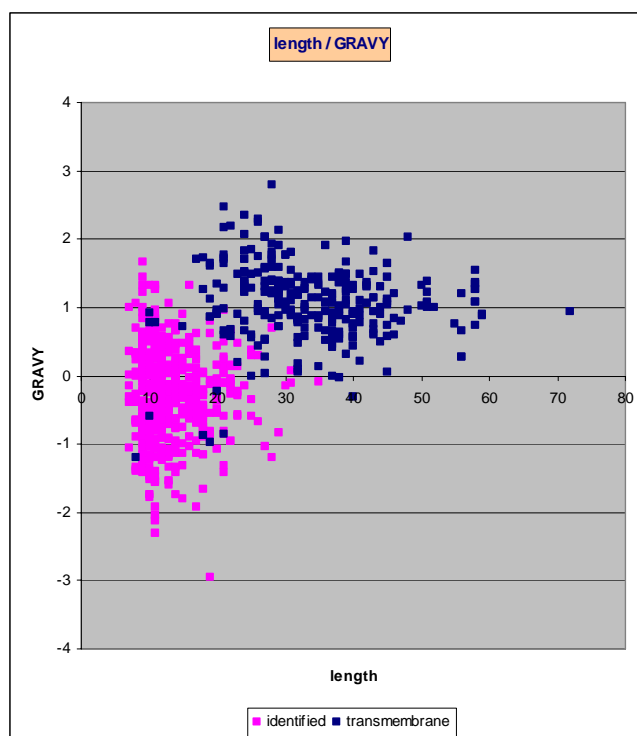
Figure 4.43: *Two-dimensional representation linking amino acid length and GRAVY scores for the 663 identified peptides and the 228 predicted transmembrane tryptic peptides. The observed tryptic peptides are marked in pink, the predicted tryptic transmembrane peptides in blue.*

Fig. 4.43 clearly confirms that the identified tryptic peptides were, in average, much shorter and somewhat less hydrophobic than their predicted tryptic transmembrane counterparts. It is evident that the GRAVY score was not discriminative to explain the lack of identification of those predicted transmembrane tryptic peptides. Indeed, a number of observed tryptic peptides bear a GRAVY score of 1 or above similarly to the predicted transmembrane tryptic peptides. Rather, the average length of the typical predicted transmembrane peptide seems to separate best the two populations. However, an overlap between those two peptides populations remains, with GRAVY values between 0 and 1 and peptide lengths between 20 and 30 amino acids, which strongly suggest that identification of at least a few of those transmembrane tryptic peptides should have been possible.

4.6.3 Discussion

One of the stated goals of this study was to ascertain at the experimental level the physicochemical properties of peptides derived from the proteolytic cleavage of membrane proteins, to determine if some of them belong to transmembrane helices, and to compare their characteristics with the predicted tryptic transmembrane peptides of the same identified proteins. The parameters that were chosen to characterize these peptides comprised the GRAVY score and the peptide length. The majority of the identified peptides were of medium

character (GRAVY scores between -0.5 and 0.5) and their lengths ranged from 10 to 20 amino acids. In contrast, the majority of the predicted tryptic transmembrane peptides (containing up to one misscleavage since the search criteria for protein identification permitted it) were of 20-30 amino acids length and the GRAVY values ranged from 0.5 to 1.5. Surprisingly, no transmembrane tryptic peptide was experimentally identified although, according to the GRAVY score and peptide length predictions, there was a high degree of likelihood that at least a few of those peptides should have been experimentally detected. Overall, only one tryptic peptide could be identified whose amino acid sequence encompassed half of a transmembrane helix due to the presence of a lysine in the middle of the alpha helix segment.

In this initial study, which was performed using a LTQ mass spectrometer, only peptides with assumed charge state of +1, +2 or +3 were selected for MS/MS analysis and peptide identification. As a consequence, peptides of mass above 4500 Da were not considered, excluding most of the potential transmembrane tryptic peptides. In subsequent analyses, peptide identification was performed using the newest generation mass spectrometer (the LTQ Orbitrap mass spectrometer) at 30,000 mass resolution and 2 ppm mass accuracy so that peptides with higher charge state could potentially be considered. Indeed, when the same samples were analyzed using the Orbitrap mass spectrometer, signals with charge state +4 and +5 were detected in the m/z range between 1000–1500 Da corresponding roughly to the mass of potential transmembrane tryptic peptides. However, due to software limitations and poor MS/MS spectral quality for highly charged precursor ions in the LTQ mass analyzer, these types of peptides were routinely excluded from the analysis in this first round of protein identification. Nevertheless, the exact nature of these high molecular mass, multiply charged peptides should be investigated in future studies.

There are many other reasons that may lead to the absence of transmembrane tryptic peptides. Assuming that trypsin cleaves off transmembrane helices, that is, the transmembrane tryptic peptide is generated during the digestion procedure, the rather hydrophobic nature of most of these peptides may hinder their extraction from the gel band or they may stick to tube walls during peptide extraction. Finally, very long and/or hydrophobic peptides typically cannot be chromatographically analyzed on a C18 column and might require a different stationary phase, such as Hydrophilic-Interaction Liquid Chromatography (HILIC) (see appendix C for a detailed discussion).

The name Hydrophilic-Interaction Chromatography (HILIC) was originally coined by Alpert (120) for the separation of hydrophilic substances such as proteins, peptides and nucleic acids using polar stationary phase, such as polyhydroxyethyl aspartamide or polyhydroxysulfotamide aspartamide. Polar peptides are retained by the polar stationary phase in a hydrophobic solvent. The solute molecule is then eluted from the chromatographic beads by increasing the polarity of the mobile phase, for example by increasing the proportion of water or by adding salt. So far, HILIC has been suggested as an alternative to cation-exchange chromatography for the separation of polar peptides, such as glycopeptides, phosphopeptides and modified histone peptides (121). It is then typically coupled off-line with a RPLC-MS for peptide characterization.

In contrast to other studies, we suggest the use of HILIC for the chromatographic separation of very hydrophobic peptides, such as typically found in transmembrane helices of membrane proteins. Such peptides, which consist of long stretches of aliphatic and hydrophobic amino acids, might not survive the initial buffer conditions of a RPLC separation: those peptides cannot be kept soluble in aqueous solutions as they form secondary structure and precipitate before being injected into the RPLC system. In addition, their hydrophobicity might cause them to stick so strongly to the stationary phase that their elution might not be possible with conventional buffer systems. In contrast, the highly organic character of the initial buffer composition of a HILIC separation would favor the stability of those kinds of peptides while the polarity of the carbonyl backbone should ensure their retention onto the polar chromatographic medium. Similarly to RPLC, HILIC can be directly coupled to an electrospray ionization interface if the salt concentration is kept to a minimum to avoid interference with the ionization process. In this configuration, HILIC-MS might provide complementary information to the conventional RPLC-MS peptide analysis.

4.7 Preparation of intact RNA from the Small Intestine

4.7.1 Monitoring RNA degradation

RNA extraction from the small intestine was prepared in sections as follows. The dissected small intestine was divided in three equal segments, which were named according to their location to the stomach. The proximal segment contained the duodenum and the proximal part of jejunum while the central segment included the central part of the jejunum. The distal segment contained the distal part of the duodenum and the ileum. This division did not follow accurately the physiological partition of the small intestine. However, it represented a good compromise for dividing quickly and reproducibly the small intestine into three representative parts while preserving most of the section specificity. Each section was then further divided into three parts to accommodate the amount of tissue in the Lysing Fast RNA Matrix D tubes. Homogenization was performed using a FastPrep Cell disrupter. Total RNA was extracted according to the protocol listed in paragraph 3.2.8, after which the RNA common to the same section was pooled. RNA quality was then assessed using a 1% agarose denaturating gel (Fig. 4.44).
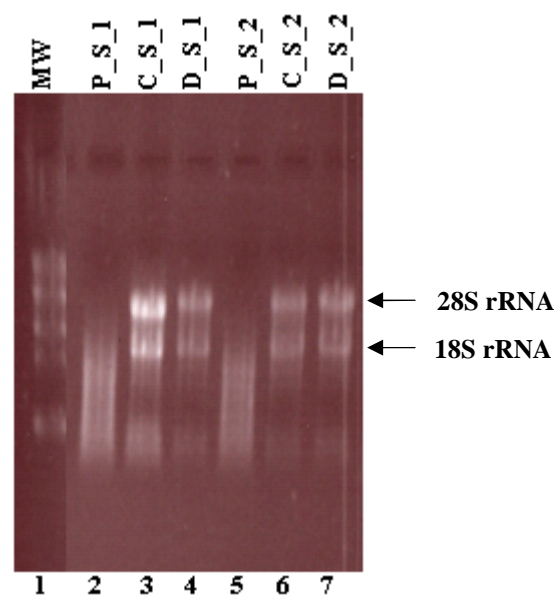
Fig.4.44: *Total RNA analysis by denaturating agarose gel electrophoresis. Approximately 1 µg total RNA extracted from the proximal, central, or distal sections of intestinal tissue were analyzed using a 1% agarose denaturating gel. The two intense bands highlighted by the arrows represent the heavy and the light subunit of ribosomal RNA.* **Lane 1**, *molecular weight marker (MW);* **Lanes 2-4** *and* **5-7**: *total RNA of the three sections from two different mice. P_S, proximal section; C_S, central section; D_S, distal section.*

Figure 4.44 shows that the 28S rRNA and 18S RNA was degraded in the proximal section of both mice while RNA quality was satisfactory for the central and distal sections. The RNA

extraction procedure was repeated on 5 additional intestinal samples to confirm that the extracted RNA was consistently degraded in the proximal sections (data not shown).

This observation led to the hypothesis that high RNAse activity might be the cause for RNA degradation in the jejunum. These RNAses might be of a different enzymatic family than in the other segments of the intestine or their specific activity could be higher in the duodenum. In order to confirm this hypothesis, the proximal part of the intestine was cut in three segments, with the most proximal part containing the duodenum while the other two segments only containing parts of the jejunum. RNA was extracted from these sub-sections and the quality of the extracted RNA was assessed again by denaturating agarose gel electrophoresis (Fig. 4.45). RNA quality was found to significantly worsen as the tissue was excised closer to the stomach. RNA extracted from the duodenum was completely degraded, while the RNA extracted from the jejunum proximal and central sections was partially degraded (jejunum proximal) or even mostly intact (jejunum central). These results confirm our hypothesis that RNAse activity is very high in the duodenum in comparison to the other sections and requires specific handling in order to isolate intact RNA for subsequent analysis.
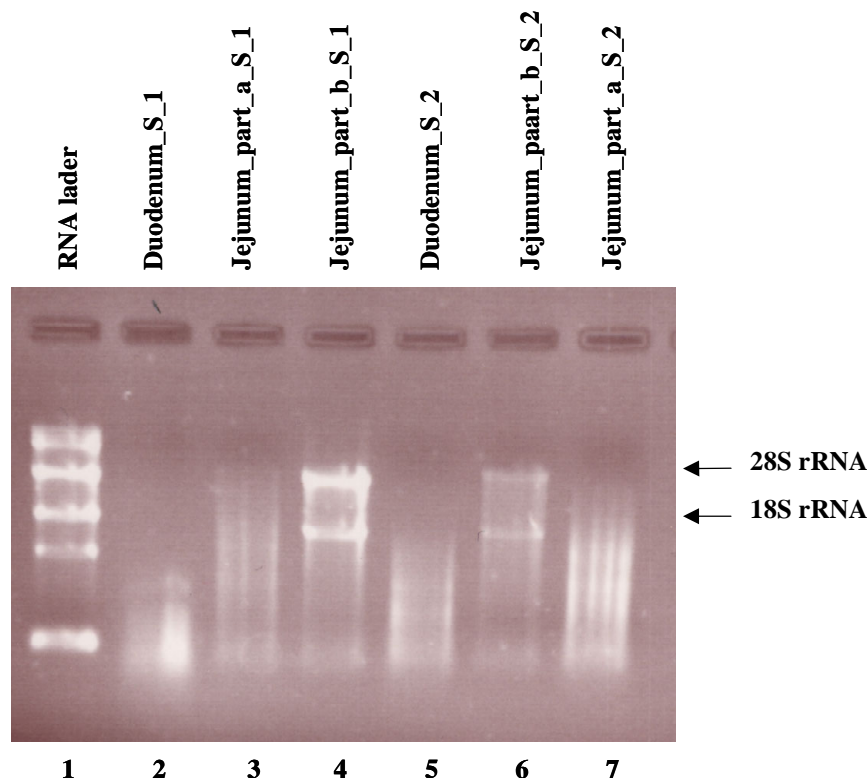
Fig.4.45: *Total RNA analysis by denaturating agarose gel electrophoresis. Approximately 1 μg total RNA extracted from the three parts of the proximal section (duodenum, jejunum a, and jejunum b) were analyzed using a 1% agarose denaturating gel. The two intense bands highlighted by the arrows represent the heavy and the light subunit of ribosomal RNA.* **Lane 1,** *molecular weight marker (RNA ladder);* **Lanes 2-4 and 5-7:** *total RNA from the three parts of the proximal section of two different mice S_1 and S_2.*

4.7.2 Inhibition of RNA degradation

The analysis of gene expression along the small intestine of a mouse, in particular with the goal to compare gene expression of differentially treated mice or of mice with different genetic background, requires a much more complete inhibition of RNA degradation in the duodenum section than previously achieved.

Attempts to improve the quality of the extracted RNA by adding RNAse inhibitors in the PBS buffer used to flush the intestine or by flushing and storing the sections in RNAlater directly did not significantly prevent RNA degradation in the duodenum. Similarly, shorter homogenization steps or setting the FastPrep Cell disrupter to lower homogenization speed did not result in any significant improvement of the RNA quality. Much better results were obtained using an alternate homogenization protocol involving tissue grinding under liquid nitrogen, pointing out to a temperature increase in the tissue during the homogenization step as being one of the major causes for RNA degradation. The main advantage of the grinding procedure consists in keeping the tissue constantly frozen until it is completely embedded in the phenol- and guanidine thiocyanate-containing TriReagent, which is an effective RNAse inhibitor. The reproducibility and the improved RNA quality using this alternate protocol were tested by processing the grinded dudenum from five different mice. The resulting RNA was assessed electrophoretically using the Experion System (Bio-Rad) as shown in Fig. 4.46. The general quality of the extracted RNA from grounded duodenum sections of five different mice was significantly better. The light and the heavy subunit of the ribosomal RNA did not appear to be degraded in any of the extractions and the ratio between the 28S and 16S RNA subunits (a quality assessment for RNA degradation) was superior to 1.5 in all the samples. These results demonstrate that significant RNA degradation in the jejunum section can be avoided provided that the samples temperature is controlled during the homogenization step.
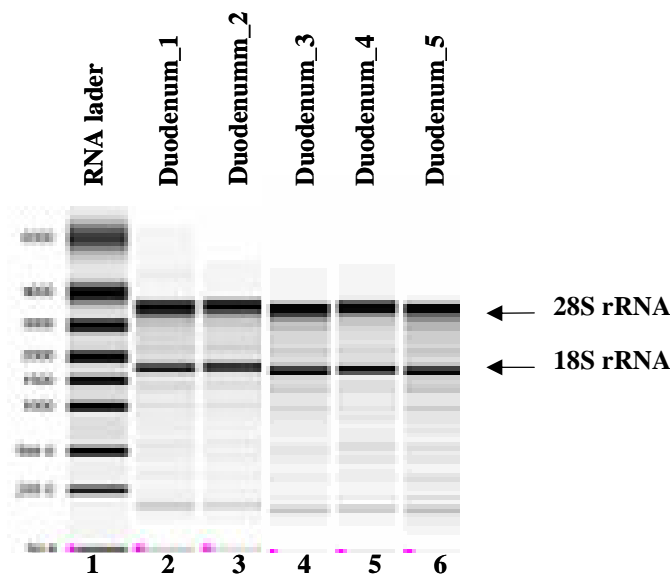
Fig.4.46: *Total RNA analysis using the automated electrophoresis Experion System (Bio-Rad).* *Approximately 1 μg total RNA was loaded onto the system. The two intense bands highlighted by the arrows represent the heavy and the light subunit of ribosomal RNA.* ***Lane 1,*** *Molecular mass marker (RNA ladder);* ***Lanes 2-6*** *extracted RNA of the dudenum section from five different mice prepared with the grinding process.*

4.7.3 Discussion

A reliable protocol to extract good quality RNA is the first and most important factor for a successful differential analysis of gene expression. The small intestine, and in particular the duodenum, contains many peptidases and nucleases whose natural role is to degrade macromolecules such as proteins, RNA and DNA, to their constituents to facilitate their absorption through the small intestine. Their abundance and their intrinsic activity make RNA extraction of the small intestine quite challenging. A standard RNA extraction protocol based on tissue homogenization with Lysing Fast RNA Matrix D tubes in a solution that contained phenol and guanidine thiocyanate could not be used for RNA extraction of the jejunum as the endogeneous, highly active RNAses could not be inhibited sufficiently rapidly to prevent significant RNA degradation. In this study I developed a protocol for the extraction of high quality RNA enabling in future studies a comparison of gene expression between differentially treated mice. First attempts to modify the original protocol, such as speeding intestinal tissue collection and snap freezing in liquid nitrogen, flushing the collected tissue with PBS containing RNAse inhibitors, or flushing and storing the tissue in RNA later, as well as varying the homogenization process (shorter disruption time, slower disruption rate, increased cooling periods) failed to deliver a better RNA quality. A major breakthrough was obtained by replacing the homogenization step with a protocol including tissue grinding in

liquid nitrogen. This alternate procedure gave optimal results for the RNA extraction from the duodenum section. The current protocol, combining a speedy tissue collection, a quick wash with PBS, snap-freezing and tissue grinding in liquid nitrogen leads to the extraction of very good quality RNA that can be stored at $-80\ ^{\circ}\text{C}$ for further use by gene chip analysis or qRT-PCR.

5. CONCLUSIONS

The study of membrane proteins is a difficult topic on its own because of the particular characteristics of these proteins. The study of membrane proteins in a complex mixture where degradation is a natural process is even more challenging.

In this study I successfully developed a robust protocol for the BBM preparation from mouse intestinal mucosa, a tissue whose main function is to degrade nutrients in order to facilitate their ingestion. The inherent protein and RNA degradation which I faced during my early work had to be considered and taken care of to obtain a reproducible BBM preparation. A significant amount of time and effort was spent on developing protocols for the inhibition of protein and RNA degradation, which at the end was partially achieved. It is worth pointing out that none of the published studies previously dealing with BBM analysis using mass spectrometry, immunoassays or microarrays had mentioned any issue regarding tissue degradation.

A triplicate identification of the proteins part of the BBM resulted in the identification of more than 1460 proteins, of which "only" 260 were integral membrane protein, an apparently disappointing result considered that the preparation was thought to be highly enriched in membrane proteins. However, a detailed analysis of the proteins identified in the BBM preparation made apparent that a large extent of the remainder was accounted for membrane-associated, membrane-anchored proteins and cytoskeleton proteins involved in protein trafficking between cytosol and plasma membrane, thus closely related functionally to the BBM function. Even some of the 330 cytosolic proteins, such as the apolipropteins which could have been discounted as contaminant at a first glance, were demonstrated to be specifically co-enriched in the BBM fraction probably due to their interaction with a BBM-specific component.

A rough estimation of protein abundance based on peptide counts confirmed that membrane proteins and membrane-associated proteins represented the dominant species in the BBM preparation. In this analysis we didn't observe any particular bias against proteins with many transmembrane helices. Nevertheless, the identification of these proteins was based on tryptic peptides exclusively located in the loops, in extracellular areas or in the intracellular domains of the proteins. I believe that the reason for the lack of identification of tryptic transmembrane peptides (assuming that trypsin was equally active in this environment) was due in many cases to the length of the generated peptides, typically of 40-50 amino acids long. Indeed, the elution of highly charges peptides of 5-6 kDa mass were occasionally observed in the elution

chromatograms. Unfortunately, due to software limitation (peptide with charge state above +3 could not be considered in an automatic data analysis), these peptides were not analyzed and further studies will be required to confirm this observation.

This study is the first proteomic approach in which numerous receptors and transporters and many relevant proteins to cholesterol absorption were identified. This is in contrast to several recently published proteomic studies (2, 122) claiming the characterization of specific membrane proteins of BBM. However, their protein lists were mostly restricted to Ras-related proteins and cytosolic or membrane associated-proteins. Aminopeptidase N and Sodium/Glucose co-transporter, two of the most abundant proteins in the BBM, were characterized in a proteomic approach for the first time in this study. Similarly, some proteomic studies have been based on the isolation of the lipid rafts (possible BBM micro domains responsible for cholesterol absorption) to enrich and identify proteins that participate in cholesterol absorption (3). None of these studies reports the identification the Niemann-Pick C1-like protein 1, the target of ezetimibe (a cholesterol absorption inhibitor) and for many researchers the likely transporter of cholesterol. In this study, the Niemann-Pick C1-like protein 1 was one of the most abundant proteins in the purified BBM fraction.

These data re-emphasize the critical importance of a reproducible isolation and fractionation of a membrane preparation in a high-throughput proteomic approach, even if using a mass spectrometer with high resolving power and excellent mass accuracy. In this study, a robust BBM isolation protocol (in which the unspecific proteolytic activity of the preparation was controlled by adding a cocktail of protease inhibitors and peptide substrates) led us to the characterization at the protein level of a large number of cholesterol- or fat absorption-related proteins which, until now, had only been characterized at the gene level or by immunoassays. Thus, a membrane proteomic approach is feasible and may yield excellent results when the analyzed membrane fraction is optimally purified. Evidently, the requirement for a complex purification and fractionation scheme is accompanied by an increasing analysis complexity. Preparation of the BBM in sections enabled the identification of some additional BBM-specific proteins and provided some examples of protein localization along the small intestine. For example, the SR-BI receptor was identified in the BBM fraction prepared from total mucosa with one single peptide (and hence was not included in the final list in appendix A1). In contrast, it was robustly characterized in the BBM fraction prepared from the duodenum segment of the small intestine with three different peptides. The benefit of the additional fractionation was obvious in respect to the confident identification of this receptor. However, it required three times more measurement time and a rather large increase of starting material.

150

The cost/benefit of the selected isolation scheme needs to be carefully considered in respect to how many biological replicates or differential experiments are planned.

The specific identification of the ApoE protein in the BBM fraction captured our interest and led us to a prototype comparison study between BBM preparations from ApoE knockout mice and wild type mice of the same genetic background. Following the same preparation protocol and a well established analytical strategy, we were able to see clear differences in the expression level of several proteins between the two different species. Since no quantification technique was applied and there was no biological replicate, we focused our interest in the "black and white" differences. Among them, the Ileal Bile Acid Transporter (IBAT) protein and the ApoAI protein, which were robustly identified in the wild type animals, couldn't be detected in the ApoE knockout mice. This finding was not described in any of the earlier published studies that have attempted to elucidate the reasons why the ApoE knockout mice appear hypercholesterolemic, have high plasma cholesterol and high cholesterol absorption. In particular, the striking downregulation of the IBAT transporter in the ApoE knockout animal strongly suggests a disruption of the bile acids metabolism. This hypothesis is supported by a separate study that showed that the bile acid pool size in the liver and the intestine of the ApoE knockout mice was significantly increased compare to the wild type mice. The increase of the bile acids pool available in the small intestine of the ApoE knockout mice might explain the higher cholesterol absorption of these animals since cholesterol is better solubilized and, consequently, transported more easily through the BBM. This part of my study provides a completely new perspective to the high cholesterol absorption of these animals and opens a new field of research in this topic.

Similarly to this prototype study, I am convinced that the analytical strategy described in this study is sufficiently mature to perform comprehensive comparative analysis of mice that have for example been treated with specific compound or that have been subjected to different diets. The simplistic analysis method used in the ApoE knockout example was sufficient to pinpoint to obvious changes. However, in a more complex dataset, a proper quantification strategy should provide a better leverage to discover differences with statistical significance. The ^{15}N full metabolic labeling of mice might represent an ideal strategy for this type of samples. The mixing of an internal standard in the form of a "heavy" BBM preparation in all the samples of interest enables very accurate protein quantification (in principle, all BBM proteins could be normalized to their specific internal standard) over the whole experiment. In addition, the co-purification of the ^{15}N internal standard with the regular BBM preparation provides a mean to compensate for variability in degradation across the experiment.

A significant amount of time and effort was spent in this study for evaluating the stability of the BBM preparation protocol, the reproducibility of the analytical steps leading to the protein identification and, further, the consistence of these analyses with an initial comparative investigation of how to extract the same information directly from the LC-MS precursor ion signals. Overall, the BBM isolation protocol that was developed in this study, combined with a reproducible analytical strategy, enabled us to identify almost 90% of all the 1639 proteins identified in the BBM fraction in at least two of three triplicates. This level of reproducibility was beyond our expectations and encouraged us to individually monitor the variability of each of the technical steps involved in the procedure. The data confirmed the critical stability and reproducibility of the Orbitrap mass spectrometer for the high throughput analysis of such samples. The study pinpointed to an unexpected source of variation in the form of the LC buffers, which need to be tightly controlled to obtain reproducible chromatographic separation conditions. Ideally, samples to be compared must be analyzed using identical buffer and column batches on the LC system.

My study in this respect initiated a discussion for better understanding the analytical requirements and the type of software tools that might allow in the future label-free quantification approaches based on LC-MS data alone. Quantification at the precursor mass signal intensity might provide a more general method of comparison for LC-MS-based proteomics studies than the comparison of protein lists, which only consider approximately 10 to 20% of the available data. Protein identification of signals of interest can then be obtained using MS/MS inclusion lists, where an ion signal of specific m/z and RT is targeted for tandem mass spectrometric analysis. Most importantly, the ability to differentially compare samples across an experiment first requires the knowledge of what is common to the technical and biological replicates of samples so that the significance of a given change can be statistically appreciated. Within this study, we determined what was achievable with regards to reproducibility using standard protein purification technique and advanced analytical strategies but we also described the current limitations that will need to be addressed in the future to fully enable this strategy.

6. LITERATURE

1. Hofer D, Jons T, Kraemer J and Drenckhahn D: From cytoskeleton to polarity and chemoreception in the gut epithelium. Ann N Y Acad Sci 859: 75-84, 1998.

2. Babusiak M, Man P, Petrak J and Vyoral D: Native proteomic analysis of protein complexes in murine intestinal brush border membranes. Proteomics 7: 121-9, 2007.

3. Nguyen HT, Amine AB, Lafitte D, Waheed AA, Nicoletti C, Villard C, Letisse M, Deyris V, Roziere M, Tchiakpe L, Danielle CD, Comeau L and Hiol A: Proteomic characterization of lipid rafts markers from the rat intestinal brush border. Biochem Biophys Res Commun 342: 236-44, 2006.

4. Simons K and Toomre D: Lipid rafts and signal transduction. Nat Rev Mol Cell Biol 1: 31-9, 2000.

5. Brown DA and London E: Structure and function of sphingolipid- and cholesterol-rich membrane rafts. J Biol Chem 275: 17221-4, 2000.

6. Schuck S and Simons K: Polarized sorting in epithelial cells: raft clustering and the biogenesis of the apical membrane. J Cell Sci 117: 5955-64, 2004.

7. Dietrich C, Bagatolli LA, Volovyk ZN, Thompson NL, Levi M, Jacobson K and Gratton E: Lipid rafts reconstituted in model membranes. Biophys J 80: 1417-28, 2001.

8. Danielsen EM and Hansen GH: Lipid rafts in epithelial brush borders: atypical membrane microdomains with specialized functions. Biochim Biophys Acta 1617: 1-9, 2003.

9. Lucero HA and Robbins PW: Lipid rafts-protein association and the regulation of protein activity. Arch Biochem Biophys 426: 208-24, 2004.

10. Ikonen E: Cellular cholesterol trafficking and compartmentalization. Nat Rev Mol Cell Biol 9: 125-38, 2008.

11. Dietschy JM, Turley SD and Spady DK: Role of liver in the maintenance of cholesterol and low density lipoprotein homeostasis in different animal species, including humans. J Lipid Res 34: 1637-59, 1993.

12. Ikonen E: Mechanisms for cellular cholesterol transport: defects and human disease. Physiol Rev 86: 1237-61, 2006.

13. Altmann SW, Davis HR, Jr., Zhu LJ, Yao X, Hoos LM, Tetzloff G, Iyer SP, Maguire M, Golovko A, Zeng M, Wang L, Murgolo N and Graziano MP: Niemann-Pick C1 Like 1 protein is critical for intestinal cholesterol absorption. Science 303: 1201-4, 2004.

14. Davis HR, Jr., Zhu LJ, Hoos LM, Tetzloff G, Maguire M, Liu J, Yao X, Iyer SP, Lam MH, Lund EG, Detmers PA, Graziano MP and Altmann SW: Niemann-Pick C1 Like 1 (NPC1L1) is the intestinal phytosterol and cholesterol transporter and a key modulator of whole-body cholesterol homeostasis. J Biol Chem 279: 33586-92, 2004.

15. Garcia-Calvo M, Lisnock J, Bull HG, Hawes BE, Burnett DA, Braun MP, Crona JH, Davis HR, Jr., Dean DC, Detmers PA, Graziano MP, Hughes M, Macintyre DE, Ogawa A, O'Neill K A, Iyer SP, Shevell DE, Smith MM, Tang YS, Makarewicz AM, Ujjainwalla F, Altmann SW, Chapman KT and Thornberry NA: The target of ezetimibe is Niemann-Pick C1-Like 1 (NPC1L1). Proc Natl Acad Sci U S A 102: 8132-7, 2005.

16. Graf GA, Li WP, Gerard RD, Gelissen I, White A, Cohen JC and Hobbs HH: Coexpression of ATP-binding cassette proteins ABCG5 and ABCG8 permits their transport to the apical surface. J Clin Invest 110: 659-69, 2002.

17. Hui DY and Howles PN: Molecular mechanisms of cholesterol absorption and transport in the intestine. Semin Cell Dev Biol 16: 183-92, 2005.

18. Klett EL and Patel SB: Biomedicine. Will the real cholesterol transporter please stand up. Science 303: 1149-50, 2004.

19. Wilkins MR, Gasteiger E, Sanchez JC, Appel RD and Hochstrasser DF: Protein identification with sequence tags. Curr Biol 6: 1543-4, 1996.

20. Olsen JV, de Godoy LM, Li G, Macek B, Mortensen P, Pesch R, Makarov A, Lange O, Horning S and Mann M: Parts per million mass accuracy on an Orbitrap mass spectrometer via lock mass injection into a C-trap. Mol Cell Proteomics 4: 2010-21, 2005.

21. Wallin E and von Heijne G: Genome-wide analysis of integral membrane proteins from eubacterial, archaean, and eukaryotic organisms. Protein Sci 7: 1029-38, 1998.

22. Santoni V, Molloy M and Rabilloud T: Membrane proteins and proteomics: un amour impossible? Electrophoresis 21: 1054-70, 2000.

23. Sinha P, Poland J, Schnolzer M and Rabilloud T: A new silver staining apparatus and procedure for matrix-assisted laser desorption/ionization-time of flight analysis of proteins after two-dimensional electrophoresis. Proteomics 1: 835-40, 2001.

24. Galeva N and Altermann M: Comparison of one-dimensional and two-dimensional gel electrophoresis as a separation tool for proteomic analysis of rat liver microsomes: cytochromes P450 and other membrane proteins. Proteomics 2: 713-22, 2002.

25. Brookes PS, Pinner A, Ramachandran A, Coward L, Barnes S, Kim H and Darley-Usmar VM: High throughput two-dimensional blue-native electrophoresis: a tool for functional proteomics of mitochondria and signaling complexes. Proteomics 2: 969-77, 2002.

26. Rais I, Karas M and Schagger H: Two-dimensional electrophoresis for the isolation of integral membrane proteins and mass spectrometric identification. Proteomics 4: 2567-71, 2004.

27. Washburn MP, Wolters D and Yates JR, 3rd: Large-scale analysis of the yeast proteome by multidimensional protein identification technology. Nat Biotechnol 19: 242-7, 2001.

28. Han DK, Eng J, Zhou H and Aebersold R: Quantitative profiling of differentiation-induced microsomal proteins using isotope-coded affinity tags and mass spectrometry. Nat Biotechnol 19: 946-51, 2001.

29. Yu YQ, Gilar M and Gebler JC: A complete peptide mapping of membrane proteins: a novel surfactant aiding the enzymatic digestion of bacteriorhodopsin. Rapid Commun Mass Spectrom 18: 711-5, 2004.

30. Blonder J, Conrads TP, Yu LR, Terunuma A, Janini GM, Issaq HJ, Vogel JC and Veenstra TD: A detergent- and cyanogen bromide-free method for integral membrane proteomics: application to Halobacterium purple membranes and the human epidermal membrane proteome. Proteomics 4: 31-45, 2004.

31. Wu CC, MacCoss MJ, Howell KE and Yates JR, 3rd: A method for the comprehensive proteomic analysis of membrane proteins. Nat Biotechnol 21: 532-8, 2003.

32. Lottspeich F and Zorbas H: Bioanalytik. Spektrum Akademischer Verlag 1998.

33. Barber M, Bordoli RS, Garner GV, Gordon DB, Sedgwick RD, Tetler LW and Tyler AN: Fast-atom-bombardment mass spectra of enkephalins. Biochem J 197: 401-4, 1981.

34. Håkansson PK, I.; Sundqvist, B.; Fohlman, J.; Peterson, P.; McNeal, C.J.; MacFarlane, R.D. : 127I-Plasma Desorption Mass Spectrometry of Insulin. Journal of American Chemical Society 104: 2948-2949, 1982.

35. Macfarlane RD and Torgerson DF: Californium-252 plasma desorption mass spectroscopy. Science 191: 920-5, 1976.

36. Karas MB, D.; Bahr, U.; Hillenkamp, F.: Matrix-Assisted Ultraviolet-Laser Desorption of Nonvolatile Compounds. International Journal of Mass Spectrometry and Ion Processes 78: 53-68, 1987.

37. Tanaka K: [Development of high resolution scanning electron microscopy and its application in medical biology]. Nippon Rinsho 45: 920-4, 1987.

38. Karas M and Hillenkamp F: Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. Anal Chem 60: 2299-301, 1988.

39. Mann M, Hendrickson RC and Pandey A: Analysis of proteins and proteomes by mass spectrometry. Annu Rev Biochem 70: 437-73, 2001.

40. Aebersold R and Mann M: Mass spectrometry-based proteomics. Nature 422: 198-207, 2003.

41. McLuckey SA, Wu J, Bundy JL, Stephenson JL, Jr. and Hurst GB: Oligonucleotide mixture analysis via electrospray and ion/ion reactions in a quadrupole ion trap. Anal Chem 74: 976-84, 2002.

42. Makarov A: Electrostatic axially harmonic orbital trapping: a high-performance technique of mass analysis. Anal Chem 72: 1156-62, 2000.

43. Hardman M and Makarov AA: Interfacing the orbitrap mass analyzer to an electrospray ion source. Anal Chem 75: 1699-705, 2003.

44. Schwartz JC, Wade AP, Enke CG and Cooks RG: Systematic delineation of scan modes in multidimensional mass spectrometry. Anal Chem 62: 1809-18, 1990.

45. Hu Q, Noll RJ, Li H, Makarov A, Hardman M and Graham Cooks R: The Orbitrap: a new mass spectrometer. J Mass Spectrom 40: 430-43, 2005.

46. James P, Quadroni M, Carafoli E and Gonnet G: Protein identification by mass profile fingerprinting. Biochem Biophys Res Commun 195: 58-64, 1993.

47. Pappin DJ, Hojrup P and Bleasby AJ: Rapid identification of proteins by peptide-mass fingerprinting. Curr Biol 3: 327-32, 1993.

48. Song Z, Chen L, Ganapathy A, Wan XF, Brechenmacher L, Tao N, Emerich D, Stacey G and Xu D: Development and assessment of scoring functions for protein identification using PMF data. Electrophoresis 28: 864-70, 2007.

49. Monigatti F and Berndt P: Algorithm for accurate similarity measurements of peptide mass fingerprints and its application. J Am Soc Mass Spectrom 16: 13-21, 2005.

50. Perkins DN, Pappin DJ, Creasy DM and Cottrell JS: Probability-based protein identification by searching sequence databases using mass spectrometry data. Electrophoresis 20: 3551-67, 1999.

51. Nesvizhskii AI, Vitek O and Aebersold R: Analysis and validation of proteomic data generated by tandem mass spectrometry. Nat Methods 4: 787-97, 2007.

52. Elias JE, Haas W, Faherty BK and Gygi SP: Comparative evaluation of mass spectrometry platforms used in large-scale proteomics investigations. Nat Methods 2: 667-75, 2005.

53. Stein S, Kilpatrick, L., Neta, P., and Roth, J. : Building and using reference libraries of peptide mass spectra. Proceedings of the 53rd ASMS Conference on Mass Spectrometry, A051573: 2005.

54. Nesvizhskii AI, Keller A, Kolker E and Aebersold R: A statistical model for identifying proteins by tandem mass spectrometry. Anal Chem 75: 4646-58, 2003.

55. Ong SE, Foster LJ and Mann M: Mass spectrometric-based approaches in quantitative proteomics. Methods 29: 124-30, 2003.

56. Mueller LN, Brusniak MY, Mani DR and Aebersold R: An assessment of software solutions for the analysis of mass spectrometry based quantitative proteomics data. J Proteome Res 7: 51-61, 2008.

57. Colinge J, Chiappe D, Lagache S, Moniatte M and Bougueleret L: Differential proteomics via probabilistic peptide identification scores. Anal Chem 77: 596-606, 2005.

58. Gao J, Opiteck GJ, Friedrichs MS, Dongre AR and Hefta SA: Changes in the protein expression of yeast as a function of carbon source. J Proteome Res 2: 643-9, 2003.

59. Ishihama Y, Oda Y, Tabata T, Sato T, Nagasu T, Rappsilber J and Mann M: Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein. Mol Cell Proteomics 4: 1265-72, 2005.

60. Liu H, Sadygov RG and Yates JR, 3rd: A model for random sampling and estimation of relative protein abundance in shotgun proteomics. Anal Chem 76: 4193-201, 2004.

61. Kuster B, Schirle M, Mallick P and Aebersold R: Scoring proteomes with proteotypic peptide probes. Nat Rev Mol Cell Biol 6: 577-83, 2005.

62. Higgs RE, Knierman MD, Gelfanova V, Butler JP and Hale JE: Comprehensive label-free method for the relative quantification of proteins from biological samples. J Proteome Res 4: 1442-50, 2005.

63. Kusmierz JJ, Sumrada R and Desiderio DM: Fast atom bombardment mass spectrometric quantitative analysis of methionine-enkephalin in human pituitary tissues. Anal Chem 62: 2395-400, 1990.

64. Choe L, D'Ascenzo M, Relkin NR, Pappin D, Ross P, Williamson B, Guertin S, Pribil P and Lee KH: 8-plex quantitation of changes in cerebrospinal fluid protein expression in subjects undergoing intravenous immunoglobulin treatment for Alzheimer's disease. Proteomics 7: 3651-60, 2007.

65. Gygi SP, Rist B, Gerber SA, Turecek F, Gelb MH and Aebersold R: Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. Nat Biotechnol 17: 994-9, 1999.

66. Ong SE, Blagoev B, Kratchmarova I, Kristensen DB, Steen H, Pandey A and Mann M: Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. Mol Cell Proteomics 1: 376-86, 2002.

67. Wang S, Zhang X and Regnier FE: Quantitative proteomics strategy involving the selection of peptides containing both cysteine and histidine from tryptic digests of cell lysates. J Chromatogr A 949: 153-62, 2002.

68. Stewart, II, Thomson T and Figeys D: 18O labeling: a tool for proteomics. Rapid Commun Mass Spectrom 15: 2456-65, 2001.

69. Lahm HW and Langen H: Mass spectrometry: a tool for the identification of proteins separated by gels. Electrophoresis 21: 2105-14, 2000.

70. Langen H, Fountoulakis M, Evers S, Wipf B and Berndt P: In From Genome to Proteome. in Proceedings from the 3rd Siena 2-D Electrophoresis Meeting p58, 1998.

71. McClatchy DB, Dong MQ, Wu CC, Venable JD and Yates JR, 3rd: 15N metabolic labeling of mammalian tissue with slow protein turnover. J Proteome Res 6: 2005-10, 2007.

72. Stahl-Zeng J, Lange V, Ossola R, Eckhardt K, Krek W, Aebersold R and Domon B: High sensitivity detection of plasma proteins by multiple reaction monitoring of N-glycosites. Mol Cell Proteomics 6: 1809-17, 2007.

73. Listgarten J and Emili A: Statistical and computational methods for comparative proteomic profiling using liquid chromatography-tandem mass spectrometry. Mol Cell Proteomics 4: 419-34, 2005.

74. Mueller LN, Rinner O, Schmidt A, Letarte S, Bodenmiller B, Brusniak MY, Vitek O, Aebersold R and Muller M: SuperHirn - a novel tool for high resolution LC-MS-based peptide/protein profiling. Proteomics 7: 3470-80, 2007.

75. Jaffe JD, Mani DR, Leptos KC, Church GM, Gillette MA and Carr SA: PEPPeR, a platform for experimental proteomic pattern recognition. Mol Cell Proteomics 5: 1927-41, 2006.

76. Rinner O, Mueller LN, Hubalek M, Muller M, Gstaiger M and Aebersold R: An integrated mass spectrometric and computational framework for the analysis of protein interaction networks. Nat Biotechnol 25: 345-52, 2007.

77. Zimmer JS, Monroe ME, Qian WJ and Smith RD: Advances in proteomics data analysis and display using an accurate mass and time tag approach. Mass Spectrom Rev 25: 450-82, 2006.

78. Bodenmiller B, Malmstrom J, Gerrits B, Campbell D, Lam H, Schmidt A, Rinner O, Mueller LN, Shannon PT, Pedrioli PG, Panse C, Lee HK, Schlapbach R and Aebersold

R: PhosphoPep--a phosphoproteome resource for systems biology research in Drosophila Kc167 cells. Mol Syst Biol 3: 139, 2007.

79. Kessler M, Acuto O, Storelli C, Murer H, Muller M and Semenza G: A modified procedure for the rapid preparation of efficiently transporting vesicles from small intestinal brush border membranes. Their use in investigating some properties of D-glucose and choline transport systems. In: Biochim Biophys Acta. 1978, pp 136-54.

80. Hooper NM: Families of zinc metalloproteases. FEBS Lett 354: 1-6, 1994.

81. Hooper NM, Hesp RJ and Tieku S: Metabolism of aspartame by human and pig intestinal microvillar peptidases. Biochem J 298 Pt 3: 635-9, 1994.

82. Tarentino AL and Plummer TH, Jr.: Enzymatic deglycosylation of asparagine-linked glycans: purification, properties, and specificity of oligosaccharide-cleaving enzymes from Flavobacterium meningosepticum. Methods Enzymol 230: 44-57, 1994.

83. Edge AS: Deglycosylation of glycoproteins with trifluoromethanesulphonic acid: elucidation of molecular structure and function. Biochem J 376: 339-50, 2003.

84. Sojar HT and Bahl OP: A chemical method for the deglycosylation of proteins. Arch Biochem Biophys 259: 52-7, 1987.

85. Neuhoff V, Arold N, Taube D and Ehrhardt W: Improved staining of proteins in polyacrylamide gels including isoelectric focusing gels with clear background at nanogram sensitivity using Coomassie Brilliant Blue G-250 and R-250. Electrophoresis 9: 255-62, 1988.

86. Gharahdaghi F, Weinberg CR, Meagher DA, Imai BS and Mische SM: Mass spectrometric identification of proteins from silver-stained polyacrylamide gel: a method for the removal of silver ions to enhance sensitivity. Electrophoresis 20: 601-5, 1999.

87. Shevchenko A, Wilm M, Vorm O and Mann M: Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. Anal Chem 68: 850-8, 1996.

88. Wilm M, Shevchenko A, Houthaeve T, Breit S, Schweigerer L, Fotsis T and Mann M: Femtomole sequencing of proteins from polyacrylamide gels by nano-electrospray mass spectrometry. Nature 379: 466-9, 1996.

89. Birney E, Clamp M and Durbin R: GeneWise and Genomewise. Genome Res 14: 988-95, 2004.

90. Birney E and Durbin R: Using GeneWise in the Drosophila annotation experiment. Genome Res 10: 547-8, 2000.

91. Yates JR, 3rd, Eng JK and McCormack AL: Mining genomes: correlating tandem mass spectra of modified and unmodified peptides to sequences in nucleotide databases. Anal Chem 67: 3202-10, 1995.

92. Elias JE and Gygi SP: Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. Nat Methods 4: 207-14, 2007.

93. Balgley BM, Laudeman T, Yang L, Song T and Lee CS: Comparative evaluation of tandem MS search algorithms using a target-decoy search strategy. Mol Cell Proteomics 6: 1599-608, 2007.

94. Nielsen H, Engelbrecht J, Brunak S and von Heijne G: A neural network method for identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. Int J Neural Syst 8: 581-99, 1997.

95. Klein P, Kanehisa M and DeLisi C: The detection and classification of membrane-spanning proteins. Biochim Biophys Acta 815: 468-76, 1985.

96. Krogh A, Larsson B, von Heijne G and Sonnhammer EL: Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol 305: 567-80, 2001.

97. Chomczynski P and Sacchi N: Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Anal Biochem 162: 156-9, 1987.

98. Albert A, Margarit J, Julia V, Sancho MA, Galan X, Lopez D and Morales L: Morphology and mucosal biochemistry of gastroschisis intestine in urine-free amniotic fluid. J Pediatr Surg 38: 1217-20, 2003.

99. Bauvois B and Dauzonne D: Aminopeptidase-N/CD13 (EC 3.4.11.2) inhibitors: chemistry, biological evaluations, and therapeutic prospects. Med Res Rev 26: 88-130, 2006.

100. Randall A, Cheng J, Sweredoski M and Baldi P: TMBpro: secondary structure, beta-contact and tertiary structure prediction of transmembrane beta-barrel proteins. Bioinformatics 24: 513-20, 2008.

101. Liu J and Rost B: Comparing function and structure between entire proteomes. Protein Sci 10: 1970-9, 2001.

102. Kramer W, Girbig F, Corsiero D, Pfenninger A, Frick W, Jahne G, Rhein M, Wendler W, Lottspeich F, Hochleitner EO, Orso E and Schmitz G: Aminopeptidase N (CD13) is a molecular target of the cholesterol absorption inhibitor ezetimibe in the enterocyte brush border membrane. J Biol Chem 280: 1306-20, 2005.

103. Field FJ, Born E, Murthy S and Mathur SN: Caveolin is present in intestinal cells: role in cholesterol trafficking? J Lipid Res 39: 1938-50, 1998.

104. Hauser H, Dyer JH, Nandy A, Vega MA, Werder M, Bieliauskaite E, Weber FE, Compassi S, Gemperli A, Boffelli D, Wehrli E, Schulthess G and Phillips MC: Identification of a receptor mediating absorption of dietary cholesterol in the intestine. Biochemistry 37: 17843-50, 1998.

105. Bietrix F, Yan D, Nauze M, Rolland C, Bertrand-Michel J, Comera C, Schaak S, Barbaras R, Groen AK, Perret B, Terce F and Collet X: Accelerated lipid absorption in mice overexpressing intestinal SR-BI. J Biol Chem 281: 7214-9, 2006.

106. Levy E, Spahis S, Sinnett D, Peretti N, Maupas-Schwalm F, Delvin E, Lambert M and Lavoie MA: Intestinal cholesterol transport proteins: an update and beyond. Curr Opin Lipidol 18: 310-8, 2007.

107. Croy JE, Brandon T and Komives EA: Two apolipoprotein E mimetic peptides, ApoE(130-149) and ApoE(141-155)2, bind to LRP1. Biochemistry 43: 7328-35, 2004.

108. Christensen EI and Birn H: Megalin and cubilin: synergistic endocytic receptors in renal proximal tubule. Am J Physiol Renal Physiol 280: F562-73, 2001.

109. Takano M, Koyama Y, Nishikawa H, Murakami T and Yumoto R: Segment-selective absorption of lysozyme in the intestine. Eur J Pharmacol 502: 149-55, 2004.

110. Kramer W, Girbig F, Corsiero D, Burger K, Fahrenholz F, Jung C and Muller G: Intestinal cholesterol absorption: identification of different binding proteins for cholesterol and cholesterol absorption inhibitors in the enterocyte brush border membrane. Biochim Biophys Acta 1633: 13-26, 2003.

111. Hansen GH, Immerdal L, Thorsen E, Niels-Christiansen LL, Nystrom BT, Demant EJ and Danielsen EM: Lipid rafts exist as stable cholesterol-independent microdomains in the brush border membrane of enterocytes. J Biol Chem 276: 32338-44, 2001.

112. Yu L, Bharadwaj S, Brown JM, Ma Y, Du W, Davis MA, Michaely P, Liu P, Willingham MC and Rudel LL: Cholesterol-regulated translocation of NPC1L1 to the cell surface facilitates free cholesterol uptake. J Biol Chem 281: 6616-24, 2006.

113. Field FJ and Mathur SN: Intestinal lipoprotein synthesis and secretion. Prog Lipid Res 34: 185-98, 1995.

114. Sehayek E, Shefer S, Nguyen LB, Ono JG, Merkel M and Breslow JL: Apolipoprotein E regulates dietary cholesterol absorption and biliary cholesterol excretion: studies in C57BL/6 apolipoprotein E knockout mice. Proc Natl Acad Sci U S A 97: 3433-7, 2000.

115. Hakansson P, Andersson I, Nystrom S, Lofgren L, Amrot LF and Li H: Ontogenetic development and spatial distribution of the ileal apical sodium-dependent bile acid transporter and the ileal lipid-binding protein in apoE knockout and C57BL/6 mice. Scand J Gastroenterol 37: 1089-96, 2002.

116. Woollett LA, Osono Y, Herz J and Dietschy JM: Apolipoprotein E competitively inhibits receptor-dependent low density lipoprotein uptake by the liver but has no effect on cholesterol absorption or synthesis in the mouse. Proc Natl Acad Sci U S A 92: 12500-4, 1995.

117. Deutsch EW, Eng JK, Zhang H, King NL, Nesvizhskii AI, Lin B, Lee H, Yi EC, Ossola R and Aebersold R: Human Plasma PeptideAtlas. Proteomics 5: 3497-500, 2005.

118. Rauch A, Bellew M, Eng J, Fitzgibbon M, Holzman T, Hussey P, Igra M, Maclean B, Lin CW, Detter A, Fang R, Faca V, Gafken P, Zhang H, Whiteaker J, States D, Hanash S, Paulovich A and McIntosh MW: Computational Proteomics Analysis System (CPAS): an extensible, open-source analytic system for evaluating and publishing proteomic data and high throughput biological experiments. J Proteome Res 5: 112-21, 2006.

119. Kyte J and Doolittle RF: A simple method for displaying the hydropathic character of a protein. J Mol Biol 157: 105-32, 1982.

120. Alpert AJ: Hydrophilic-interaction chromatography for the separation of peptides, nucleic acids and other polar compounds. J Chromatogr 499: 177-96, 1990.

121. Boersema PJ, Mohammed S and Heck AJ: Hydrophilic interaction liquid chromatography (HILIC) in proteomics. Anal Bioanal Chem 391: 151-9, 2008.

122. Cutillas PR, Biber J, Marks J, Jacob R, Stieger B, Cramer R, Waterfield M, Burlingame AL and Unwin RJ: Proteomic analysis of plasma membrane vesicles isolated from the rat renal cortex. Proteomics 5: 101-12, 2005.

123. Molnar I and Horvath C: Reverse-phase chromatography of polar biological substances: separation of catechol compounds by high-performance liquid chromatography. Clin Chem 22: 1497-502, 1976.

124. Strege MA, Stevenson S and Lawrence SM: Mixed-mode anion-cation exchange/hydrophilic interaction liquid chromatography-electrospray mass spectrometry as an alternative to reversed phase for small molecule drug discovery. Anal Chem 72: 4629-33, 2000.

125. Yoshida T: Peptide separation in normal phase liquid chromatography. Anal Chem 69: 3038-43, 1997.

126. Zhu BY, Mant CT and Hodges RS: Hydrophilic-interaction chromatography of peptides on hydrophilic and strong cation-exchange columns. J Chromatogr 548: 13-24, 1991.

127. el Rassi Z: Capillary electrophoresis of carbohydrates. Adv Chromatogr 34: 177-250, 1994.

128. Naidong W: Bioanalytical liquid chromatography tandem mass spectrometry methods on underivatized silica columns with aqueous/organic mobile phases. J Chromatogr B Analyt Technol Biomed Life Sci 796: 209-24, 2003.

129. Grumbach ES, Wagrowski-Diehl DM, Mazzeo JR, Alden B and Iraneta PC: Hydrophilic Interaction Chromatography Using Silica Columns for the Retention of Polar Analytes and Enhanced ESI-MS Sensitivity. LCGC NORTH AMERICA 22: 1010-1023, 2004.

130. Horie K, Ikegami T, Hosoya K, Saad N, Fiehn O and Tanaka N: Highly efficient monolithic silica capillary columns modified with poly(acrylic acid) for hydrophilic interaction chromatography. J Chromatogr A 1164: 198-205, 2007.

7. ABBREVIATION

ABC	Ammonium Bicarbonate
ACN	Acetonitrile
a-CYANO	a-Cyano-4-hydroxy-cynnamic-acid
AF	Ammonium Formate
BBM	Brush Border Membrane
Bicine	N,N-Bis(2-hydroxyethyl) Glycine
Bis-Tris	Bis(2-hydroxyethyl)amino-tris(hydroxymethyl)methane
BSA	Bovin Seum Albumin
βME	beta Mercaptoethanol
DTT	1.4 Dithioerythritol
EDTA	Ethylenediaminetetraacetic acid
ESI	Electrospray ionization
EtBr	Ethidium Bromide
FA	Formic Acid
GC	Gas Chromatography
GRAVY	Grand average of hydropathicity
HILIC	Hydrophilic Interaction Chromatography
IAA	Iodoacetamide
KO	Knockout
LC	Liquid Chromatography
LDS	Lithium dodecyl sulfate
MALDI	Matrix-assisted laser desorption/ionization
MOPS	3-(N-morpholino) ethane sulfonic acid
PMSF	Phenylmethylsulfonylfluorid
PVDF	Polyvinylidene difluoride
RPLC	Reverse Phase chromatography
RT	Retention time
SDS	Sodium dodecyl sulfate
TBS	Tris Buffer Saline
TFA	Trifluoroacetic acid
THF	Tetrahydrofuran
TOF	Time of flight
TTBS	Tween Tris Buffer Saline
WT	Wild type

8. ACKNOWLEDGEMENTS

APPENDIX

Appendix A. Protein identification

A1. BBM protein identification from whole mucosa

Total list of the 1306 identified BBM proteins from whole mucosa with a valid Swissprot database entries, listed by topology, then alphabetically according to their gene symbol. The color coding is as follows: yellow, "membrane" protein as predicted by the ALOM tool; blue, "secreted" protein as predicted by the "signal_anchor" software tool; green, "non-membrane, non-secreted" (=cytosolic) protein; red, proteins not analyzed by those software packages and manually added to the table.

Gene ID: internal Roche gene ID number

Gene Symbol, Swissprot protein ID, Swissprot protein AC: associated gene symbol, protein ID and protein accession number in the Swissprot database.

Gene Description: description of gene product as provided by the internal Roche database entry

Max Diff peptide: maximum number of different peptides observed in anyone of the triplicate LC-MS/MS analysis

Avg Pep count: sum of peptide counts for all the LC-MS analysis of this protein divided by the number of LC-MS analysis it has been detected.

Rank order: protein abundance ranking based on average peptide counts in the tripliocate LC-MS analysis.

Ortho Gene Symbol, Ortho Swissprot protein ID, Ortho Swissprot protein AC: associated orthologous (human) gene symbol, protein ID and protein accession number in the Swissprot database, as provided by the Roche internal database.

Topology: predicted protein topology: c=cytosolic, tm=transmembrane, s, secreted. Proteins labelled "ad" were manually added to the table after the topology analysis was performed.

GO term: GO annotation for the protein cellular location

Protein length: length of protein in amino acids

Number of TM: predicted number of transmembrane domains according to the TMHMM software package.

Gene ID (mouse)	Gene Symbol (mouse)	Swissprot protein ID (mouse)	Swissprot protein AC (mouse)	Gene Description	Max Diff peptide	Avg Pep count	Rank order	Ortho Gene ID	Ortho Gene Symbol	Swissprot protein ID	Swissprot protein AC	Topology	GO term	Protein length	Number of TM
74556	9130404H23Rik	Q9CU24_MOUSE	Q9CU24	RIKEN cDNA 9130404H23 gene	2	2	1087	-	-	-	-	ad	intrinsic to peroxisomal membrane	569	0
76279	Cyp2d26	CP2DQ_MOUSE	Q8CIM7	cytochrome P450, family 2, subfamily d, polypeptide 26	8	8	483	1564	CYP2D7P1	Pseudogene	-	ad	associated to endoplasmic reticulum membrane	500	2
14972	H2-K1	HA1B_MOUSE	P01901	histocompatibility 2, K1, K region	12	24	203	3107	HLA-C	1C07_HUMAN	P10321	ad	external side of plasma membrane	369	1
232714	Mgam	-	-	maltase-glucoamylase	81	1196	4	8972	MGAM	MGA_HUMAN	O43451	ad	integral to apical plasma membrane	1857	1
381417	RP23-357I14.1	A2AWR5_MOUSE	A2AWR5	novel protein similar to solute carrier family 28 (sodium-coupled nucleoside transporter) member 2 (Slc28a2)	4	6	565	-	-	-	-	ad	membrane	660	14
22599	Slc6a20b	S620B_MOUSE	O88575	solute carrier family 6 (neurotransmitter transporter), member 20B	3	4	789	54716	SLC6A20	S6A20_HUMAN	Q9NP91	ad	integral to plasma membrane	635	12
22187	Ubb	UBIQ_MOUSE	P62991	ubiquitin B	8	226	21	-	-	-	-	ad	cytoplasm	76	0
52398	Sep 11	SEP11_MOUSE	Q8C1B7	septin 11	5	5	635	55752	Sep 11	SEP11_HUMAN	Q9NVA2	c	septin complex	431	0
18951	Sept5	SEPT5_MOUSE	Q9Z2Q6	septin 5	3	3	895	5413	SEPT5	SEPT5_HUMAN	Q99719	c	plasma membrane	369	0
235072	Sept7	SEPT7_MOUSE	O55131	septin 7	4	6	558	989	SEPT7	SEPT7_HUMAN	Q16181	c	nucleus	436	0
66060	0610010O12Rik	CE032_MOUSE	Q8K353	UPF0467 protein C5orf32 homolog	2	7	538	84418	C5orf32	CE032_HUMAN	Q9H1C7	c	-	363	0
74190	1200009I06Rik	CN073_MOUSE	Q6DIA2	SEC6-like protein C14orf73 homolog	5	8	474	91828	C14orf73	CN073_HUMAN	Q17RC7	c	-	762	0
75471	1700009N14Rik	Q14AA6_MOUSE	Q14AA6	novel Ras protein	5	27	185	-	-	-	-	c	cytoplasm	216	0
112419	2010002M12Rik	Q3U687_MOUSE	Q3U687	homolog to Interferon-induced protein with tetratricopeptide repeats 1 (IFIT1_MOUSE)	20	3	870	3434	IFIT1	IFIT1_HUMAN	P09914	c	cytoplasm	466	0
70153	2210016F16Rik	Q9D7X2_MOUSE	Q9D7X2	hypothetical protein	3	2	1032	84267	C9orf64	CI064_HUMAN	Q5T6V5	c	cellular_component	159	0
70420	2610034B18Rik	CO038_MOUSE	Q9D0A3	UPF0552 protein C15orf38 homolog	4	2	961	348110	C15orf38	CO038_HUMAN	Q7Z6K5	c	associated to plasma membrane at apical part	226	0
66469	2810405K02Rik	CA093_MOUSE	Q9DB60	Uncharacterized protein C1orf93 homolog	7	12	360	127281	C1orf93	CA093_HUMAN	Q8TBF2	c	-	216	0
67268	2900073G15Rik	Q6ZWQ9_MOUSE	Q6ZWQ9	myosin light chain, regulatory B-like	3	4	769	10398	MYL9	MLRN_HUMAN	P24844	c	muscle myosin complex	172	0
73205	3110043O21Rik	CI072_MOUSE	Q6DFW0	Uncharacterized protein C9orf72 homolog	3	2	1038	203228	C9orf72	CI072_MOUSE	Q96LT7	c	-	821	0
70984	4931406C07Rik	CK054_MOUSE	Q91V76	Ester hydrolase C11orf54 homolog	6	5	622	28970	C11orf54	CK054_HUMAN	Q9H0W9	c	nucleus	315	0
78906	9130017N09Rik	CS021_MOUSE	Q9D279	Uncharacterized protein C19orf21 homolog	8	29	170	126353	C19orf21	CS021_HUMAN	Q8IVT2	c	-	153	0
231151	AA474455	-	-	transcriptional adaptor 2 (ADA2 homolog, yeast)-beta	2	2	1165	93624	TADA2B	TAD2B_HUMAN	Q86TJ2	c	nucleus	420	0
78894	Aacs	AACS_MOUSE	Q9D2R0	acetoacetyl-CoA synthetase	6	4	737	65985	AACS	AACS_HUMAN	Q86V21	c	cytoplasm	672	0
234734	Aars	SYAC_MOUSE	Q8BGQ7	alanyl-tRNA synthetase	2	2	1167	16	AARS	SYAC_HUMAN	P49588	c	cytoplasm	968	0
69684	Aarsd1	AASD1_MOUSE	Q3THG9	alanyl-tRNA synthetase domain containing 1	3	2	1031	80755	AARSD1	AASD1_HUMAN	Q9BTE6	c	cytoplasm	412	0
24015	Abce1	ABCE1_MOUSE	P61222	ATP-binding cassette, sub-family E (OABP), member 1	11	8	488	6059	ABCE1	ABCE1_HUMAN	P61221	c	cytoplasm	599	0
76491	Abhd14b	ABHEB_MOUSE	Q8VCR7	abhydrolase domain containing 14b	4	6	598	84836	ABHD14B	ABHEB_HUMAN	Q96IU4	c	cytoplasm	210	0
52538	Acaa2	THIM_MOUSE	Q8BWT1	acetyl-Coenzyme A acyltransferase 2 (mitochondrial 3-oxoacyl-Coenzyme A thiolase)	7	6	574	10449	ACAA2	THIM_HUMAN	P42765	c	mitochondrion	397	0
104112	Acly	ACLY_MOUSE	Q91V92	ATP citrate lyase	3	3	820	47	ACLY	ACLY_HUMAN	P53396	c	cytoplasm	1091	0
11428	Aco1	Q8VDC3_MOUSE	Q8VDC3	aconitase 1	3	3	929	48	ACO1	Q5VZA6_HUMAN	Q5VZA6	c	cytoplasm	899	0
11429	Aco2	ACON_MOUSE	Q99KI0	aconitase 2, mitochondrial	8	7	522	50	ACO2	ACON_HUMAN	Q99798	c	mitochondrion	780	0
329910	Acot11	ACO11_MOUSE	Q8VHQ9	acyl-CoA thioesterase 11	3	2	1050	26027	ACOT11	ACO11_HUMAN	Q8WXI4	c	cytoplasm	594	0
11430	Acox1	ACOX1_MOUSE	Q9R0H0	acyl-Coenzyme A oxidase 1, palmitoyl	5	7	517	51	ACOX1	ACOX1_HUMAN	Q15067	c	mitochondrion	661	0
11431	Acp1	PPAC_MOUSE	Q9D358	acid phosphatase 1, soluble	2	2	949	52	ACP1	PPAC_HUMAN	P24666	c	cytoplasm	158	0
11459	Acta1	ACTS_MOUSE	P68134	actin, alpha 1, skeletal muscle	14	280	13	58	ACTA1	ACTS_HUMAN	P68133	c	cytoskeleton	377	0
11475	Acta2	ACTA_MOUSE	P62737	actin, alpha 2, smooth muscle, aorta	14	250	17	72	ACTG2	ACTH_HUMAN	P63267	c	cytoskeleton	377	0
11461	Actb	ACTB_MOUSE	P60710	actin, beta, cytoplasmic	29	1737	1	60	ACTB	ACTB_HUMAN	P60709	c	cytoskeleton	375	0
238880	Actbl2	ACTBL_MOUSE	Q8BFZ3	actin, beta-like 2	8	37	131	345651	ACTBL2	ACTBL_HUMAN	Q562R1	c	cytoskeleton	376	0
11464	Actc1	ACTC_MOUSE	P68033	actin, alpha, cardiac	14	280	14	70	ACTC1	ACTC_HUMAN	P68032	c	cytoskeleton	377	0
11465	Actg1	ACTG_MOUSE	P63260	actin, gamma, cytoplasmic 1	29	983	5	71	ACTG1	ACTG_HUMAN	P63261	c	cytoskeleton	375	0
11468	Actg2	ACTH_MOUSE	P63268	actin, gamma 2, smooth muscle, enteric	14	250	16	-	-	-	-	c	cytoskeleton	376	0
109711	Actn1	ACTN1_MOUSE	Q7TPR4	actinin, alpha 1	8	7	541	87	ACTN1	ACTN1_HUMAN	P12814	c	cytoskeleton	892	0
11472	Actn2	ACTN2_MOUSE	Q9JI91	actinin alpha 2	2	2	1128	88	ACTN2	ACTN2_HUMAN	P35609	c	cytoskeleton	894	0
11474	Actn3	ACTN3_MOUSE	O88990	actinin alpha 3	2	2	1129	89	ACTN3	ACTN3_HUMAN	Q08043	c	actin filament	900	0
60595	Actn4	ACTN4_MOUSE	P57780	actinin alpha 4	18	25	196	81	ACTN4	ACTN4_HUMAN	O43707	c	cytosol	912	0
66713	Actr2	ARP2_MOUSE	P61161	ARP2 actin-related protein 2 homolog (yeast)	8	8	489	10097	ACTR2	ARP2_HUMAN	P61160	c	actin cytoskeleton	394	0
74117	Actr3	ARP3_MOUSE	Q99JY9	ARP3 actin-related protein 3 homolog (yeast)	14	15	290	10096	ACTR3	ARP3_HUMAN	P61158	c	actin cytoskeleton	418	0
109652	Acy1	ACY1_MOUSE	Q99JW2	aminoacylase 1	6	3	823	95	ACY1	ACY1_HUMAN	Q03154	c	cytoplasm	408	0
11564	Adsl	PUR8_MOUSE	P54822	adenylosuccinate lyase	2	2	981	158	ADSL	PUR8_HUMAN	P30566	c	cytoplasm	484	0
76223	Agbl3	CBPC3_MOUSE	Q6CDP0	ATP/GTP binding protein-like 3	2	3	862	340351	AGBL3	CBPC3_HUMAN	Q8NEM8	c	cytoplasm	1006	0
269378	Ahcy	SAHH_MOUSE	P50247	S-adenosylhomocysteine hydrolase	12	13	334	191	AHCY	SAHH_HUMAN	P23526	c	cytoplasm	432	0
229709	Ahcyl1	SAHH2_MOUSE	Q80SW1	S-adenosylhomocysteine hydrolase-like 1	14	21	236	10768	AHCYL1	SAHH2_HUMAN	O43865	c	endoplasmic reticulum	530	0
74340	Ahcyl2	SAHH3_MOUSE	Q68FL4	S-adenosylhomocysteine hydrolase-like 2	13	32	157	23382	AHCYL2	SAHH3_HUMAN	Q96HN2	c	-	303	0
102502	AI427122	Q3V0K9_MOUSE	Q3V0K9	Plastin 1 (I isoform) homolog	34	122	37	5357	PLS1	PLSI_HUMAN	Q14651	c	cytoplasm	630	0
432442	Akap7	AKAP7_MOUSE	O55074	A kinase (PRKA) anchor protein 7	3	2	969	9465	AKAP7	AKA7A_HUMAN	O43687	c	anchored to apical plasma membrane	81	0
58810	Akr1a4	AK1A1_MOUSE	Q9JII6	aldo-keto reductase family 1, member A4 (aldehyde reductase)	16	31	160	10327	AKR1A1	AK1A1_HUMAN	P14550	c	apical plasma membrane	325	0
11677	Akr1b3	ALDR_MOUSE	P45376	aldo-keto reductase family 1, member B3 (aldose reductase)	4	6	609	231	AKR1B1	ALDR_HUMAN	P15121	c	cytoplasm	316	0
11997	Akr1b7	ALD1_MOUSE	P21300	aldo-keto reductase family 1, member B7	8	9	450	57016	AKR1B10	AK1BA_HUMAN	O60218	c	cytoplasm	316	0
622402	Akr1c12	Q9JLI0_MOUSE	Q9JLI0	aldo-keto reductase family 1, member C12	7	7	543	-	-	-	-	c	-	323	0
27384	Akr1c13	AK1CD_MOUSE	Q9VC28	aldo-keto reductase family 1, member C13	8	10	412	-	-	-	-	c	-	323	0
432720	Akr1c19	Q5I0T6_MOUSE	Q5I0T6	aldo-keto reductase family 1, member C19	7	8	493	-	-	-	-	c	-	323	0
110198	Akr7a5	ARK72_MOUSE	Q8CG76	aldo-keto reductase family 7, member A5 (aflatoxin aldehyde reductase)	5	4	775	246181	AFAR3	ARK74_HUMAN	Q8NHP1	c	mitochondrion	367	0
11668	Aldh1a1	AL1A1_MOUSE	P24549	aldehyde dehydrogenase family 1, subfamily A1	15	15	287	216	ALDH1A1	AL1A1_HUMAN	P00352	c	cytoplasm	501	0
72535	Aldh1b1	AL1B1_MOUSE	Q9CZS1	aldehyde dehydrogenase 1 family, member B1	14	27	184	219	ALDH1B1	AL1B1_HUMAN	P30837	c	mitochondrion	519	0
107747	Aldh1l1	FTHFD_MOUSE	Q8R0Y6	aldehyde dehydrogenase 1 family, member L1	3	2	1043	10840	ALDH1L1	FTHFD_HUMAN	O75891	c	cytoplasm	902	0
11669	Aldh2	ALDH2_MOUSE	P47738	aldehyde dehydrogenase 2, mitochondrial	6	7	518	217	ALDH2	ALDH2_HUMAN	P05091	c	mitochondrion	519	0
56752	Aldh9a1	AL9A1_MOUSE	Q9JLJ2	aldehyde dehydrogenase 9, subfamily A1	10	12	359	223	ALDH9A1	AL9A1_HUMAN	P49189	c	cytoplasm	494	0
11674	Aldoa	ALDOA_MOUSE	P05064	aldolase 1, A isoform	9	12	357	226	ALDOA	ALDOA_HUMAN	P04075	c	cytoplasm	196	0
230163	Aldob	ALDOB_MOUSE	Q91Y97	aldolase 2, B isoform	13	48	110	229	ALDOB	ALDOB_HUMAN	P05062	c	cytoplasm	364	0
72074	Anks4b	ANS4B_MOUSE	Q8K3X6	ankyrin repeat and sterile alpha motif domain containing 4B	20	98	54	257629	ANKS4B	ANS4B_HUMAN	Q8N8V4	c	plasma membrane	423	0
11737	Anp32a	AN32A_MOUSE	O35381	acidic (leucine-rich) nuclear phosphoprotein 32 family, member A	4	3	836	8125	ANP32A	AN32A_HUMAN	P39687	c	cytoplasm	247	0
67628	Anp32b	AN32B_MOUSE	Q9EST5	acidic nuclear phosphoprotein 32 family, member B	3	4	734	10541	ANP32B	AN32B_HUMAN	Q92688	c	nucleus	272	0
16952	Anxa1	ANXA1_MOUSE	P10107	annexin A1	3	3	844	301	ANXA1	ANXA1_HUMAN	P04083	c	basolateral plasma membrane	346	0
11744	Anxa11	ANX11_MOUSE	P97384	annexin A11	7	19	248	311	ANXA11	ANX11_HUMAN	P50995	c	cytoplasm	503	0
69787	Anxa13	ANX13_MOUSE	Q99JG3	annexin A13	18	69	70	312	ANXA13	ANX13_HUMAN	P27216	c	anchored to plasma membrane	317	0
12306	Anxa2	ANXA2_MOUSE	P07356	annexin A2	29	115	39	302	ANXA2	ANXA2_HUMAN	P07355	c	plasma membrane	339	0
11746	Anxa4	ANXA4_MOUSE	P97429	annexin A4	22	99	53	307	ANXA4	ANXA4_HUMAN	P09525	c	cytoplasm	319	0
11747	Anxa5	ANXA5_MOUSE	P48036	annexin A5	9	6	553	308	ANXA5	ANXA5_HUMAN	P08758	c	cytoplasm	319	0
11750	Anxa7	ANXA7_MOUSE	Q07076	annexin A7	7	10	417	310	ANXA7	ANXA7_HUMAN	P20073	c	plasma membrane	463	0
11764	Ap1b1	AP1B1_MOUSE	O35643	adaptor protein complex AP-1, beta 1 subunit	9	22	221	162	AP1B1	AP1B1_HUMAN	Q10567	c	clathrin adaptor complex	943	0
11765	Ap1g1	AP1G1_MOUSE	P22892	adaptor protein complex AP-1, gamma 1 subunit	7	8	484	164	AP1G1	AP1G1_HUMAN	O43747	c	clathrin adaptor complex	822	0
11766	Ap1g2	AP1G2_MOUSE	O88512	adaptor protein complex AP-1, gamma 2 subunit	6	4	688	8906	AP1G2	AP1G2_HUMAN	O75843	c	clathrin adaptor complex	791	0
11768	Ap1m2	AP1M2_MOUSE	Q9WVP1	adaptor protein complex AP-1, mu 2 subunit	13	13	323	10053	AP1M2	AP1M2_HUMAN	Q9Y6Q5	c	clathrin adaptor complex	423	0
11769	Ap1s1	AP1S1_MOUSE	P61967	adaptor protein complex AP-1, sigma 1	4	5	628	1174	AP1S1	AP1S1_HUMAN	P61966	c	clathrin adaptor complex	158	0
11772	Ap2a2	AP2A2_MOUSE	P17427	adaptor protein complex AP-2, alpha 2 subunit	3	3	711	161	AP2A2	AP2A2_HUMAN	O94973	c	clathrin adaptor complex	938	0
11773	Ap2m1	AP2M1_MOUSE	P84091	adaptor protein complex AP-2, mu1	11	10	409	1173	AP2M1	AP2M1_HUMAN	Q96CW1	c	clathrin adaptor complex	435	0
232910	Ap2s1	AP2S1_MOUSE	P62743	adaptor-related protein complex 2, sigma 1 subunit	2	2	967	1175	AP2S1	AP2S1_HUMAN	P53680	c	clathrin adaptor complex	142	0
55946	Ap3m1	AP3M1_MOUSE	Q9JKC8	adaptor-related protein complex 3, mu 1 subunit	3	2	1019	26985	AP3M1	AP3M1_HUMAN	Q9Y2T2	c	clathrin adaptor complex	418	0
245282	Apol10a	Q8CCA5_MOUSE	Q8CCA5	apolipoprotein L 10a	5	4	702	-	-	-	-	c	extracellular region	318	0
71896	Apol9b	Q8C7I4_MOUSE	Q8C7I4	apolipoprotein L 9b	3	2	1107	-	-	-	-	c	extracellular region	310	0
11836	Araf	ARAF_MOUSE	P04627	v-raf murine sarcoma 3611 viral oncogene homolog	5	4	755	369	ARAF	ARAF_HUMAN	P10398	c	cellular_component	604	0
213827	Arcn1	COPD_MOUSE	Q5XJY5	archain 1	4	5	644	372	ARCN1	COPD_HUMAN	P48444	c	COPI vesicle coat	511	0
11840	Arf1	ARF1_MOUSE	P84078	ADP-ribosylation factor 1	8	30	163	375	ARF1	ARF1_HUMAN	P84077	c	Golgi membrane	181	0
11841	Arf2	Q8BSL7_MOUSE	Q8BSL7	ADP-ribosylation factor 2	6	8	516	-	-	-	-	c	Golgi apparatus	181	0
11843	Arf4	ARF4_MOUSE	P61750	ADP-ribosylation factor 4	7	23	216	378	ARF4	ARF4_HUMAN	P18085	c	Golgi apparatus	180	0
11844	Arf5	ARF5_MOUSE	P84084	ADP-ribosylation factor 5	7	28	172	381	ARF5	ARF5_HUMAN	P84085	c	Golgi apparatus	180	0
11845	Arf6	ARF6_MOUSE	P62331	ADP-ribosylation factor 6	5	17	272	382	ARF6	ARF6_HUMAN	P62330	c	Golgi apparatus	175	0
11847	Arg2	ARGI2_MOUSE	O08691	arginase type II	5	4	712	384	ARG2	ARGI2_HUMAN	P78540	c	mitochondrion	354	0
228359	Arhgap1	RHG01_MOUSE	Q5FWK3	Rho GTPase activating protein 1	5	6	624	392	ARHGAP1	RHG01_HUMAN	Q07960	c	cytoplasm	439	1
544817	Arhgap27	RHG27_MOUSE	A2AB59	Rho GTPase activating protein 27	2	2	1054	201176	ARHGAP27	RHG27_HUMAN	Q6ZUM4	c	associated to membrane	869	0
192662	Arhgdia	GDIR1_MOUSE	Q99PT1	Rho GDP dissociation inhibitor (GDI) alpha	3	5	643	396	ARHGDIA	GDIR1_HUMAN	P52565	c	cytoplasm	204	0
230972	Arhgef16	ARHGG_MOUSE	Q3U5C8	Rho guanine nucleotide exchange factor (GEF) 16	2	4	777	27237	ARHGEF16	ARHGG_HUMAN	Q5VV41	c	cytoplasm	713	0
104303	Arl1	ARL1_MOUSE	P61211	ADP-ribosylation factor-like 1	4	4	699	400	ARL1	ARL1_HUMAN	P40616	c	Golgi apparatus	181	0
71619	Arl14	ARL14_MOUSE	Q3SXC5	ADP-ribosylation factor-like 14	2	2	1034	80117	ARL14	ARL14_HUMAN	Q8N4G2	c	cytoplasm	192	0
218639	Arl15	ARL15_MOUSE	Q8BGR6	ADP-ribosylation factor-like 15	3	5	1115	54622	ARL15	ARL15_HUMAN	Q9NXU5	c	-	204	0
56527	Arl2	ARL2_MOUSE	Q9D0J4	ADP-ribosylation factor-like 2	4	3	854	402	ARL2	ARL2_HUMAN	P36404	c	intracellular	184	0
75423	Arl5a	ARL5A_MOUSE	Q80ZU0	ADP-ribosylation factor-like 5A	2	2	1109	26225	ARL5A	ARL5A_HUMAN	Q9Y689	c	intracellular	179	0
75869	Arl5b	Q9D4P0_MOUSE	Q9D4P0	ADP-ribosylation factor-like 5B	2	2	1110	221079	ARL5B	ARL5B_HUMAN	Q96KC2	c	anchored to membrane	179	0
68724	Arl8a	ARL8A_MOUSE	Q8VEH3	ADP-ribosylation factor-like 8A	5	8	507	127829	ARL8A	ARL8A_HUMAN	Q96BM9	c	late endosome membrane	186	0
67166	Arl8b	ARL8B_MOUSE	Q9CQW2	ADP-ribosylation factor-like 8B	4	8	490	55207	ARL8B	ARL8B_HUMAN	Q9NVJ2	c	anchored to late endosome membrane	186	0

170

Gene ID (mouse)	Gene Symbol (mouse)	Swissprot protein ID (mouse)	Swissprot protein AC (mouse)	Gene Description	Max Diff peptide	Avg Pep count	Rank order	Ortho Gene ID	Ortho Gene Symbol	Swissprot protein ID	Swissprot protein AC	Topology	GO term	Protein length	Number of TM
11867	Arpc1b	ARC1B_MOUSE	Q9WV32	actin related protein 2/3 complex, subunit 1B	6	6	566	10095	ARPC1B	ARC1B_HUMAN	O15143	c	cytoskeleton	372	0
76709	Arpc2	ARPC2_MOUSE	Q9CVB6	actin related protein 2/3 complex, subunit 2	8	15	294	10109	ARPC2	ARPC2_HUMAN	O15144	c	cytoskeleton	300	0
56378	Arpc3	ARPC3_MOUSE	Q9JM76	actin related protein 2/3 complex, subunit 3	4	7	537	10094	ARPC3	ARPC3_HUMAN	O15145	c	cytoskeleton	178	0
68089	Arpc4	ARPC4_MOUSE	P59999	actin related protein 2/3 complex, subunit 4	5	12	361	10093	ARPC4	ARPC4_HUMAN	P59998	c	cytoskeleton	168	0
67771	Arpc5	ARPC5_MOUSE	Q9CPW4	actin related protein 2/3 complex, subunit 5	5	5	673	10092	ARPC5	ARPC5_HUMAN	O15511	c	cytoskeleton	151	0
74192	Arpc5l	ARP5L_MOUSE	Q9D898	actin related protein 2/3 complex, subunit 5-like	2	2	1108	81873	ARPC5L	ARP5L_HUMAN	Q9BPX5	c	cytoskeleton	153	0
27053	Asns	ASNS_MOUSE	Q61024	asparagine synthetase	3	3	848	440	ASNS	ASNS_HUMAN	P08243	c	cytosol	561	0
11946	Atp5a1	ATPA_MOUSE	Q03265	ATP synthase, H+ transporting, mitochondrial F1 complex, alpha subunit, isoform 1	6	11	371	498	ATP5A1	ATPA_HUMAN	P25705	c	associated to mitochondrial inner membrane	553	0
11947	Atp5b	ATPB_MOUSE	P56480	ATP synthase, H+ transporting mitochondrial F1 complex, beta subunit	7	4	689	506	ATP5B	ATPB_HUMAN	P06576	c	integral to mitochondrial inner membrane	529	0
11949	Atp5c1	ATPG_MOUSE	Q91VR2	ATP synthase, H+ transporting, mitochondrial F1 complex, gamma polypeptide 1	5	4	714	509	ATP5C1	ATPG_HUMAN	P36542	c	associated to mitochondrial inner membrane	298	0
71679	Atp5h	ATP5H_MOUSE	Q9DCX2	ATP synthase, H+ transporting, mitochondrial F0 complex, subunit d	2	2	1157	10476	ATP5H	ATP5H_HUMAN	O75947	c	mitochondrial inner membrane	161	0
28080	Atp5o	ATPO_MOUSE	Q9DB20	ATP synthase, H+ transporting, mitochondrial F1 complex, O subunit	3	3	900	539	ATP5O	ATPO_HUMAN	P48047	c	associated to mitochondrial inner membrane	213	0
11972	Atp6v0d1	VA0D1_MOUSE	P51863	ATPase, H+ transporting, lysosomal V0 subunit D1	6	8	464	9114	ATP6V0D1	VA0D1_HUMAN	P61421	c	associated to endomembrane system	351	0
11964	Atp6v1a	VATA_MOUSE	P50516	ATPase, H+ transporting, lysosomal V1 subunit A	6	4	756	523	ATP6V1A	VATA_HUMAN	P38606	c	associated to endomembrane system	617	0
11966	Atp6v1b2	VATB2_MOUSE	P62814	ATPase, H+ transporting, lysosomal V1 subunit B2	3	4	757	526	ATP6V1B2	VATB2_HUMAN	P21281	c	associated to endomembrane system	511	0
73834	Atp6v1d	VATD_MOUSE	P57746	ATPase, H+ transporting, lysosomal V1 subunit D	4	4	796	51382	ATP6V1D	VATD_HUMAN	Q9Y5K8	c	proton-transporting two-sector ATPase complex	247	0
11973	Atp6v1e1	VATE1_MOUSE	P50518	VATPase, H+ transporting, lysosomal V1 subunit E1	7	6	567	529	ATP6V1E1	VATE1_HUMAN	P36543	c	associated to endomembrane system	226	0
381476	B930007M17Rik	CD037_MOUSE	Q8C8J0	Uncharacterized protein C4orf37 homolog	5	9	456	285555	C4orf37	CD037_HUMAN	Q8N412	c	-	534	0
66898	Baiap2l1	BI2L1_MOUSE	Q9DBJ3	BAI1-associated protein 2-like 1	25	55	90	55971	BAIAP2L1	BI2L1_HUMAN	Q9UHR4	c	cellular_component	514	0
207495	Baiap2l2	BI2L2_MOUSE	Q80Y61	BAI1-associated protein 2-like 2	7	7	529	80115	BAIAP2L2	BI2L2_HUMAN	Q6UXY1	c	-	578	0
53817	Bat1a	UAP56_MOUSE	Q9Z1N5	HLA-B-associated transcript 1A	3	4	726	7919	BAT1	UAP56_HUMAN	Q13838	c	nucleus	428	0
66813	Bcl2l14	B2L14_MOUSE	Q9CPT0	Bcl2-like 14 (apoptosis facilitator)	5	3	816	79370	BCL2L14	B2L14_HUMAN	Q9BZR8	c	cytoplasm	328	0
170752	Bco2	BCDO2_MOUSE	Q99NF1	beta-carotene oxygenase 2	4	3	940	83875	BCDO2	BCDO2_HUMAN	Q9BYV7	c	intracellular	532	0
109778	Blvra	BIEA_MOUSE	Q9CY64	biliverdin reductase A	4	3	868	644	BLVRA	BIEA_HUMAN	P53004	c	cytoplasm	295	0
233016	Blvrb	BLVRB_MOUSE	Q923D2	biliverdin reductase B (flavin reductase (NADPH))	4	3	872	645	BLVRB	BLVRB_HUMAN	P30043	c	cytoplasm	206	0
12183	Bpgm	PMGE_MOUSE	P15327	2,3-bisphosphoglycerate mutase	4	4	790	669	BPGM	PMGE_HUMAN	P07738	c	actin cytoskeleton	259	0
23827	Bpnt1	BPNT1_MOUSE	Q9Z0S1	bisphosphate 3'-nucleotidase 1	12	21	235	10380	BPNT1	BPNT1_HUMAN	O95861	c	cytosol	308	0
71678	Brox	BROX_MOUSE	Q8K2Q7	BRO1 domain-containing protein BROX	5	4	771	148362	BROX	BROX_HUMAN	Q5VW32	c	membrane	411	0
238386	Btbd7	BTBD7_MOUSE	Q8CFE5	BTB (POZ) domain containing 7	4	6	581	55727	BTBD7	BTBD7_HUMAN	Q9P203	c	-	164	0
12237	Bub3	BUB3_MOUSE	Q9WVA3	budding uninhibited by benzimidazoles 3 homolog (S. cerevisiae)	5	4	715	9184	BUB3	BUB3_HUMAN	O43684	c	nucleus	326	0
66882	Bzw1	BZW1_MOUSE	Q6CQC6	basic leucine zipper and W2 domains 1	3	2	1025	9689	BZW1	BZW1_HUMAN	Q7L1Q6	c	cytoplasm	419	0
66912	Bzw2	BZW2_MOUSE	Q91VK1	basic leucine zipper and W2 domains 2	2	2	1102	28969	BZW2	BZW2_HUMAN	Q9Y6E2	c	-	201	0
12283	Cab39	CAB39_MOUSE	Q06138	calcium binding protein 39	6	4	716	51719	CAB39	CAB39_HUMAN	Q9Y376	c	cytoplasm	341	0
12301	Cacybp	CYBP_MOUSE	Q9CXW3	calcyclin binding protein	2	2	982	27101	CACYBP	CYBP_HUMAN	Q9HB71	c	cytoplasm	229	0
12313	Calm1	CALM_MOUSE	P62204	calmodulin 1	8	34	147	801	CALM1	CALM_HUMAN	P62158	c	plasma membrane	149	0
12314	Calm2	CALM_MOUSE	P62204	calmodulin 2	8	34	148	805	CALM2	CALM_HUMAN	P62158	c	plasma membrane	149	0
12315	Calm3	CALM_MOUSE	P62204	calmodulin 3	8	34	149	808	CALM3	CALM_HUMAN	P62158	c	plasma membrane	149	0
75600	Calml4	CALL4_MOUSE	Q91WQ9	calmodulin-like 4	15	25	197	91860	CALML4	CALL4_HUMAN	Q96GE6	c	-	567	0
12331	Cap1	CAP1_MOUSE	P40124	CAP, adenylate cyclase-associated protein 1 (yeast)	4	5	661	10487	CAP1	CAP1_HUMAN	Q01518	c	plasma membrane	474	0
12333	Capn1	CAN1_MOUSE	O35350	calpain 1	6	5	662	823	CAPN1	CAN1_HUMAN	P07384	c	plasma membrane	713	0
12337	Capn5	CAN5_MOUSE	O08688	calpain 5	2	2	1089	726	CAPN5	CAN5_HUMAN	O15484	c	cytosol	640	0
12340	Capza1	CAZA1_MOUSE	P47753	capping protein (actin filament) muscle Z-line, alpha 1	6	12	358	829	CAPZA1	CAZA1_HUMAN	P52907	c	actin cytoskeleton	286	0
12343	Capza2	CAZA2_MOUSE	P47754	capping protein (actin filament) muscle Z-line, alpha 2	5	10	402	830	CAPZA2	CAZA2_HUMAN	P47755	c	actin cytoskeleton	286	0
12345	Capzb	CAPZB_MOUSE	P47757	capping protein (actin filament) muscle Z-line, beta	12	15	292	832	CAPZB	CAPZB_HUMAN	P47756	c	actin cytoskeleton	277	0
12350	Car3	CAH3_MOUSE	P16015	carbonic anhydrase 3	2	2	1130	761	CA3	CAH3_HUMAN	P07451	c	cytoplasm	260	0
12362	Casp1	CASP1_MOUSE	P29452	caspase 1	14	18	258	834	CASP1	CASP1_HUMAN	P29466	c	cytoplasm	402	0
12367	Casp3	CASP3_MOUSE	P70677	caspase 3	5	11	391	836	CASP3	CASP3_HUMAN	P42574	c	cytoplasm	277	0
12368	Casp6	CASP6_MOUSE	O08738	caspase 6	8	13	319	839	CASP6	CASP6_HUMAN	P55212	c	cytoplasm	276	0
12369	Casp7	CASP7_MOUSE	P97864	caspase 7	6	7	533	840	CASP7	CASP7_HUMAN	P55210	c	mitochondrial inner membrane	303	0
12370	Casp8	CASP8_MOUSE	O89110	caspase 8	10	9	434	841	CASP8	CASP8_HUMAN	Q14790	c	cytoplasm	480	0
12408	Cbr1	CBR1_MOUSE	P48758	carbonyl reductase 1	10	17	261	873	CBR1	CBR1_HUMAN	P16152	c	cytoplasm	277	0
12461	Cct2	TCPB_MOUSE	P80314	chaperonin subunit 2 (beta)	2	3	930	10576	CCT2	TCPB_HUMAN	P78371	c	cytoplasm	535	0
12462	Cct3	TCPG_MOUSE	P80318	chaperonin subunit 3 (gamma)	5	5	629	7203	CCT3	TCPG_HUMAN	P49368	c	cytoskeleton	545	0
12464	Cct4	TCPD_MOUSE	P80315	chaperonin subunit 4 (delta)	10	10	418	10575	CCT4	TCPD_HUMAN	P50991	c	cytoplasm	539	0
12465	Cct5	TCPE_MOUSE	P80316	chaperonin subunit 5 (epsilon)	2	1	1184	22948	CCT5	TCPE_HUMAN	P48643	c	cytoplasm	541	0
12466	Cct6a	TCPZ_MOUSE	P80317	chaperonin subunit 6a (zeta)	9	9	440	908	CCT6A	TCPZ_HUMAN	P40227	c	cytoplasm	531	0
12468	Cct7	TCPH_MOUSE	P80313	chaperonin subunit 7 (eta)	7	9	463	10574	CCT7	TCPH_HUMAN	Q99832	c	cytoplasm	544	0
72269	Cda	CDD_MOUSE	P56389	cytidine deaminase	2	3	819	978	CDA	CDD_HUMAN	P32320	c	cytoplasm	146	0
12534	Cdc2a	CDC2_MOUSE	P11440	cell division cycle 2 homolog A (S. pombe)	4	5	663	983	CDC2	CDC2_HUMAN	P06493	c	cytosol	297	0
12540	Cdc42	CDC42_MOUSE	P60766	cell division cycle 42 homolog (S. cerevisiae)	7	28	176	643751	CDC42	-	-	c	anchored to plasma membrane	191	0
12566	Cdk2	CDK2_MOUSE	P97377	cyclin-dependent kinase 2	2	2	1131	1017	CDK2	CDK2_HUMAN	P24941	c	cytoplasm	346	0
69681	Cdk3	CDK3_MOUSE	Q80YP0	cyclin-dependent kinase 3	2	2	1155	1018	CDK3	CDK3_HUMAN	Q00526	c	-	274	0
12568	Cdk5	CDK5_MOUSE	P49615	cyclin-dependent kinase 5	3	3	894	1020	CDK5	CDK5_HUMAN	Q00535	c	cytoplasm	292	0
12631	Cfl1	COF1_MOUSE	P18760	cofilin 1, non-muscle	12	37	134	1072	CFL1	COF1_HUMAN	P23528	c	cytoskeleton	166	0
234852	Chmp1a	CHM1A_MOUSE	Q921W0	chromatin modifying protein 1A	4	7	530	5119	CHMP1A	CHM1A_HUMAN	Q9HD42	c	associated to endosome membrane	196	0
67064	Chmp1b	CH1B1_MOUSE	Q99LU0	chromatin modifying protein 1B	2	2	619	57132	CHMP1B	CHM1B_HUMAN	Q7LBR1	c	cytoplasm	199	0
68953	Chmp2a	CHM2A_MOUSE	Q9DB34	chromatin modifying protein 2A	7	12	343	27243	CHMP2A	CHM2A_HUMAN	O43633	c	associated to late endosome membrane	222	0
76608	Chmp4b	CHM4B_MOUSE	Q9DB83	chromatin modifying protein 4B	9	25	201	128866	CHMP4B	CHM4B_HUMAN	Q9H444	c	cytoplasm	224	0
66371	Chmp4c	CHM4C_MOUSE	Q9D7F7	chromatin modifying protein 4C	6	13	332	92421	CHMP4C	CHM4C_HUMAN	Q96CF2	c	cytoplasm	232	0
76959	Chmp5	CHMP5_MOUSE	Q9D7S9	chromatin modifying protein 5	4	5	674	51510	CHMP5	CHMP5_HUMAN	Q9NZZ3	c	cytoplasm	219	0
208092	Chmp6	CHMP6_MOUSE	P0C0A3	chromatin modifying protein 6	4	8	491	79643	CHMP6	CHMP6_HUMAN	Q96FZ7	c	endosome membrane	200	0
56398	Chp	CHP1_MOUSE	P61022	Calcium-binding protein p22	5	5	638	11261	CHP	CHP1_HUMAN	Q99653	c	cytoplasm	195	0
70261	Chp2	CHP2_MOUSE	Q9D869	Calcineurin B homologous protein 2	15	48	109	63928	CHP2	CHP2_HUMAN	O43745	c	-	613	0
23991	Cib1	CIB1_MOUSE	Q9Z0F4	calcium and integrin binding 1 (calmyrin)	11	18	254	10519	CIB1	CIB1_HUMAN	Q99828	c	anchored to endoplasmic reticulum membrane	191	0
12709	Ckb	KCRB_MOUSE	Q04447	creatine kinase, brain	12	37	135	1152	CKB	KCRB_HUMAN	P12277	c	cytoplasm	381	0
114584	Clic1	CLIC1_MOUSE	Q9Z1Q5	chloride intracellular channel 1	13	15	295	1192	CLIC1	CLIC1_HUMAN	O00299	c	integral to apical plasma membrane	241	0
224796	Clic5	CLIC5_MOUSE	Q8BXK9	chloride intracellular channel 5	12	86	62	53405	CLIC5	CLIC5_HUMAN	Q9NZA1	c	actin cytoskeleton	251	0
12757	Clta	CLCA_MOUSE	O08585	clathrin, light polypeptide (Lca)	3	3	837	1211	CLTA	CLCA_HUMAN	P09496	c	clathrin coat	235	0
74325	Cltb	CLCB_MOUSE	Q6IRU5	clathrin, light polypeptide (Lcb)	5	5	641	1212	CLTB	CLCB_HUMAN	P09497	c	clathrin coat	229	0
67300	Cltc	CLH_MOUSE	Q68FD5	clathrin, heavy polypeptide (Hc)	73	303	12	1213	CLTC	CLH_HUMAN	Q00610	c	clathrin coat	1675	0
69574	Cmbl	CMBL_MOUSE	Q8R1G2	carboxymethylenebutenolidase-like (Pseudomonas)	2	2	1030	134147	CMBL	CMBL_HUMAN	Q96DG6	c	-	648	0
22169	Cmpk2	CMPK2_MOUSE	Q3U5Q7	cytidine monophosphate (UMP-CMP) kinase 2, mitochondrial	7	9	435	129607	CMPK2	CMPK2_HUMAN	Q5EBM0	c	mitochondrion	447	0
12785	Cnbp	CNBP_MOUSE	P53996	cellular nucleic acid binding protein	2	2	1132	7555	CNBP	CNBP_HUMAN	P62633	c	cytoplasm	178	0
66054	Cndp2	CNDP2_MOUSE	Q9D1A2	CNDP dipeptidase 2 (metallopeptidase M20 family)	13	18	255	55748	CNDP2	CNDP2_HUMAN	Q96KP4	c	cytoplasm	548	0
12799	Cnp	CN37_MOUSE	P16330	2',3'-cyclic nucleotide 3' phosphodiesterase	3	3	950	1267	CNP	CN37_HUMAN	P09543	c	membrane	420	0
12808	Cobl	COBL_MOUSE	Q5NBX1	cordon-bleu	14	68	71	23242	COBL	COBL_HUMAN	O75128	c	-	487	0
68018	Col4a3bp	C43BP_MOUSE	Q9EQG9	collagen, type IV, alpha 3 (Goodpasture antigen) binding protein	2	3	859	10087	COL4A3BP	C43BP_HUMAN	Q9Y5P4	c	Golgi apparatus	624	0
12238	Commd3	COMD3_MOUSE	Q63829	COMM domain containing 3	2	2	1088	23412	COMMD3	COMD3_HUMAN	Q9UBI1	c	-	340	0
54161	Copg	COPG_MOUSE	Q9QZE5	coatomer protein complex, subunit gamma	2	3	850	22820	COPG	COPG_HUMAN	Q9Y678	c	COPI vesicle coat	874	0
12848	Cops2	CSN2_MOUSE	P61202	COP9 (constitutive photomorphogenic) homolog, subunit 2 (Arabidopsis thaliana)	3	2	985	9318	COPS2	CSN2_HUMAN	P61201	c	cytoplasm	443	0
26754	Cops5	CSN5_MOUSE	O35864	COP9 (constitutive photomorphogenic) homolog, subunit 5 (Arabidopsis thaliana)	2	2	1146	10987	COPS5	CSN5_HUMAN	Q92905	c	cytoplasm	334	0
26893	Cops6	CSN6_MOUSE	O88545	COP9 (constitutive photomorphogenic) homolog, subunit 6 (Arabidopsis thaliana)	2	2	1011	10980	COPS6	CSN6_HUMAN	Q7L5N1	c	cytoplasm	324	0
26894	Cops7a	CSN7A_MOUSE	Q9CZ04	COP9 (constitutive photomorphogenic) homolog, subunit 7a (Arabidopsis thaliana)	2	2	1012	50813	COPS7A	CSN7A_HUMAN	Q9UBW8	c	cytoplasm	275	0
56447	Copz1	COPZ1_MOUSE	P61924	coatomer protein complex, subunit zeta 1	2	2	1149	22818	COPZ1	COPZ1_HUMAN	P61923	c	COPI vesicle coat	177	0
12721	Coro1a	COR1A_MOUSE	O89053	coronin, actin binding protein 1A	2	2	984	11151	CORO1A	COR1A_HUMAN	P31146	c	cytoskeleton	461	0
72042	Cotl1	COTL1_MOUSE	Q9CQI6	coactosin-like 1 (Dictyostelium)	3	3	1036	23406	COTL1	COTL1_HUMAN	Q14019	c	cytoskeleton	142	0
234577	Cpne2	CPNE2_MOUSE	P59108	copine II	4	5	647	221184	CPNE2	CPNE2_HUMAN	Q96FN4	c	-	621	0
227231	Cps1	CPSM_MOUSE	Q8C196	carbamoyl-phosphate synthetase 1	5	5	675	1373	CPS1	CPSM_HUMAN	P31327	c	mitochondrion	1500	0
12904	Crabp2	RABP2_MOUSE	P22935	cellular retinoic acid binding protein II	3	3	838	1382	CRABP2	RABP2_HUMAN	P29373	c	mitochondrion	138	0
214897	Csnk1g1	KC1G1_MOUSE	Q8BTH8	casein kinase 1, gamma 1	2	2	1162	53944	CSNK1G1	KC1G1_HUMAN	Q9HCP0	c	cytoplasm	459	0
70425	Csnk1g3	KC1G3_MOUSE	Q8C4X2	casein kinase 1, gamma 3	11	10	413	1456	CSNK1G3	KC1G3_HUMAN	Q9Y6M4	c	cytoplasm	424	0
13007	Csrp1	CSRP1_MOUSE	P97315	cysteine and glycine-rich protein 1	4	3	839	1465	CSRP1	CSRP1_HUMAN	P21291	c	nucleus	193	0
227292	Ctdsp1	CTDS1_MOUSE	P58466	CTD (carboxy-terminal domain, RNA polymerase II, polypeptide A) small phosphatase 1	4	5	646	58190	CTDSP1	CTDS1_HUMAN	Q9GZU7	c	nucleus	261	0
69274	Ctdspl	CTDSL_MOUSE	P58465	CTD (carboxy-terminal domain, RNA polymerase II, polypeptide A) small phosphatase-like	2	2	1029	10217	CTDSPL	CTDSL_HUMAN	O15194	c	nucleus	276	0
107869	Cth	CGL_MOUSE	Q8VCN5	cystathionase (cystathionine gamma-lyase)	3	3	912	1491	CTH	CGL_HUMAN	P32929	c	cytoplasm	398	0
12388	Ctnnd1	CTND1_MOUSE	P30999	catenin (cadherin associated protein), delta 1	2	2	983	1500	CTNND1	CTND1_HUMAN	O60716	c	plasma membrane	938	0
26554	Cul3	CUL3_MOUSE	Q9JLV5	cullin 3	2	2	1145	8452	CUL3	CUL3_HUMAN	Q13618	c	Golgi apparatus	768	0
109672	Cyb5	CYB5_MOUSE	P56395	cytochrome b-5	2	2	867	1528	CYB5A	CYB5_HUMAN	P00167	c	integral to endoplasmic reticulum membrane	134	1
109754	Cyb5r3	NB5R3_MOUSE	Q9DCN2	cytochrome b5 reductase 3	11	14	305	1727	CYB5R3	NB5R3_HUMAN	P00387	c	anchored to endoplasmic reticulum membrane	301	0
13063	Cycs	CYC_MOUSE	P62897	cytochrome c, somatic	2	1	1185	54205	CYCS	CYC_HUMAN	P99999	c	mitochondrial intermembrane space	105	0
28088	D10Wsu52e	CV028_MOUSE	Q99LF4	DNA segment, Chr 10, Wayne State University 52, expressed	3	3	849	51493	C22orf28	CV028_HUMAN	Q9Y3I0	c	-	192	0

171

Gene ID (mouse)	Gene Symbol (mouse)	Swissprot protein ID (mouse)	Swissprot protein AC (mouse)	Gene Description	Max Diff peptide	Avg Pep count	Rank order	Ortho Gene ID	Ortho Gene Symbol	Swissprot protein ID	Swissprot protein AC	Topology	GO term	Protein length	Number of TM
110957	D1Pas1	DDX3L_MOUSE	P16381	DNA segment, Chr 1, Pasteur Institute 1	8	14	316	8653	DDX3Y	DDX3Y_HUMAN	O15523	c	cytoplasm	660	0
225913	Dak	DHAK_MOUSE	Q8VC30	dihydroxyacetone kinase 2 homolog (yeast)	17	58	84	26007	DAK	DHAK_HUMAN	Q3LXA3	c		1308	0
226414	Dars	SYDC_MOUSE	Q922B2	aspartyl-tRNA synthetase	11	9	443	1615	DARS	SYDC_HUMAN	P14868	c	cytoplasm	501	0
13195	Ddc	DDC_MOUSE	O88533	dopa decarboxylase	2	1	1186	1644	DDC	DDC_HUMAN	P20711	c	-	624	0
13202	Ddt	DOPD_MOUSE	O35215	D-dopachrome tautomerase	2	2	1133	1652	DDT	DOPD_HUMAN	P30046	c		118	0
104721	Ddx1	DDX1_MOUSE	Q91VR5	DEAD (Asp-Glu-Ala-Asp) box polypeptide 1	3	3	911	1653	DDX1	DDX1_HUMAN	Q92499	c	cellular_component	740	0
68278	Ddx39	DDX39_MOUSE	Q8VDW0	DEAD (Asp-Glu-Ala-Asp) box polypeptide 39	4	6	577	10212	DDX39	DDX39_HUMAN	O00148	c	nucleus	427	0
13205	Ddx3x	DDX3X_MOUSE	Q62167	DEAD/H (Asp-Glu-Ala-Asp/His) box polypeptide 3, X-linked	13	36	140	1654	DDX3X	DDX3X_HUMAN	O00571	c	cytoplasm	662	0
26900	Ddx3y	DDX3Y_MOUSE	Q62095	DEAD (Asp-Glu-Ala-Asp) box polypeptide 3, Y-linked	11	6	573	8653	DDX3Y	DDX3Y_HUMAN	O15523	c	cytoplasm	658	0
13207	Ddx5	DDX5_MOUSE	Q61656	DEAD (Asp-Glu-Ala-Asp) box polypeptide 5	4	7	534	1655	DDX5	DDX5_HUMAN	P17844	c	nucleus	614	0
232449	Dera	DEOC_MOUSE	Q9JHU4	2-deoxyribose-5-phosphate aldolase homolog (C. elegans)	11	8	508	51071	DERA	DEOC_HUMAN	Q9Y315	c	cytoplasm	318	0
54722	Dfna5h	DFNA5_MOUSE	Q9Z2D3	deafness, autosomal dominant 5 homolog (human)	3	3	851	1687	DFNA5	DFNA5_HUMAN	O60443	c	-	419	0
13361	Dhfr	DYR_MOUSE	P00375	dihydrofolate reductase	2	2	986	1719	DHFR	DYR_HUMAN	P00374	c	cytoplasm	187	0
52585	Dhrs1	DHRS1_MOUSE	Q99L04	dehydrogenase/reductase (SDR family) member 1	10	10	425	115817	DHRS1	DHRS1_HUMAN	Q96LJ7	c	endoplasmic reticulum	313	0
235339	Dlat	ODP2_MOUSE	Q8BMF4	dihydrolipoamide S-acetyltransferase (E2 component of pyruvate dehydrogenase complex)	2	2	968	1737	DLAT	ODP2_HUMAN	P10515	c	mitochondrion	642	0
15502	Dnaja1	DNJA1_MOUSE	P63037	DnaJ (Hsp40) homolog, subfamily A, member 1	11	11	373	3301	DNAJA1	DNJA1_HUMAN	P31689	c	membrane	397	0
56445	Dnaja2	DNJA2_MOUSE	Q9QYJ0	DnaJ (Hsp40) homolog, subfamily A, member 2	9	8	472	10294	DNAJA2	DNJA2_HUMAN	O60884	c	membrane	412	0
58233	Dnaja4	DNJA4_MOUSE	Q9JMC3	DnaJ (Hsp40) homolog, subfamily A, member 4	2	1	1191	55466	DNAJA4	DNJA4_HUMAN	Q8WW22	c	membrane	397	0
56354	Dnajc7	DNJC7_MOUSE	Q9QYI3	DnaJ (Hsp40) homolog, subfamily C, member 7	5	3	904	7266	DNAJC7	DNJC7_HUMAN	Q99615	c	-	505	0
13430	Dnm2	DYN2_MOUSE	P39054	dynamin 2	6	11	372	1785	DNM2	DYN2_HUMAN	P50570	c	cytoskeleton	870	0
13437	Dnpep	DNPEP_MOUSE	Q9Z2W0	aspartyl aminopeptidase	11	12	351	23549	DNPEP	DNPEP_HUMAN	Q9ULA0	c	cytoplasm	473	0
93838	Dqx1	DQX1_MOUSE	Q924H9	DEAQ RNA-dependent ATPase	2	2	1042	165545	DQX1	DQX1_HUMAN	Q8TE96	c	nucleus	718	0
109620	Dsp	Q8BP77_MOUSE	Q8BP77	desmoplakin homolog	38	59	244	1832	DSP	DESP_HUMAN	P15924	c	cytoskeleton	482	0
56431	Dstn	DEST_MOUSE	Q9R0P5	destrin	12	36	138	11034	DSTN	DEST_HUMAN	P60981	c	actin cytoskeleton	165	0
13424	Dync1h1	DYHC1_MOUSE	Q9JHU4	dynein cytoplasmic 1 heavy chain 1	73	71	69	1778	DYNC1H1	DYHC1_HUMAN	Q14204	c	microtubule	4644	0
235661	Dync1li1	DC1L1_MOUSE	Q8R1Q8	dynein cytoplasmic 1 light intermediate chain 1	2	3	825	51143	DYNC1LI1	DC1L1_HUMAN	Q9Y6G9	c	microtubule	523	0
13627	Eef1a1	EF1A1_MOUSE	P10126	eukaryotic translation elongation factor 1 alpha 1	18	97	56	1915	EEF1A1	EF1A1_HUMAN	P68104	c	cytoplasm	462	0
13628	Eef1a2	EF1A2_MOUSE	P62631	eukaryotic translation elongation factor 1 alpha 2	3	13	320	1917	EEF1A2	EF1A2_HUMAN	Q05639	c	cytoplasm	463	0
66656	Eef1d	EF1D_MOUSE	P57776	eukaryotic translation elongation factor 1 delta (guanine nucleotide exchange protein)	3	2	958	1936	EEF1D	EF1D_HUMAN	P29692	c	cytosol	281	0
67160	Eef1g	EF1G_MOUSE	Q9D8N0	eukaryotic translation elongation factor 1 gamma	5	4	768	1937	EEF1G	EF1G_HUMAN	P26641	c	cytosol	437	0
13629	Eef2	EF2_MOUSE	P58252	eukaryotic translation elongation factor 2	18	52	139	1938	EEF2	EF2_HUMAN	P13639	c	cytoplasm	858	0
27984	Efhd2	EFHD2_MOUSE	Q9D8Y0	EF hand domain containing 2	9	26	194	79180	EFHD2	EFHD2_HUMAN	Q96C19	c	plasma membrane	494	0
76740	Efr3a	EFR3A_MOUSE	Q8BG67	EFR3 homolog A (S. cerevisiae)	3	3	863	23167	EFR3A	EFR3A_HUMAN	Q14156	c	-	819	0
433923	EG433923	-	-	solute carrier family 25 (mitochondrial carrier; adenine nucleotide translocator), member 6	2	2	1053	293	SLC25A6	ADT3_HUMAN	P12236	c	integral to mitochondrion inner membrane	298	2
13660	Ehd1	EHD1_MOUSE	Q9WVK4	EH-domain containing 1	20	27	182	10938	EHD1	EHD1_HUMAN	Q9H4M9	c	anchored to plasma membrane	534	0
98878	Ehd4	EHD4_MOUSE	Q9EQP2	EH-domain containing 4	9	9	442	30844	EHD4	EHD4_HUMAN	Q9H223	c	endoplasmic reticulum	541	0
74147	Ehhadh	ECHP_MOUSE	Q9DBM2	enoyl-Coenzyme A, hydratase/3-hydroxyacyl Coenzyme A dehydrogenase	5	3	861	1962	EHHADH	ECHP_HUMAN	Q08426	c	mitochondrion	718	0
13681	Eif4a1	IF4A1_MOUSE	P60843	eukaryotic translation initiation factor 4A1	17	28	173	1973	EIF4A1	IF4A1_HUMAN	P60842	c	cytoplasm	406	0
13682	Eif4a2	IF4A2_MOUSE	P10630	eukaryotic translation initiation factor 4A2	2	3	840	1974	EIF4A2	IF4A2_HUMAN	Q14240	c	cytosol	407	0
217869	Eif5	IF5_MOUSE	P59325	eukaryotic translation initiation factor 5	2	2	1163	1983	EIF5	IF5_HUMAN	P55010	c	cytoplasm	429	0
16418	Eif6	IF6_MOUSE	O55135	eukaryotic translation initiation factor 6	5	5	664	3692	EIF6	IF6_HUMAN	P56537	c	cytoplasm	245	0
13806	Eno1	ENOA_MOUSE	P17182	enolase 1, alpha non-neuron	17	47	111	2023	ENO1	ENOA_HUMAN	P06733	c	plasma membrane	434	0
54357	Epb4.1l4b	E41LB_MOUSE	Q9JMC8	erythrocyte protein band 4.1-like 4b	5	5	687	54566	EPB41L4B	E41LB_HUMAN	Q9H329	c	cytoskeleton	527	0
226352	Epb4.1l5	E41L5_MOUSE	Q8BGS1	erythrocyte protein band 4.1-like 5	2	2	1164	57669	EPB41L5	E41L5_HUMAN	Q9HCM4	c	cytoskeleton	731	0
13860	Eps8	EPS8_MOUSE	Q08509	epidermal growth factor receptor pathway substrate 8	11	49	103	2059	EPS8	EPS8_HUMAN	Q12929	c	-	574	0
98845	Eps8l2	ESBL2_MOUSE	Q99K30	EPS8-like 2	16	48	107	64787	EPS8L2	ESBL2_HUMAN	Q9H6S3	c	cytoskeleton	729	0
99662	Eps8l3	ESBL3_MOUSE	Q91WL0	EPS8-like 3	28	140	31	79574	EPS8L3	ESBL3_HUMAN	Q8TE67	c	cytoskeleton	600	0
13885	Esd	ESTD_MOUSE	Q9R0P3	esterase D/formylglutathione hydrolase	4	4	690	2098	ESD	ESTD_HUMAN	P10768	c	cytoplasmic membrane-bounded vesicle	282	0
56226	Espn	ESPN_MOUSE	Q9ET47	espin	2	3	853	83715	ESPN	ESPN_HUMAN	B1AK53	c	cytoskeleton	871	0
225363	Etf1	ERF1_MOUSE	Q8BWY3	eukaryotic translation termination factor 1	2	2	1046	2107	ETF1	ERF1_HUMAN	P62495	c	cytoplasm	437	0
110826	Etfb	ETFB_MOUSE	Q9DCW4	electron transferring flavoprotein, beta polypeptide	4	3	869	2109	ETFB	ETFB_HUMAN	P38117	c	mitochondrion	255	0
22350	Ezr	EZRI_MOUSE	P26040	ezrin	37	131	35	7430	EZR	EZRI_HUMAN	P15311	c	cytoskeleton	586	0
14080	Fabp1	FABPL_MOUSE	P12710	fatty acid binding protein 1, liver	9	27	183	2168	FABP1	FABPL_HUMAN	P07148	c	cytoplasm	127	0
14079	Fabp2	FABPI_MOUSE	P55050	fatty acid binding protein 2, intestinal	10	19	249	2169	FABP2	FABPI_HUMAN	P12104	c	cytoplasm	132	0
16592	Fabp5	FABP5_MOUSE	Q05816	fatty acid binding protein 5, epidermal	2	1	1188	2171	FABP5	FABP5_HUMAN	Q01469	c	cytoplasm	135	0
16204	Fabp6	FABP6_MOUSE	P51162	fatty acid binding protein 6, ileal (gastrotropin)	3	5	614	2172	FABP6	FABP6_HUMAN	P51161	c	cytoplasm	128	0
216169	Fam108a	F108A_MOUSE	Q99JW1	Abhydrolase domain-containing protein FAM108A	5	5	645	81926	FAM108A3	F1A8A1_HUMAN	Q96GS6	c	extracellular region	310	0
226016	Fam108b	F108B_MOUSE	Q7M759	Abhydrolase domain-containing protein FAM108B1 (Precursor)	4	2	966	51104	FAM108B1	F108B_HUMAN	Q5VST6	c	extracellular region	288	0
66590	Farsa	SYFA_MOUSE	Q8C0C7	phenylalanyl-tRNA synthetase, alpha subunit	4	3	857	2193	FARSA	SYFA_HUMAN	Q9Y285	c	cytoplasm	508	0
14120	Fbp2	F16P2_MOUSE	P70695	fructose bisphosphatase 2	4	4	691	8789	FBP2	F16P2_HUMAN	O00757	c	cytosol	339	0
110196	Fdps	FPPS_MOUSE	Q920E5	farnesyl diphosphate synthetase	4	4	797	2224	FDPS	FPPS_HUMAN	P14324	c	cytoplasm	353	0
14194	Fh1	FUMH_MOUSE	P97807	fumarate hydratase 1	4	4	758	2271	FH	FUMH_HUMAN	P07954	c	cytoplasm	507	0
14199	Fhl1	FHL1_MOUSE	P97447	four and a half LIM domains 1	2	1	1187	2273	FHL1	FHL1_HUMAN	Q13642	c	cytoplasm	280	0
14228	Fkbp4	FKBP4_MOUSE	P30416	FK506 binding protein 4	7	10	419	2288	FKBP4	FKBP4_HUMAN	Q02790	c	cytoplasm	458	0
286940	Flnb	FLNB_MOUSE	Q80X90	filamin, beta	14	9	444	2317	FLNB	FLNB_HUMAN	O75369	c	cytoskeleton	2602	0
67457	Frmd8	FRMD8_MOUSE	Q3UFK8	FERM domain containing 8	2	2	1154	83786	FRMD8	FRMD8_HUMAN	Q9BZ67	c	cytoskeleton	466	0
14381	G6pdx	G6PD1_MOUSE	Q00612	glucose-6-phosphate dehydrogenase X-linked	3	2	987	2539	G6PD	G6PD1_HUMAN	P11413	c	cytoplasm	515	0
74246	Gale	GALE_MOUSE	Q8R059	galactose-4-epimerase, UDP	14	14	303	2582	GALE	GALE_HUMAN	Q14376	c	cytoplasm	347	0
14635	Galk1	GALK1_MOUSE	Q9R0N0	galactokinase 1	6	5	613	2584	GALK1	GALK1_HUMAN	P51570	c	cytoplasm	391	0
319625	Galm	GALM_MOUSE	Q8K157	galactose mutarotase	2	2	1049	130589	GALM	GALM_HUMAN	Q96C23	c	cytoplasm	342	0
14630	Gclm	GSH0_MOUSE	O09172	glutamate-cysteine ligase , modifier subunit	5	5	630	2730	GCLM	GSH0_HUMAN	P48507	c	cytosol	274	0
14569	Gdi2	GDIB_MOUSE	Q61598	guanosine diphosphate (GDP) dissociation inhibitor 2	15	21	225	2665	GDI2	GDIB_HUMAN	P50395	c	associated to membrane	445	0
54120	Gipc2	GIPC2_MOUSE	Q9Z2H7	GIPC PDZ domain containing family, member 2	11	9	436	54810	GIPC2	GIPC2_HUMAN	Q8TF65	c	cytoplasm	314	0
384009	Glipr2	GAPR1_MOUSE	Q9CYL5	GLI pathogenesis-related 2	2	2	1052	152007	C9orf19	GAPR1_HUMAN	Q9H4G4	c	associated to Golgi membrane	154	0
109801	Glo1	LGUL_MOUSE	Q9CPU0	glyoxalase 1	3	2	1113	2739	GLO1	LGUL_HUMAN	Q04760	c	cytoplasm	184	0
69824	Glod5	GLOD5_MOUSE	Q9DBI3	glyoxalase domain containing 5	6	4	696	392465	GLOD5	GLOD5_HUMAN	A6NK44	c	-	475	0
30926	Glrx3	GLRX3_MOUSE	Q9CQM9	glutaredoxin 3	3	3	901	10539	GLRX3	GLRX3_HUMAN	O76003	c	cytoplasm	337	0
330440	Gm766	Q3UN63_MOUSE	Q3UN63	hypothetical Sulfotransferase	2	2	1118	-	-	-	-	c	-	171	0
218138	Gmds	GMDS_MOUSE	Q8K0C9	GDP-mannose 4, 6-dehydratase	13	11	388	2762	GMDS	GMDS_HUMAN	O60547	c	cytoplasm	372	0
14672	Gna11	GNA11_MOUSE	P21278	guanine nucleotide binding protein, alpha 11	28	236	20	2767	GNA11	GNA11_HUMAN	P29992	c	associated to plasma membrane	359	0
14673	Gna12	GNA12_MOUSE	P27600	guanine nucleotide binding protein, alpha 12	2	2	951	2768	GNA12	GNA12_HUMAN	Q03113	c	plasma membrane	379	0
14674	Gna13	GNA13_MOUSE	P27601	guanine nucleotide binding protein, alpha 13	19	43	119	10672	GNA13	GNA13_HUMAN	Q14344	c	membrane	377	0
14675	Gna14	GNA14_MOUSE	P30677	guanine nucleotide binding protein, alpha 14	7	58	85	9630	GNA14	GNA14_HUMAN	O95837	c	plasma membrane	355	0
14681	Gnao1	GNAO_MOUSE	P18872	guanine nucleotide binding protein, alpha O	5	24	204	2775	GNAO1	GNAO_HUMAN	P09471	c	membrane fraction	354	0
14682	Gnaq	GNAQ_MOUSE	P21279	guanine nucleotide binding protein, alpha q polypeptide	20	115	40	2776	GNAQ	GNAQ_HUMAN	P50148	c	associated to plasma membrane	353	0
14688	Gnb1	GBB1_MOUSE	P62874	guanine nucleotide binding protein (G protein), beta 1	15	116	38	2782	GNB1	GBB1_HUMAN	P62873	c	associated to membrane	130	0
14693	Gnb2	GBB2_MOUSE	P62880	guanine nucleotide binding protein (G protein), beta 2	12	107	46	2783	GNB2	GBB2_HUMAN	P62879	c	associated to membrane	493	0
50798	Gne	GLCNE_MOUSE	Q91WG8	glucosamine	4	3	814	10020	GNE	GLCNE_HUMAN	Q9Y223	c	cytoplasm	722	0
26384	Gnpda1	GNPI_MOUSE	O88958	glucosamine-6-phosphate deaminase 1	5	4	723	10007	GNPDA1	GNPI_HUMAN	P46926	c	cytoplasm	289	0
57437	Golga7	GOGA7_MOUSE	Q9JW53	golgi autoantigen, golgin subfamily a, 7	3	4	766	51125	GOLGA7	GOGA7_HUMAN	Q7Z5G4	c	Golgi apparatus	137	0
66629	Golph3	GOLP3_MOUSE	Q9CRA5	golgi phosphoprotein 3	4	4	733	64083	GOLPH3	GOLP3_HUMAN	Q9H4A6	c	associated to Golgi membrane	298	0
229593	Golph3l	GLP3L_MOUSE	Q8R088	golgi phosphoprotein 3-like	4	4	700	55204	GOLPH3L	GLP3L_HUMAN	Q9H4A5	c	associated to Golgi membrane	285	0
14718	Got1	AATC_MOUSE	P05201	glutamate oxaloacetate transaminase 1, soluble	4	2	952	2805	GOT1	AATC_HUMAN	P17174	c	cytoplasm	413	0
14555	Gpd1	GPDA_MOUSE	P13707	glycerol-3-phosphate dehydrogenase 1 (soluble)	15	22	219	2819	GPD1	GPDA_HUMAN	P21695	c	cytoplasm	349	0
333433	Gpd1l	GPD1L_MOUSE	Q3ULJ0	glycerol-3-phosphate dehydrogenase 1-like	9	7	542	23171	GPD1L	GPD1L_HUMAN	Q8N335	c	cytoplasm	351	0
76282	Gpt	ALAT1_MOUSE	Q8QZR5	glutamic pyruvic transaminase, soluble	7	6	597	2875	GPT	ALAT1_HUMAN	P24298	c	cytoplasm	496	0
14775	Gpx1	GPX1_MOUSE	P11352	glutathione peroxidase 1	7	8	485	2876	GPX1	GPX1_HUMAN	P07203	c	cytoplasm	201	0
14776	Gpx2	GPX2_MOUSE	Q9JHC0	glutathione peroxidase 2	15	19	243	2877	GPX2	GPX2_HUMAN	P18283	c	cytoplasm	190	0
14784	Grb2	GRB2_MOUSE	Q60631	growth factor receptor bound protein 2	7	3	806	2885	GRB2	GRB2_HUMAN	P62993	c	Golgi apparatus	217	0
69146	Gsdmd	GSDC1_MOUSE	Q9D8T2	gasdermin D	10	10	256	79792	GSDMDC1	GSDC1_HUMAN	P57764	c	-	420	0
14854	Gss	GSHB_MOUSE	P51855	glutathione synthetase	2	2	988	2937	GSS	GSHB_HUMAN	P48637	c	-	512	0
14857	Gsta1	GSTA1_MOUSE	Q7TRV2	glutathione S-transferase, alpha 1 (Ya)	6	8	466	21357 ; 2953	GSTA1 ; GSTA5	GSTA5_HUMAN	Q7RTV2	c	cytoplasm	223	0
14858	Gsta2	GSTA2_MOUSE	P10648	glutathione S-transferase, alpha 2 (Yc2)	6	8	466	221345	GSTA5	GSTA5_HUMAN	Q7RTV2	c	cytoplasm	222	0
14860	Gsta4	GSTA4_MOUSE	P24472	glutathione S-transferase, alpha 4	2	2	1134	-	-	-	-	c	cytoplasm	222	0
14862	Gstm1	GSTM1_MOUSE	P10649	glutathione S-transferase, mu 1	6	6	717	2944	GSTM1	GSTM1_HUMAN	P09488	c	cytoplasm	218	0
14864	Gstm3	GSTM4_MOUSE	P19639	glutathione S-transferase, mu 3	11	7	523	-	-	-	-	c	cytoplasm	218	0

172

Gene ID (mouse)	Gene Symbol (mouse)	Swissprot protein ID (mouse)	Swissprot protein AC (mouse)	Gene Description	Max Diff peptide	Avg Pep count	Rank order	Ortho Gene ID	Ortho Gene Symbol	Swissprot protein ID	Swissprot protein AC	Topology	GO term	Protein length	Number of TM
14873	Gsto1	GSTO1_MOUSE	O09131	glutathione S-transferase omega 1	7	6	594	9446	GSTO1	GSTO1_HUMAN	P78417	c	cytoplasm	240	0
14870	Gstp1	GSTP1_MOUSE	P19157	glutathione S-transferase, pi 1	6	9	451	2950	GSTP1	GSTP1_HUMAN	P09211	c	cytoplasm	210	0
14871	Gstt1	GSTT1_MOUSE	Q64471	glutathione S-transferase, theta 1	2	2	1090	2952	GSTT1	GSTT1_HUMAN	P30711	c	cytoplasm	240	0
103140	Gstt3	Q6P6I4_MOUSE	Q6P6I4	glutathione S-transferase, theta 3	5	4	698	25774	GSTTP1	GSTT4_HUMAN	A8MPT4	c	cytoplasm	241	0
14874	Gstz1	MAAI_MOUSE	Q9WVL0	glutathione transferase zeta 1 (maleylacetoacetate isomerase)	2	2	989	2954	GSTZ1	MAAI_HUMAN	O43708	c	cytoplasm	216	0
15107	Hadh	HCDH_MOUSE	Q61425	hydroxyacyl-Coenzyme A dehydrogenase	4	3	842	3033	HADH	HCDH_HUMAN	Q16836	c	mitochondrial inner membrane	314	0
231086	Hadhb	ECHB_MOUSE	Q99JY0	hydroxyacyl-Coenzyme A dehydrogenase/3-ketoacyl-Coenzyme A thiolase/enoyl-Coenzyme A hydratase (trifunctional protein), beta subunit	4	3	871	3032	HADHB	ECHB_HUMAN	P55084	c	cytosol	475	0
14651	Hagh	GLO2_MOUSE	Q99KB8	hydroxyacyl glutathione hydrolase	5	6	610	3029	HAGH	GLO2_HUMAN	Q16775	c	-	245	0
15109	Hal	HUTH_MOUSE	P35492	histidine ammonia lyase	4	4	718	3034	HAL	HUTH_HUMAN	P42357	c	cytoplasm	657	0
15115	Hars	SYHC_MOUSE	Q61035	histidyl-tRNA synthetase	2	2	1091	3035	HARS	SYHC_HUMAN	P12081	c	cytoplasm	509	0
107435	Hat1	HAT1_MOUSE	Q8BY71	histone aminotransferase 1	2	2	1112	8520	HAT1	HAT1_HUMAN	O14929	c	cytoplasm	416	0
15122	Hba-a1	HBA_MOUSE	P01942	hemoglobin alpha, adult chain 1	4	7	524	3039	HBA1	HBA_HUMAN	P69905	c	hemoglobin complex	142	0
15129	Hbb-b1	HBB1_MOUSE	P02088	hemoglobin, beta adult major chain	11	21	226	3043	HBB	HBB_HUMAN	P68871	c	hemoglobin complex	147	0
15130	Hbb-b2	HBB1_MOUSE	P02088	hemoglobin, beta adult minor chain	11	21	227	3043	HBB	HBB_HUMAN	P68871	c	hemoglobin complex	147	0
15254	Hint1	HINT1_MOUSE	P70349	histidine triad nucleotide binding protein 1	3	5	631	3094	HINT1	HINT1_HUMAN	P49773	c	cytoplasm	126	0
29816	Hip1r	HIP1R_MOUSE	Q9JKY5	huntingtin interacting protein 1 related	2	3	813	9026	HIP1R	HIP1R_HUMAN	O75146	c	clathrin-coated vesicle	1068	0
216019	Hkdc1	HKDC1_MOUSE	Q91W97	hexokinase domain containing 1	3	2	964	80201	HKDC1	HKDC1_HUMAN	Q2TB90	c	-	267	0
15289	Hmgb1	HMGB1_MOUSE	P63158	high mobility group box 1	4	4	791	10357	HMG1L1	-	-	c	cytoplasm	215	0
53379	Hnrnpa2b1	ROA2_MOUSE	O88569	heterogeneous nuclear ribonucleoprotein A2/B1	4	10	426	3181	HNRNPA2B1	ROA2_HUMAN	P22626	c	nucleus	353	0
98758	Hnrnpf	HNRPF_MOUSE	Q9Z2X1	heterogeneous nuclear ribonucleoprotein F	5	4	773	3185	HNRNPF	HNRPF_HUMAN	P52597	c	nucleus	415	0
59013	Hnrnph1	HNRH1_MOUSE	O35737	heterogeneous nuclear ribonucleoprotein H1	2	3	905	3187	HNRNPH1	HNRH1_HUMAN	P31943	c	nucleus	449	0
56258	Hnrnph2	HNRH2_MOUSE	P70333	heterogeneous nuclear ribonucleoprotein H2	2	3	903	3188	HNRNPH2	HNRH2_HUMAN	P55795	c	nucleus	449	0
15387	Hnrnpk	HNRPK_MOUSE	P61979	heterogeneous nuclear ribonucleoprotein K	2	3	843	3190	HNRNPK	HNRPK_HUMAN	P61978	c	nucleus	463	0
15444	Hpca	HPCA_MOUSE	P84075	hippocalcin	2	2	991	3208	HPCA	HPCA_HUMAN	P84074	c	anchored to membrane	480	0
53602	Hpcal1	HPCL1_MOUSE	P62748	hippocalcin-like 1	2	2	1015	3241	HPCAL1	HPCL1_HUMAN	P37235	c	anchored to membrane	364	0
15446	Hpgd	PGDH_MOUSE	Q8VCC1	hydroxyprostaglandin dehydrogenase 15 (NAD)	7	4	719	3248	HPGD	PGDH_HUMAN	P15428	c	cytoplasm	269	0
15452	Hprt1	HPRT_MOUSE	P00493	hypoxanthine guanine phosphoribosyl transferase 1	4	4	720	3251	HPRT1	HPRT_HUMAN	P00492	c	cytoplasm	218	0
15461	Hras1	RASH_MOUSE	Q61411	Harvey rat sarcoma virus oncogene 1	2	2	1135	3265	HRAS	RASH_HUMAN	P01112	c	anchored to plasma membrane	189	0
193740	Hspa1a	HS71A_MOUSE	Q61696	heat shock protein 1A	12	13	326	3303	HSPA1A	HSP71_HUMAN	P08107	c	cytoplasm	641	0
15511	Hspa1b	HS71B_MOUSE	P17879	heat shock protein 1B	12	13	324	3303	HSPA1A	HSP71_HUMAN	P08107	c	cytoplasm	642	0
15512	Hspa2	HSP72_MOUSE	P17156	heat shock protein 2	2	3	931	3306	HSPA2	HSP72_HUMAN	P54652	c	cytoplasm	633	0
15481	Hspa8	HSP7C_MOUSE	P63017	heat shock protein 8	33	104	48	3312	HSPA8	HSP7C_HUMAN	P11142	c	cytoplasm	646	0
15505	Hsph1	HS105_MOUSE	Q61699	heat shock 105kDa/110kDa protein 1	2	2	992	10808	HSPH1	HS105_HUMAN	Q92598	c	cytoplasm	858	0
15926	Idh1	IDHC_MOUSE	O88844	isocitrate dehydrogenase 1 (NADP+), soluble	18	29	169	3417	IDH1	IDHC_HUMAN	O75874	c	cytoplasm	414	0
269951	Idh2	IDHP_MOUSE	P54071	isocitrate dehydrogenase 2 (NADP+), mitochondrial	9	12	362	3418	IDH2	IDHP_HUMAN	P48735	c	mitochondrial inner membrane	452	0
67834	Idh3a	IDH3A_MOUSE	Q9D6R2	isocitrate dehydrogenase 3 (NAD+) alpha	3	3	938	3419	IDH3A	IDH3A_HUMAN	P50213	c	mitochondrial matrix	366	0
170718	Idh3b	-	-	isocitrate dehydrogenase 3 (NAD+) beta	3	2	1045	3420	IDH3B	IDH3B_HUMAN	O43837	c	mitochondrion	385	0
15929	Idh3g	IDH3G_MOUSE	P70404	isocitrate dehydrogenase 3 (NAD+), gamma	3	2	993	3421	IDH3G	IDH3G_HUMAN	P51553	c	mitochondrion	393	0
319554	Idi1	IDI1_MOUSE	P58044	isopentenyl-diphosphate delta isomerase	2	2	1168	3422	IDI1	IDI1_HUMAN	Q13907	c	cytoplasm	227	0
15953	Ifi47	Q61635_MOUSE	Q61635	interferon gamma inducible protein 47	8	6	595	-	-	-	c	endoplasmic reticulum	420	0
15957	Ifit1	IFIT1_MOUSE	Q64282	interferon-induced protein with tetratricopeptide repeats 1	10	11	374	3434	IFIT1	IFIT1_HUMAN	P09914	c	cytoplasm	463	0
54396	Iigp2	Q58E63_MOUSE	Q58E63	interferon inducible GTPase 2	11	12	342	-	-	-	c	-	435	0
16201	Ilf3	ILF3_MOUSE	Q9Z1X4	interleukin enhancer binding factor 3	6	11	375	3609	ILF3	ILF3_HUMAN	Q12906	c	nucleus	898	0
55980	Impa1	IMPA1_MOUSE	O55023	inositol (myo)-1(or 4)-monophosphatase 1	5	4	728	3612	IMPA1	IMPA1_HUMAN	P29218	c	cytoplasm	277	0
12695	Inadl	INADL_MOUSE	Q63ZW7	InaD-like (Drosophila)	5	15	293	10207	INADL	INADL_HUMAN	Q8NI35	c	apical plasma membrane	1834	0
29875	Iqgap1	IQGA1_MOUSE	Q9JKF1	IQ motif containing GTPase activating protein 1	14	12	354	8826	IQGAP1	IQGA1_HUMAN	P46940	c	associated to plasma membrane	1657	0
544963	Iqgap2	fragment	-	IQ motif containing GTPase activating protein 2	23	39	127	10788	IQGAP2	IQGA2_HUMAN	Q13576	c	actin cytoskeleton	1575	0
631323	Irgb10	Q0GUM3_MOUSE	Q0GUM3	interferon-gamma-inducible p47 GTPase	14	13	327	-	-	-	c	-	417	0
15944	Irgm	IRGM_MOUSE	Q60766	immunity-related GTPase family, M	9	18	252	345611	IRGM	IRGM_HUMAN	A1A4Y4	c	associated to Golgi membrane	409	0
16480	Jup	PLAK_MOUSE	Q02257	junction plakoglobin	16	12	352	3728	JUP	PLAK_HUMAN	P14923	c	actin cytoskeleton	745	0
107351	Kank1	Q6AXG6_MOUSE	Q6AXG6	KN motif and ankyrin repeat domains 1	2	3	864	23189	KANK1	KANK1_HUMAN	Q14678	c	cytoplasm	1202	0
16572	Kif5a	KIF5A_MOUSE	P33175	kinesin family member 5A	3	3	807	3798	KIF5A	KIF5A_HUMAN	Q12840	c	cytoplasm	1027	0
16573	Kif5b	KINH_MOUSE	Q61768	kinesin family member 5B	3	3	808	3799	KIF5B	KINH_HUMAN	P33176	c	microtubule associated complex	963	0
16574	Kif5c	KIF5C_MOUSE	P28738	kinesin family member 5C	3	3	809	3800	KIF5C	KIF5C_HUMAN	O60282	c	cytoskeleton	956	0
16211	Kpnb1	IMB1_MOUSE	P70168	karyopherin (importin) beta 1	2	2	953	3837	KPNB1	IMB1_HUMAN	Q14974	c	cytoplasm	876	0
16653	Kras	RASK_MOUSE	P32883	v-Ki-ras2 Kirsten rat sarcoma viral oncogene homolog	7	10	420	3845	KRAS	RASK_HUMAN	P01116	c	anchored to plasma membrane	189	0
16678	Krt1	K2C1_MOUSE	P04104	keratin 1	4	25	195	3848	KRT1	K2C1_HUMAN	P04264	c	cytoskeleton	637	0
16663	Krt13	K1C13_MOUSE	P08730	keratin 13	5	16	279	3860	KRT13	K1C13_HUMAN	P13646	c	intermediate filament	437	0
16664	Krt14	K1C14_MOUSE	Q61781	keratin 14	18	28	175	3861	KRT14	K1C14_HUMAN	P02533	c	cytoskeleton	484	0
16667	Krt17	K1C17_MOUSE	Q9QWL7	keratin 17	5	13	321	3872	KRT17	K1C17_HUMAN	Q04695	c	intermediate filament	433	0
16669	Krt19	K1C19_MOUSE	P19001	keratin 19	8	18	259	3880	KRT19	K1C19_HUMAN	P08727	c	intermediate filament	403	0
16681	Krt2	K22E_MOUSE	Q3TTY5	keratin 2	5	12	349	3849	KRT2	K22E_HUMAN	P35908	c	intermediate filament	707	0
68239	Krt42	K1C42_MOUSE	Q6IFX2	keratin 42	15	27	189	-	-	-	c	intermediate filament	452	0
110308	Krt5	K2C5_MOUSE	Q922U2	keratin 5	14	16	285	3852	KRT5	K2C5_HUMAN	P13647	c	intermediate filament	580	0
77055	Krt76	K2C7_MOUSE	Q3UV17	keratin 76	6	17	266	51350	KRT76	K2O_HUMAN	Q01546	c	intermediate filament	594	0
16691	Krt8	K2C8_MOUSE	P11679	keratin 8	8	33	152	3856	KRT8	K2C8_HUMAN	P05787	c	cytoplasm	490	0
16763	Lad1	LAD1_MOUSE	P57016	ladinin	11	49	105	3898	LAD1	LAD1_HUMAN	O00515	c	basement membrane	528	0
71835	Lancl2	LANC2_MOUSE	Q9JJK2	LanC (bacterial lantibiotic synthetase component C)-like 2	2	2	1158	55915	LANCL2	LANC2_HUMAN	Q9NS86	c	plasma membrane	450	0
66988	Lap3	AMPL_MOUSE	Q9CPY7	leucine aminopeptidase 3	16	25	199	51056	LAP3	AMPL_HUMAN	P28838	c	cytoplasm	519	0
16796	Lasp1	LASP1_MOUSE	Q61792	LIM and SH3 protein 1	12	45	115	3927	LASP1	LASP1_HUMAN	Q14847	c	cortical actin cytoskeleton	263	0
30949	Lcmt1	A2RTH5_MOUSE	A2RTH5	leucine carboxyl methyltransferase 1	4	4	793	51451	LCMT1	LCMT1_HUMAN	Q9UIC8	c	-	321	0
24131	Ldb3	LDB3_MOUSE	Q9JKS4	LIM domain binding 3	2	2	1008	11155	LDB3	LDB3_HUMAN	O75112	c	cytoskeleton	723	0
107753	Lgals2	LEG2_MOUSE	Q9CQW5	lectin, galactose-binding, soluble 2	10	36	139	3957	LGALS2	LEG2_HUMAN	P05162	c	-	254	0
16854	Lgals3	LEG3_MOUSE	P16110	lectin, galactose binding, soluble 3	5	8	467	3958	LGALS3	LEG3_HUMAN	P17931	c	plasma membrane	264	0
16855	Lgals4	LEG4_MOUSE	Q8K419	lectin, galactose binding, soluble 4	11	80	65	3960	LGALS4	LEG4_HUMAN	P56470	c	plasma membrane	326	0
16859	Lgals9	LEG9_MOUSE	O08573	lectin, galactose binding, soluble 9	7	21	233	284194	LOC284194	LE9LA_HUMAN	Q3B8N2	c	extracellular region	353	0
65970	Lima1	LIMA1_MOUSE	Q9ERG0	LIM domain and actin binding 1	18	136	33	51474	LIMA1	LIMA1_HUMAN	Q9UHB6	c	actin cytoskeleton	753	0
22343	Lin7c	LIN7C_MOUSE	O88952	lin-7 homolog C (C. elegans)	4	4	693	55327	LIN7C	LIN7C_HUMAN	Q9NUP9	c	associated to basolateral plasma membrane	197	0
214854	Lincr	LINCR_MOUSE	Q8CJC5	lung-inducible neuralized-related C3HC4 RING domain protein	2	2	1114	93082	LINCR	LINCR_HUMAN	Q96EH8	c	-	483	0
16890	Lipe	LIPS_MOUSE	P54310	lipase, hormone sensitive	2	2	994	3991	LIPE	LIPS_HUMAN	Q05469	c	membrane	759	0
214345	Lrrc1	LRRC1_MOUSE	Q80VQ1	leucine rich repeat containing 1	2	2	1161	55227	LRRC1	LRRC1_HUMAN	Q9BTT6	c	associated to membrane	524	0
66508	Lrrc51	LRC51_MOUSE	Q9DAK8	Leucine-rich repeat-containing protein 51	3	3	815	220074	LRRC51	LRC51_HUMAN	Q96E66	c	-	104	0
66606	Lrrc57	LRC57_MOUSE	Q9D1G5	leucine rich repeat containing 57	7	7	540	255252	LRRC57	LRC57_HUMAN	Q8N9N7	c	-	271	0
16993	Lta4h	LKHA4_MOUSE	P24527	leukotriene A4 hydrolase	25	32	154	4048	LTA4H	LKHA4_HUMAN	P09960	c	cytoplasm	611	0
17096	Lyn	LYN_MOUSE	P25911	Tyrosine-protein kinase Lyn	13	25	198	4067	LYN	LYN_HUMAN	P07948	c	anchored to plasma membrane	512	0
18777	Lypla1	LYPA1_MOUSE	P97823	lysophospholipase 1	3	7	519	10434	LYPLA1	LYPA1_HUMAN	O75608	c	cytoplasm	230	0
26394	Lypla2	LYPA2_MOUSE	Q9WTL7	lysophospholipase 2	2	2	1142	11313	LYPLA2	LYPA2_HUMAN	O95372	c	cytoplasm	231	0
66905	M6prbp1	M6PBP_MOUSE	Q9DBG5	mannose-6-phosphate receptor binding protein 1	3	3	858	10226	M6PRBP1	M6PBP_HUMAN	O60664	c	associated to endosome membrane	437	0
109731	Maob	AOFB_MOUSE	Q8BW75	monoamine oxidase B	6	5	623	4129	MAOB	AOFB_HUMAN	P27338	c	integral to mitochondrial membrane	526	0
26395	Map2k1	MP2K1_MOUSE	P31938	mitogen-activated protein kinase kinase 1	6	7	528	5604	MAP2K1	MP2K1_HUMAN	Q02750	c	cytosol	393	0
26397	Map2k3	MP2K3_MOUSE	O09110	mitogen-activated protein kinase kinase 3	2	2	1143	5606	MAP2K3	MP2K3_HUMAN	P46734	c	cytoplasm	148	0
71751	Map3k13	M3K13_MOUSE	Q1HKZ5	mitogen-activated protein kinase kinase kinase 13	2	2	1035	9175	MAP3K13	M3K13_HUMAN	O43283	c	associated to membrane	959	0
26921	Map4k4	M4K4_MOUSE	P97820	mitogen-activated protein kinase kinase kinase kinase 4	2	2	1095	9448	MAP4K4	M4K4_HUMAN	O95819	c	cellular_component	1233	0
83409	Mapbpip	MAPIP_MOUSE	Q9JHS3	mitogen-activated protein binding protein interacting protein	2	2	1041	28956	RP11-336K24.1	MAPIP_HUMAN	Q9Y2Q5	c	associated to late endosome membrane	125	0
26413	Mapk1	MK01_MOUSE	P63085	mitogen-activated protein kinase 1	8	16	278	5594	MAPK1	MK01_HUMAN	P28482	c	cytoplasm	358	0
26414	Mapk10	MK10_MOUSE	Q61831	mitogen-activated protein kinase 10	2	2	1193	5602	MAPK10	MK10_HUMAN	P53779	c	cytoplasm	464	0
26415	Mapk13	MK13_MOUSE	Q9Z1B7	mitogen-activated protein kinase 13	12	11	394	5603	MAPK13	MK13_HUMAN	O15264	c	cytoplasm	404	0
26416	Mapk14	MK14_MOUSE	P47811	mitogen-activated protein kinase 14	4	4	694	1432	MAPK14	MK14_HUMAN	Q16539	c	cytoplasm	360	0
26417	Mapk3	MK03_MOUSE	Q63844	mitogen-activated protein kinase 3	17	31	161	5595	MAPK3	MK03_HUMAN	P27361	c	cytosol	380	0
26419	Mapk8	MK08_MOUSE	Q91Y86	mitogen-activated protein kinase 8	2	2	899	5599	MAPK8	MK08_HUMAN	P45983	c	cytoplasm	384	0
227743	Mapkap1	SIN1_MOUSE	Q8BKH7	mitogen-activated protein kinase associated protein 1	2	2	1096	79109	MAPKAP1	SIN1_HUMAN	Q9BPZ7	c	associated to plasma membrane	522	0
108645	Mat2b	MAT2B_MOUSE	Q99LB6	methionine adenosyltransferase II, beta	3	4	738	27430	MAT2B	MAT2B_HUMAN	Q9NZL9	c	nucleus	334	0
17449	Mdh1	MDHC_MOUSE	P14152	malate dehydrogenase 1, NAD (soluble)	8	17	265	4190	MDH1	MDHC_HUMAN	P40925	c	cytoplasm	334	0
17448	Mdh2	MDHM_MOUSE	P08249	malate dehydrogenase 2, NAD (mitochondrial)	13	11	392	4191	MDH2	MDHM_HUMAN	P40926	c	mitochondrial inner membrane	338	0
67881	mdp1	MGDP1_MOUSE	Q9D967	Magnesium-dependent phosphatase 1	2	2	1028	145553	MDP-1	Q86V88	c	-	212	0
108098	Med21	MED21_MOUSE	Q9CQ39	mediator complex subunit 21	2	3	866	9412	MED21	MED21_HUMAN	Q13503	c	nucleus	144	0
29808	Mga	A2AWL7_MOUSE	A2AWL7	MAX gene associated	2	2	1096	23269	MGA	fragment	-	c	nucleus	3003	0
171580	Mical1	MICA1_MOUSE	Q8VDP3	microtubule associated monoxygenase, calponin and LIM domain containing 1	6	10	432	64780	MICAL1	MICA1_HUMAN	Q8TDZ2	c	cytoskeleton	1048	0

Gene ID (mouse)	Gene Symbol (mouse)	Swissprot protein ID (mouse)	Swissprot protein AC (mouse)	Gene Description	Max Diff peptide	Avg Pep count	Rank order	Ortho Gene ID	Ortho Gene Symbol	Swissprot protein ID	Swissprot protein AC	Topology	GO term	Protein length	Number of TM
50932	Mink1	MINK1_MOUSE	Q9JM52	misshapen-like kinase 1 (zebrafish)	2	2	1097	50488	MINK1	MINK1_HUMAN	Q8N4C8	c	-	332	0
68473	Mobkl1a	MOL1A_MOUSE	Q8BPB0	MOB1, Mps One Binder kinase activator-like 1A (yeast)	2	3	909	92597	MOBKL1A	MOL1A_HUMAN	Q7L9L4	c	cytoplasm	216	0
232157	Mobkl1b	MOL1B_MOUSE	Q921Y0	MOB1, Mps One Binder kinase activator-like 1B (yeast)	2	3	913	55233	MOBKL1B	MOL1B_HUMAN	Q9H8S9	c	-	240	0
17524	Mpp1	EM55_MOUSE	P70290	membrane protein, palmitoylated	12	21	228	4354	MPP1	EM55_HUMAN	Q00013	c	anchored to plasma membrane	466	0
56217	Mpp5	MPP5_MOUSE	Q9JLB2	membrane protein, palmitoylated 5 (MAGUK p55 subfamily member 5)	10	15	239	64398	MPP5	MPP5_HUMAN	Q8N3R9	c	associated to plasma membrane	675	0
17698	Msn	MOES_MOUSE	P26041	moesin	2	2	995	4478	MSN	MOES_HUMAN	P26038	c	cytoskeleton	577	0
70415	Mst4	MST4_MOUSE	Q99JT2	Serine/threonine-protein kinase MST4	7	6	556	51765	MST4	MST4_HUMAN	Q9P289	c	Golgi apparatus	416	0
108156	Mthfd1	C1TC_MOUSE	Q922D8	methylenetetrahydrofolate dehydrogenase (NADP+ dependent), methenyltetrahydrofolate cyclohydrolase, formyltetrahydrofolate synthase	3	3	963	4522	MTHFD1	C1TC_HUMAN	P11586	c	cytoplasm	935	0
17772	Mtm1	MTM1_MOUSE	Q9Z2C5	X-linked myotubular myopathy gene 1	7	5	665	4534	MTM1	MTM1_HUMAN	Q13496	c	cellular_component	603	0
17831	Muc2	Q8CIA2_MOUSE	Q8CIA2	mucin 2	7	7	525	4583	MUC2	MUC2_HUMAN	Q02817	c	extracellular region	1096	0
78388	Mvp	MVP_MOUSE	Q9EQK5	major vault protein	12	17	264	9961	MVP	MVP_HUMAN	Q14764	c	cytoplasm	861	0
17886	Myh9	MYH9_MOUSE	Q8VDD5	myosin, heavy polypeptide 9, non-muscle	12	12	341	4627	MYH9	MYH9_HUMAN	P35579	c	actin cytoskeleton	1960	0
17904	Myl6	MYL6_MOUSE	Q60605	myosin, light polypeptide 6, alkali, smooth muscle and non-muscle	10	50	100	4637	MYL6	MYL6_HUMAN	P60660	c	myosin complex	151	0
67938	Mylc2b	MLRB_MOUSE	Q3THE2	myosin light chain, regulatory B	3	4	735	103910	MRLC2	Q13182_HUMAN	Q13182	c	myosin complex	172	0
107589	Mylk	MYLK_MOUSE	Q6PDN3	myosin, light polypeptide kinase	6	10	430	4638	MYLK	MYLK_HUMAN	Q15746	c	cytoplasm	1941	0
217328	Myo15b	-	-	myosin XVB	10	42	122	80022	MYO15B	MY15B_HUMAN	Q96JP2	c	cytoskeleton	1530	0
432516	Myo1a	MYO1A_MOUSE	O88329	myosin IA	75	780	7	4640	MYO1A	MYO1A_HUMAN	Q9UBC5	c	apical plasma membrane	1043	0
17913	Myo1c	MYO1C_MOUSE	Q9WTI7	myosin IC	7	8	468	4641	MYO1C	MYO1C_HUMAN	O00159	c	basal plasma membrane	1028	0
338367	Myo1d	MYO1D_MOUSE	Q5SYD0	myosin ID	54	172	28	4642	MYO1D	MYO1D_HUMAN	O94832	c	myosin complex	1006	0
71602	Myo1e	Q80X36_MOUSE	Q80X36	myosin IE	2	2	1106	4643	MYO1E	MYO1E_HUMAN	Q12965	c	actin cytoskeleton	1107	0
17919	Myo5b	MYO5B_MOUSE	P21271	myosin Vb	3	6	568	4645	MYO5B	MYO5B_HUMAN	Q9ULV0	c	myosin complex	1818	0
17920	Myo6	MYO6_MOUSE	Q64331	myosin VI	29	67	73	4646	MYO6	MYO6_HUMAN	Q9UM54	c	clathrin-coated endocytic vesicle	1265	0
17921	Myo7a	MYO7A_MOUSE	P97479	myosin VIIa	36	59	82	4647	MYO7A	MYO7A_HUMAN	Q13402	c	cytoskeleton	2215	0
234664	Nae1	ULA1_MOUSE	Q8VBW6	NEDD8 activating enzyme E1 subunit 1	2	2	1166	8883	NAE1	ULA1_HUMAN	Q13564	c	associated to plasma membrane	534	0
59027	Nampt	NAMPT_MOUSE	Q99KQ4	nicotinamide phosphoribosyltransferase	8	10	427	10135	NAMPT	NAMPT_HUMAN	P43490	c	cytoplasm	491	0
108124	Napa	SNAA_MOUSE	Q9DB05	N-ethylmaleimide sensitive fusion protein attachment protein alpha	11	14	304	8775	NAPA	SNAA_HUMAN	P54920	c	associated to membrane	295	0
108123	Napg	SNAG_MOUSE	Q9CWZ7	N-ethylmaleimide sensitive fusion protein attachment protein gamma	3	2	962	8774	NAPG	SNAG_HUMAN	Q99747	c	associated to endoplasmic reticulum membrane	312	0
223646	Naprt1	PNCB_MOUSE	Q8CC86	nicotinate phosphoribosyltransferase domain containing 1	6	6	601	93100	NAPRT1	PNCB_HUMAN	Q6XQN6	c	cytoplasm	538	0
70223	Nars	SYNC_MOUSE	Q8BP47	asparaginyl-tRNA synthetase	10	8	473	4677	NARS	SYNC_HUMAN	O43776	c	cytoplasm	547	0
17961	Nat2	ARY2_MOUSE	P50295	N-acetyltransferase 2 (arylamine N-acetyltransferase)	5	5	666	9	NAT1	ARY1_HUMAN	P18440	c	cytoplasm	290	0
68342	Ndufb10	NDUBA_MOUSE	Q9DCS9	NADH dehydrogenase (ubiquinone) 1 beta subcomplex, 10	5	2	960	4716	NDUFB10	NDUBA_HUMAN	O96000	c	associated to mitochondrial inner membrane	176	0
66495	Ndufb3	NDUB3_MOUSE	Q9CQZ6	NADH dehydrogenase (ubiquinone) 1 beta subcomplex 3	2	2	1101	4709	NDUFB3	NDUB3_HUMAN	O43676	c	mitochondrial inner membrane	104	0
66916	Ndufb7	NDUB7_MOUSE	Q9CR61	NADH dehydrogenase (ubiquinone) 1 beta subcomplex, 7	3	2	1026	4713	NDUFB7	NDUB7_HUMAN	P17568	c	associated to mitochondrial inner membrane	137	0
59126	Nek6	NEK6_MOUSE	Q9ES70	NIMA (never in mitosis gene a)-related expressed kinase 6	4	3	906	10783	NEK6	NEK6_HUMAN	Q9HC98	c	cytoplasm	313	0
27045	Nit1	NIT1_MOUSE	Q8VDK1	nitrilase 1	2	2	1147	4817	NIT1	NIT1_HUMAN	Q86X76	c	-	239	0
268973	Nlrc4	Q3UP24_MOUSE	Q3UP24	NLR family, CARD domain containing 4	3	3	874	58484	NLRC4	NLRC4_HUMAN	Q9NPP4	c	cytoplasm	1024	0
446099	Nlrp4e	NAL4E_MOUSE	Q66X19	NLR family, pyrin domain containing 4E	7	6	602	147945	NLRP4	NALP4_HUMAN	Q96MN2	c		193	0
18102	Nme1	NDKA_MOUSE	P15532	non-metastatic cells 1, protein (NM23A) expressed in	6	6	569	4830	NME1	NDKA_HUMAN	P15531	c	cytoplasm	152	0
18103	Nme2	NDKB_MOUSE	Q01768	non-metastatic cells 2, protein (NM23B) expressed in	6	8	486	4831	NME2	NDKB_HUMAN	O60361	c	mitochondrion	152	0
18126	Nos2	NOS2_MOUSE	P29477	nitric oxide synthase 2, inducible, macrophage	16	24	210	4843	NOS2A	NOS2A_HUMAN	P35228	c	cytoplasm	1144	0
19155	Npepps	PSA_MOUSE	Q11011	aminopeptidase puromycin sensitive	5	5	668	9520	NPEPPS	PSA_HUMAN	P55786	c	cytoplasm	920	0
18176	Nras	RASN_MOUSE	P08556	neuroblastoma ras oncogene	5	3	845	4893	NRAS	RASN_HUMAN	P01111	c	anchored to plasma membrane	189	0
18195	Nsf	NSF_MOUSE	P46460	N-ethylmaleimide sensitive fusion protein	9	6	554	4905	NSF	NSF_HUMAN	P46459	c	cytoplasm	744	0
50773	Nt5c	NT5C_MOUSE	Q9JM14	5',3'-nucleotidase, cytosolic	2	2	1148	30833	NT5C	NT5C_HUMAN	Q8TCD5	c	cytoplasm	200	0
26426	Nubp2	Q9R061_MOUSE	Q9R061	nucleotide binding protein 2	2	2	1144	10101	NUBP2	NUBP2_HUMAN	Q9Y5Y2	c	nucleus	275	0
209586	Nudcd3	NUDC3_MOUSE	Q8R1N4	NudC domain containing 3	2	1	1193	23386	NUDCD3	NUDC3_HUMAN	Q8IVD9	c	-	471	0
53893	Nudt5	NUDT5_MOUSE	Q9JKX6	nudix (nucleoside diphosphate linked moiety X)-type motif 5	4	3	902	11164	NUDT5	NUDT5_HUMAN	Q9UKK9	c	intracellular	218	0
18222	Numb	NUMB_MOUSE	Q9QZS3	numb gene homolog (Drosophila)	4	6	555	8650	NUMB	NUMB_HUMAN	P49757	c	integral to plasma membrane	653	0
246730	Oas1a	OAS1A_MOUSE	P11928	2'-5' oligoadenylate synthetase 1A	8	9	437	4938	OAS1	OAS1_HUMAN	P00973	c	mitochondrion	367	0
23960	Oas1g	Q8K469_MOUSE	Q8K469	2'-5' oligoadenylate synthetase 1G	4	6	572	4938	OAS1	OAS1_HUMAN	P00973	c	mitochondrion	367	0
18293	Ogdh	ODO1_MOUSE	Q60597	oxoglutarate dehydrogenase (lipoamide)	2	2	1136	4967	OGDH	ODO1_HUMAN	Q02218	c	mitochondrial membrane	1023	0
20409	Ostf1	OSTF1_MOUSE	Q62422	osteoclast stimulating factor 1	9	9	453	26578	OSTF1	OSTF1_HUMAN	Q92882	c	cytoplasm	215	0
107260	Otub1	OTUB1_MOUSE	Q7TQI3	OTU domain, ubiquitin aldehyde binding 1	5	6	580	55611	OTUB1	OTUB1_HUMAN	Q96FW1	c	-	271	0
108737	Oxsr1	OXSR1_MOUSE	Q6P9R2	oxidative-stress responsive 1	4	4	739	9943	OXSR1	OXSR1_HUMAN	O95747	c	-	639	0
23970	Pacsin2	PACN2_MOUSE	Q9WVE8	protein kinase C and casein kinase substrate in neurons 2	5	7	527	11252	PACSIN2	PACN2_HUMAN	Q9UNF0	c	cytoplasm	486	0
18472	Pafah1b1	LIS1_MOUSE	P63005	platelet-activating factor acetylhydrolase, isoform 1b, beta1 subunit	14	14	291	5048	PAFAH1B1	LIS1_HUMAN	P43034	c	cytoskeleton	410	0
18476	Pafah1b3	PA1B3_MOUSE	Q61205	platelet-activating factor acetylhydrolase, isoform 1b, alpha1 subunit	3	3	810	5050	PAFAH1B3	PA1B3_HUMAN	Q15102	c	cytoplasm	232	0
100163	Pafah2	PAFA2_MOUSE	Q8VDG7	platelet-activating factor acetylhydrolase 2	12	6	429	5051	PAFAH2	PAFA2_HUMAN	Q99487	c	-	390	0
23971	Papss1	PAPS1_MOUSE	Q60967	3'-phosphoadenosine 5'-phosphosulfate synthase 1	9	10	424	9061	PAPSS1	PAPS1_HUMAN	O43252	c	-	147	0
23972	Papss2	PAPS2_MOUSE	O88428	3'-phosphoadenosine 5'-phosphosulfate synthase 2	33	140	30	9060	PAPSS2	PAPS2_HUMAN	O95340	c	-	366	0
58220	Pard6b	PAR6B_MOUSE	Q9JK83	par-6 (partitioning defective 6) homolog beta (C. elegans)	2	2	1100	84612	PARD6B	PAR6B_HUMAN	Q9BYG5	c	associated to plasma membrane	371	0
57320	Park7	PARK7_MOUSE	Q99LX0	Parkinson disease (autosomal recessive, early onset) 7	3	2	1021	11315	PARK7	PARK7_HUMAN	Q99497	c	cytoplasm	189	0
67307	Pbld2	PBLD2_MOUSE	Q9CXN7	Phenazine biosynthesis-like domain containing protein 2	4	4	770	64081	PBLD	PBLD_HUMAN	P30039	c	cellular_component	288	0
23983	Pcbp1	PCBP1_MOUSE	P60335	poly(rC) binding protein 1	6	13	331	5093	PCBP1	PCBP1_HUMAN	Q15365	c	cytoplasm	356	0
18538	Pcna	PCNA_MOUSE	P17918	proliferating cell nuclear antigen	2	2	1138	5111	PCNA	PCNA_HUMAN	P12004	c	nucleus	261	0
18557	Pctk3	PCTK3_MOUSE	Q04899	PCTAIRE-motif protein kinase 3	4	5	667	5129	PCTK3	PCTK3_HUMAN	Q07002	c	cytoplasm	451	0
68671	Pcyt2	PCY2_MOUSE	Q922E4	phosphate cytidylyltransferase 2, ethanolamine	7	5	621	5833	PCYT2	PCY2_HUMAN	Q99447	c	cellular_component	404	0
56426	Pdcd10	PDC10_MOUSE	Q9JW70	programmed cell death 10	9	18	260	11235	PDCD10	PDC10_HUMAN	Q9BUL8	c	cytoplasm	339	0
18570	Pdcd6	PDCD6_MOUSE	P12815	programmed cell death 6	5	5	615	10016	PDCD6	PDCD6_HUMAN	O75340	c	associated to endoplasmic reticulum membrane	191	0
18571	Pdcd6ip	PDC6I_MOUSE	Q9WU78	programmed cell death 6 interacting protein	21	16	280	10015	PDCD6IP	PDC6I_HUMAN	Q8WUM4	c	cytoplasm	869	0
18585	Pde9a	PDE9A_MOUSE	O70628	phosphodiesterase 9A	7	5	632	5152	PDE9A	PDE9A_HUMAN	O76083	c	cytoplasm	534	0
68263	Pdhb	ODPB_MOUSE	Q9D051	pyruvate dehydrogenase (lipoamide) beta	2	2	1103	5162	PDHB	ODPB_HUMAN	P11177	c	mitochondrion	359	0
54132	Pdlim1	PDLI1_MOUSE	O70400	PDZ and LIM domain 1 (elfin)	5	6	575	9124	PDLIM1	PDLI1_HUMAN	O00151	c	cytoskeleton	329	0
56376	Pdlim5	PDLI5_MOUSE	Q8CI51	PDZ and LIM domain 5	4	6	576	10611	PDLIM5	PDLI5_HUMAN	Q96HC4	c	actin cytoskeleton	591	0
216134	Pdxk	PDXK_MOUSE	Q8K183	pyridoxal (pyridoxine, vitamin B6) kinase	4	3	824	8566	PDXK	CU124_HUMAN	Q9H1N0	c	cytoplasm	312	0
170761	Pdzd3	PDZD3_MOUSE	Q99MJ6	PDZ domain containing 3	11	17	267	79849	PDZD3	PDZD3_HUMAN	Q86UT5	c	associated to membrane apical part of cell	498	0
59020	Pdzk1	PDZD1_MOUSE	Q9JIL4	PDZ domain containing 1 (Na/Pi cotransporter C-terminal-associated protein)	38	179	24	5174	PDZK1	PDZD1_HUMAN	Q5T2W1	c	associated to membrane	519	0
67898	Pef1	PEF1_MOUSE	Q8BFY6	penta-EF hand domain containing 1	5	5	639	553115	PEF1	PEF1_HUMAN	Q9UBV8	c	associated to membrane	275	0
18624	Pepd	PEPD_MOUSE	Q11136	peptidase D	13	12	353	5184	PEPD	PEPD_HUMAN	P12955	c	-	193	0
18641	Pfkl	K6PL_MOUSE	P12382	phosphofructokinase, liver, B-type	6	7	549	5211	PFKL	K6PL_HUMAN	P17858	c	cytoplasm	780	0
56421	Pfkp	K6PP_MOUSE	Q9WUA3	phosphofructokinase, platelet	18	24	212	5214	PFKP	K6PP_HUMAN	Q01813	c	cytoplasm	784	0
18643	Pfn1	PROF1_MOUSE	P62962	profilin 1	8	23	217	5216	PFN1	PROF1_HUMAN	P07737	c	actin cytoskeleton	140	0
18648	Pgam1	PGAM1_MOUSE	Q9DBJ1	phosphoglycerate mutase 1	11	15	286	5223	PGAM1	PGAM1_HUMAN	P18669	c	cytosol	254	0
110208	Pgd	6PGD_MOUSE	Q9DCD0	phosphogluconate dehydrogenase	17	29	177	5226	PGD	6PGD_HUMAN	P52209	c	-	1130	0
18655	Pgk1	PGK1_MOUSE	P09411	phosphoglycerate kinase 1	17	26	193	5230	PGK1	PGK1_HUMAN	P00558	c	cytoplasm	417	0
66681	Pgm2	PGM2_MOUSE	Q7TSV4	phosphoglucomutase 2	12	4	767	55276	PGM2	PGM2_HUMAN	Q96G03	c	cytoplasm	620	0
67078	Pgp	PGP_MOUSE	Q8CHP8	Phosphoglycolate phosphatase	2	2	1153	283871	PGP	PGP_HUMAN	A6NDG6	c	-	323	0
237928	Phospho1	PHOP1_MOUSE	Q8R2H9	phosphatase, orphan 1	3	3	873	162466	PHOSPHO1	PHOP1_HUMAN	Q8TCT1	c	-	267	0
18738	Pitpna	PIPNA_MOUSE	P53810	phosphatidylinositol transfer protein, alpha	3	3	811	5306	PITPNA	PIPNA_HUMAN	Q00169	c	cytoplasm	271	0
56305	Pitpnb	PIPNB_MOUSE	P53811	phosphatidylinositol transfer protein, beta	5	5	729	23760	PITPNB	PIPNB_HUMAN	P48739	c	cytoplasm	271	0
18770	Pklr	KPYR_MOUSE	P53657	pyruvate kinase liver and red blood cell	11	18	246	5313	PKLR	KPYR_HUMAN	P30613	c	-	1337	0
18746	Pkm2	KPYM_MOUSE	P52480	pyruvate kinase, muscle	26	95	57	5315	PKM2	KPYM_HUMAN	P14618	c	cytoplasm	531	0
18797	Plcb3	PLCB3_MOUSE	P51432	phospholipase C, beta 3	10	9	452	5331	PLCB3	PLCB3_HUMAN	Q01970	c	associated to membrane	1234	0
18799	Plcd1	PLCD1_MOUSE	Q8R3B1	phospholipase C, delta 1	6	4	692	5333	PLCD1	PLCD1_HUMAN	P51178	c	cytoplasm	756	0
18805	Pld1	PLD1_MOUSE	Q9Z280	phospholipase D1	15	45	114	5337	PLD1	PLD1_HUMAN	Q13393	c	anchored to endoplasmic reticulum membrane	1074	0
71801	Plekhf2	PKHF2_MOUSE	Q91WB4	pleckstrin homology domain containing, family F (with FYVE domain) member 2	2	3	860	79666	PLEKHF2	PKHF2_HUMAN	Q9H8W4	c	transport vesicle	249	0
54128	Pmm2	PMM2_MOUSE	Q9Z2M7	phosphomannomutase 2	5	5	636	5373	PMM2	PMM2_HUMAN	O15305	c	cytoplasm	242	0
18950	Pnp1	PNPH_MOUSE	P23492	purine-nucleoside phosphorylase 1	8	12	370	4860	NP	PNPH_HUMAN	P00491	c	cytosol	289	0
67895	Ppa1	IPYR_MOUSE	Q9D819	pyrophosphatase (inorganic) 1	5	8	464	5464	PPA1	IPYR_HUMAN	Q15181	c	cytoplasm	289	0
268373	Ppia	PPIA_MOUSE	P17742	peptidylprolyl isomerase A	13	28	174	5478	PPIA	PPIA_HUMAN	P62937	c	cytoplasm	164	0
19045	Ppp1ca	PP1A_MOUSE	P62137	protein phosphatase 1, catalytic subunit, alpha isoform	12	10	421	5499	PPP1CA	PP1A_HUMAN	P62136	c	cytoplasm	330	0
19046	Ppp1cb	PP1B_MOUSE	P62141	protein phosphatase 1, catalytic subunit, beta isoform	9	8	506	5500	PPP1CB	PP1B_HUMAN	P62140	c	cytoplasm	327	0
19052	Ppp2ca	PP2AA_MOUSE	P63330	protein phosphatase 2 (formerly 2A), catalytic subunit, alpha isoform	5	5	686	5515	PPP2CA	PP2AA_HUMAN	P67775	c	cytoplasm	309	0

Gene ID (mouse)	Gene Symbol (mouse)	Swissprot protein ID (mouse)	Swissprot protein AC (mouse)	Gene Description	Max Diff peptide	Avg Pep count	Rank order	Ortho Gene ID	Ortho Gene Symbol	Swissprot protein ID	Swissprot protein AC	Topology	GO term	Protein length	Number of TM
51792	Ppp2r1a	2AAA_MOUSE	Q76MZ3	protein phosphatase 2 (formerly 2A), regulatory subunit A (PR 65), alpha isoform	12	14	301	5518	PPP2R1A	2AAA_HUMAN	P30153	c	cytosol	589	0
21770	Ppp2r5d	Q7TNL5_MOUSE	Q7TNL5	protein phosphatase 2, regulatory subunit B (B56), delta isoform	3	6	596	5528	PPP2R5D	2A5D_HUMAN	Q14738	c	cytoplasm	595	0
18477	Prdx1	PRDX1_MOUSE	P35700	peroxiredoxin 1	11	24	205	5052	PRDX1	PRDX1_HUMAN	Q06830	c	cytoplasm	199	0
21672	Prdx2	PRDX2_MOUSE	Q61171	peroxiredoxin 2	6	8	487	7001	PRDX2	PRDX2_HUMAN	P32119	c	cytoplasm	198	0
54683	Prdx5	PRDX5_MOUSE	P99029	peroxiredoxin 5	9	10	404	25824	PRDX5	PRDX5_HUMAN	P30044	c	mitochondrion	210	0
11758	Prdx6	PRDX6_MOUSE	O08709	peroxiredoxin 6	14	21	231	9588	PRDX6	PRDX6_HUMAN	P30041	c	cytoplasm	224	0
105787	Prkaa1	AAPK1_MOUSE	Q5EG47	protein kinase, AMP-activated, alpha 1 catalytic subunit	5	6	599	5562	PRKAA1	AAPK1_HUMAN	Q13131	c	intracellular	548	0
108079	Prkaa2	AAPK2_MOUSE	Q8BRK8	protein kinase, AMP-activated, alpha 2 catalytic subunit	5	3	865	5563	PRKAA2	AAPK2_HUMAN	P54646	c	cytosol	552	0
19079	Prkab1	AAKB1_MOUSE	Q9R078	protein kinase, AMP-activated, beta 1 non-catalytic subunit	2	2	1092	5564	PRKAB1	AAKB1_HUMAN	Q9Y478	c	nucleus	270	0
18747	Prkaca	KAPCA_MOUSE	P05132	protein kinase, cAMP dependent, catalytic, alpha	11	11	376	5566	PRKACA	KAPCA_HUMAN	P17612	c	cytoplasm	351	0
19082	Prkag1	AAKG1_MOUSE	O54950	protein kinase, AMP-activated, gamma 1 non-catalytic subunit	9	9	441	5571	PRKAG1	AAKG1_HUMAN	P54619	c	cytoplasm	330	0
19087	Prkar2a	KAP2_MOUSE	P12367	protein kinase, cAMP dependent regulatory, type II alpha	19	57	86	5576	PRKAR2A	KAP2_HUMAN	P13861	c	plasma membrane	401	0
18750	Prkca	KPCA_MOUSE	P20444	protein kinase C, alpha	17	20	238	5578	PRKCA	KPCA_HUMAN	P17252	c	plasma membrane	672	0
18753	Prkcd	KPCD_MOUSE	P28867	protein kinase C, delta	20	36	141	5580	PRKCD	KPCD_HUMAN	Q05655	c	cytoplasm	674	0
18761	Prkcq	KPCT_MOUSE	Q02111	protein kinase C, theta	3	4	721	5588	PRKCQ	KPCT_HUMAN	Q04759	c	intracellular	707	0
18762	Prkcz	KPCZ_MOUSE	Q02956	protein kinase C, zeta	2	2	1139	5590	PRKCZ	KPCZ_HUMAN	Q05513	c	cytoplasm	592	0
19092	Prkg2	KGP2_MOUSE	Q61410	protein kinase, cGMP-dependent, type II	33	65	75	5593	PRKG2	KGP2_HUMAN	Q13237	c	-	340	0
114863	Prosc	PROSC_MOUSE	Q9Z2Y8	proline synthetase co-transcribed	2	2	1044	11212	PROSC	PROSC_HUMAN	O94903	c	cytoplasm	274	0
26440	Psma1	PSA1_MOUSE	Q9R1P4	proteasome (prosome, macropain) subunit, alpha type 1	6	6	611	5682	PSMA1	PSA1_HUMAN	P25786	c	proteasome complex	263	0
19166	Psma2	PSA2_MOUSE	P49722	proteasome (prosome, macropain) subunit, alpha type 2	3	3	932	5683	PSMA2	PSA2_HUMAN	P25787	c	proteasome complex	234	0
19167	Psma3	PSA3_MOUSE	O70435	proteasome (prosome, macropain) subunit, alpha type 3	3	3	846	5684	PSMA3	PSA3_HUMAN	P25788	c	proteasome complex	255	0
26441	Psma4	PSA4_MOUSE	Q9R1P0	proteasome (prosome, macropain) subunit, alpha type 4	2	2	1009	5685	PSMA4	PSA4_HUMAN	P25789	c	proteasome complex	261	0
26442	Psma5	PSA5_MOUSE	Q9Z2U1	proteasome (prosome, macropain) subunit, alpha type 5	5	5	617	5686	PSMA5	PSA5_HUMAN	P28066	c	proteasome complex	241	0
26443	Psma6	PSA6_MOUSE	Q9QUM9	proteasome (prosome, macropain) subunit, alpha type 6	4	4	724	5687	PSMA6	PSA6_HUMAN	P60900	c	proteasome core complex	246	0
26444	Psma7	PSA7_MOUSE	Q9Z2U0	proteasome (prosome, macropain) subunit, alpha type 7	5	4	695	5688	PSMA7	PSA7_HUMAN	O14818	c	proteasome complex	248	0
73677	Psma8	PSA7L_MOUSE	Q9CWH6	proteasome (prosome, macropain) subunit, alpha type, 8	2	2	1159	143471	PSMA8	PSA7L_HUMAN	Q8TAA3	c	proteasome core complex	250	0
19170	Psmb1	PSB1_MOUSE	O09061	proteasome (prosome, macropain) subunit, beta type 1	4	4	759	5689	PSMB1	PSB1_HUMAN	P20618	c	proteasome complex	240	0
19171	Psmb10	PSB10_MOUSE	O35955	proteasome (prosome, macropain) subunit, beta type 10	4	4	722	5699	PSMB10	PSB10_HUMAN	P40306	c	proteasome complex	273	0
26445	Psmb2	PSB2_MOUSE	Q9R1P3	proteasome (prosome, macropain) subunit, beta type 2	2	1	1190	5690	PSMB2	PSB2_HUMAN	P49721	c	proteasome complex	201	0
26446	Psmb3	PSB3_MOUSE	Q9R1P1	proteasome (prosome, macropain) subunit, beta type 3	2	2	1010	5691	PSMB3	PSB3_HUMAN	P49720	c	proteasome complex	205	0
19175	Psmb6	PSB6_MOUSE	Q60692	proteasome (prosome, macropain) subunit, beta type 6	3	2	954	5694	PSMB6	PSB6_HUMAN	P28072	c	proteasome core complex	238	0
19179	Psmc1	PRS4_MOUSE	P62192	protease (prosome, macropain) 26S subunit, ATPase1	13	16	274	5700	PSMC1	PRS4_HUMAN	P62191	c	proteasome complex	440	0
19181	Psmc2	PRS7_MOUSE	P46471	proteasome (prosome, macropain) 26S subunit, ATPase 2	16	24	206	5701	PSMC2	PRS7_HUMAN	P35998	c	proteasome complex	433	0
19182	Psmc3	PRS6A_MOUSE	O88685	proteasome (prosome, macropain) 26S subunit, ATPase 3	10	14	298	5702	PSMC3	PRS6A_HUMAN	P17980	c	proteasome complex	442	0
23996	Psmc4	PRS6B_MOUSE	P54775	proteasome (prosome, macropain) 26S subunit, ATPase, 4	8	9	454	5704	PSMC4	PRS6B_HUMAN	P43686	c	proteasome complex	418	0
19184	Psmc5	PRS8_MOUSE	P62196	protease (prosome, macropain) 26S subunit, ATPase 5	13	14	299	5705	PSMC5	PRS8_HUMAN	P62195	c	proteasome complex	406	0
67089	Psmc6	PRS10_MOUSE	P62334	proteasome (prosome, macropain) 26S subunit, ATPase, 6	12	10	428	5706	PSMC6	PRS10_HUMAN	P62333	c	proteasome complex	389	0
69077	Psmd11	PSD11_MOUSE	Q8BG32	proteasome (prosome, macropain) 26S subunit, non-ATPase, 11	12	12	344	5717	PSMD11	PSD11_HUMAN	O00231	c	proteasome complex	422	0
66997	Psmd12	PSD12_MOUSE	Q9D8W5	proteasome (prosome, macropain) 26S subunit, non-ATPase, 12	10	10	407	5718	PSMD12	PSD12_HUMAN	O00232	c	proteasome regulatory particle	456	0
23997	Psmd13	PSD13_MOUSE	Q9WVJ2	proteasome (prosome, macropain) 26S subunit, non-ATPase, 13	8	10	411	5719	PSMD13	PSD13_HUMAN	Q9UNM6	c	proteasome complex	376	0
59029	Psmd14	PSDE_MOUSE	O35593	proteasome (prosome, macropain) 26S subunit, non-ATPase, 14	3	2	956	10213	PSMD14	PSDE_HUMAN	O00487	c	proteasome complex	310	0
22123	Psmd3	PSMD3_MOUSE	P14685	proteasome (prosome, macropain) 26S subunit, non-ATPase, 3	8	7	535	5709	PSMD3	PSMD3_HUMAN	O43242	c	proteasome complex	530	0
19185	Psmd4	PSMD4_MOUSE	O35226	proteasome (prosome, macropain) 26S subunit, non-ATPase, 4	2	2	1140	5710	PSMD4	PSMD4_HUMAN	P55036	c	proteasome complex	376	0
66413	Psmd6	PSMD6_MOUSE	Q99JI4	proteasome (prosome, macropain) 26S subunit, non-ATPase, 6	5	3	907	9861	PSMD6	PSMD6_HUMAN	Q15008	c	proteasome complex	389	0
57296	Psmd8	PSMD8_MOUSE	Q9CX56	proteasome (prosome, macropain) 26S subunit, non-ATPase, 8	2	3	855	5714	PSMD8	PSMD8_HUMAN	P48556	c	proteasome complex	257	0
19186	Psme1	PSME1_MOUSE	P97371	proteasome (prosome, macropain) 28 subunit, alpha	8	7	526	5720	PSME1	PSME1_HUMAN	Q06323	c	proteasome complex	249	0
19188	Psme2	PSME2_MOUSE	P97372	proteasome (prosome, macropain) 28 subunit, beta	7	11	377	5721	PSME2	PSME2_HUMAN	Q9UL46	c	proteasome complex	239	0
19205	Ptbp1	PTBP1_MOUSE	P17225	polypyrimidine tract binding protein 1	2	2	996	5725	PTBP1	PTBP1_HUMAN	P26599	c	nucleus	527	0
20459	Ptk6	PTK6_MOUSE	Q64434	PTK6 protein tyrosine kinase 6	3	2	933	5753	PTK6	PTK6_HUMAN	Q13882	c	cytoplasm	451	0
15170	Ptpn6	PTN6_MOUSE	P29351	protein tyrosine phosphatase, non-receptor type 6	3	2	990	5777	PTPN6	PTN6_HUMAN	P29350	c	cytoplasm	595	0
19303	Pxn	PAXI_MOUSE	Q8VI36	paxillin	3	2	1093	5829	PXN	PAXI_HUMAN	P49023	c	cytoskeleton	591	0
66824	Pycard	ASC_MOUSE	Q9EPB4	PYD and CARD domain containing	3	5	672	29108	PYCARD	ASC_HUMAN	Q9ULZ3	c	cytoplasm	193	0
66194	Pycrl	P5CR3_MOUSE	Q9DCC4	pyrroline-5-carboxylate reductase-like	3	3	856	65263	PYCRL	P5CR3_HUMAN	Q53H96	c	-	204	0
97541	Qars	Q8R1V9_MOUSE	Q8R1V9	glutaminyl-tRNA synthetase	2	2	1160	5859	QARS	SYQ_HUMAN	P47897	c	cytoplasm	606	0
19324	Rab1	RAB1A_MOUSE	P62821	RAB1, member RAS oncogene family	14	89	60	5861	RAB1A	RAB1A_HUMAN	P62820	c	anchored to Golgi membrane	205	0
53869	Rab11a	RB11A_MOUSE	P62492	RAB11a, member RAS oncogene family	11	14	302	8766	RAB11A	RB11A_HUMAN	P62491	c	anchored to plasma membrane	216	0
19326	Rab11b	RB11B_MOUSE	P46638	RAB11B, member RAS oncogene family	12	60	81	9230	RAB11B	RB11B_HUMAN	Q15907	c	anchored to plasma membrane	218	0
68365	Rab14	RAB14_MOUSE	Q91V41	RAB14, member RAS oncogene family	12	28	178	51552	RAB14	RAB14_HUMAN	P61106	c	anchored to plasma membrane	215	0
104886	Rab15	RAB15_MOUSE	Q8K386	RAB15, member RAS oncogene family	2	3	821	376267	RAB15	RAB15_HUMAN	P59190	c	anchored to membrane	212	0
19331	Rab19	RAB19_MOUSE	P35294	RAB19, member RAS oncogene family	5	5	669	401409	RAB19	RAB19_HUMAN	A4D1S5	c	anchored to plasma membrane	217	0
76308	Rab1b	RAB1B_MOUSE	Q9D1G1	RAB1B, member RAS oncogene family	12	40	126	81876	RAB1B	RAB1B_HUMAN	Q9H0U4	c	anchored to plasma membrane	201	0
19334	Rab22a	RB22A_MOUSE	P35285	RAB22A, member RAS oncogene family	6	5	616	57403	RAB22A	RB22A_HUMAN	Q9UL26	c	anchored to endosome membrane	194	0
53868	Rab25	RAB25_MOUSE	Q9WTL2	RAB25, member RAS oncogene family	8	10	403	57111	RAB25	RAB25_HUMAN	P57735	c	anchored to plasma membrane	213	0
11891	Rab27a	RB27A_MOUSE	Q9ERI2	RAB27A, member RAS oncogene family	3	4	713	5873	RAB27A	RB27A_HUMAN	P51159	c	anchored to membrane	221	0
80718	Rab27b	RB27B_MOUSE	Q99P58	RAB27b, member RAS oncogene family	6	8	476	5874	RAB27B	RB27B_HUMAN	O00194	c	anchored to membrane	218	0
59021	Rab2a	RAB2A_MOUSE	P53994	RAB2A, member RAS oncogene family	12	37	137	5862	RAB2A	RAB2A_HUMAN	P61019	c	anchored to ER-Golgi intermediate compartme	212	0
75985	Rab30	RAB30_MOUSE	Q923S9	RAB30, member RAS oncogene family	5	6	557	27314	RAB30	RAB30_HUMAN	Q15771	c	anchored to plasma membrane	203	0
67844	Rab32	RAB32_MOUSE	Q9CZE3	RAB32, member RAS oncogene family	3	3	908	10981	RAB32	RAB32_HUMAN	Q13637	c	anchored to mitochondrial membrane	223	0
19338	Rab33b	RB33B_MOUSE	O35963	RAB33B, member of RAS oncogene family	6	15	288	83452	RAB33B	RB33B_HUMAN	Q9H082	c	anchored to Golgi membrane	229	0
77407	Rab35	RAB35_MOUSE	Q6PHN9	RAB35, member RAS oncogene family	10	29	167	11021	RAB35	RAB35_HUMAN	Q15286	c	anchored to plasma membrane	201	0
19339	Rab3a	RAB3A_MOUSE	P63011	RAB3A, member RAS oncogene family	3	3	896	5864	RAB3A	RAB3A_HUMAN	P20336	c	anchored to plasma membrane	220	0
19340	Rab3d	RAB3D_MOUSE	P35276	RAB3D, member RAS oncogene family	6	20	240	9545	RAB3D	RAB3D_HUMAN	O95716	c	anchored to plasma membrane	219	0
69834	Rab43	RAB43_MOUSE	Q8CG50	RAB43, member RAS oncogene family	5	6	578	339122	RAB43	RAB43_HUMAN	Q86YS6	c	anchored to plasma membrane	210	0
19341	Rab4a	RAB4A_MOUSE	P56371	RAB4A, member RAS oncogene family	10	14	311	5867	RAB4A	RAB4A_HUMAN	P20338	c	anchored to membrane	213	0
19342	Rab4b	RAB4B_MOUSE	Q91ZR1	RAB4B, member RAS oncogene family	2	3	847	53916	RAB4B	RAB4B_HUMAN	P61018	c	anchored to membrane	213	0
271457	Rab5a	RAB5A_MOUSE	Q9CQD1	RAB5A, member RAS oncogene family	8	26	191	5868	RAB5A	RAB5A_HUMAN	P20339	c	anchored to plasma membrane	215	0
19346	Rab6	RAB6A_MOUSE	P35279	RAB6, member RAS oncogene family	8	13	325	5870	RAB6A	RAB6A_HUMAN	P20340	c	anchored to Golgi membrane	208	0
270192	Rab6b	RAB6B_MOUSE	P61294	RAB6B, member RAS oncogene family	3	3	875	51560	RAB6B	RAB6B_HUMAN	Q9NRW1	c	anchored to Golgi membrane	208	0
19349	Rab7	RAB7A_MOUSE	P51150	RAB7, member RAS oncogene family	12	28	177	7879	RAB7A	RAB7A_HUMAN	P51149	c	anchored to late endosome membrane	207	0
17274	Rab8a	RAB8A_MOUSE	P55258	RAB8A, member RAS oncogene family	14	32	156	4218	RAB8A	RAB8A_HUMAN	P61006	c	anchored to plasma membrane	207	0
56382	Rab9	RAB9A_MOUSE	Q9R0M6	RAB9, member RAS oncogene family	6	5	637	9367	RAB9A	RAB9A_HUMAN	P51151	c	anchored to plasma membrane	201	0
56187	Rabggta	PGTA_MOUSE	Q9JHK4	Rab geranylgeranyl transferase, a subunit	5	4	794	5875	RABGGTA	PGTA_HUMAN	Q92696	c	-	195	0
19354	Rac2	RAC2_MOUSE	Q05144	RAS-related C3 botulinum substrate 2	2	4	760	5880	RAC2	RAC2_HUMAN	P15153	c	associated to plasma membrane	192	0
170758	Rac3	RAC3_MOUSE	P60764	RAS-related C3 botulinum substrate 3	2	4	776	5881	RAC3	RAC3_HUMAN	P60763	c	anchored to plasma membrane	192	0
56044	Rala	RALA_MOUSE	P63321	v-ral simian leukemia viral oncogene homolog A (ras related)	10	32	155	5898	RALA	RALA_HUMAN	P11233	c	anchored to plasma membrane	206	0
64143	Ralb	RALB_MOUSE	Q9JIW9	v-ral simian leukemia viral oncogene homolog B (ras related)	8	17	263	5899	RALB	RALB_HUMAN	P11234	c	anchored to plasma membrane	206	0

|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
19384	Ran	RAN_MOUSE	P62827	RAN, member RAS oncogene family	6	29	166	5901	RAN	RAN_HUMAN	P62826	c	cytoplasm	216	0
19385	Ranbp1	RANG_MOUSE	P34022	RAN binding protein 1	2	3	897	5902	RANBP1	RANG_HUMAN	P43487	c	cytoplasm	203	0
109905	Rap1a	RAP1A_MOUSE	P62835	RAS-related protein-1a	11	21	229	5906	RAP1A	RAP1A_HUMAN	P62834	c	anchored to plasma membrane	184	0
215449	Rap1b	RAP1B_MOUSE	Q99JI6	RAS related protein 1b	11	16	282	5908	RAP1B	RAP1B_HUMAN	P61224	c	anchored to plasma membrane	184	0
74012	Rap2b	RAP2B_MOUSE	P61226	RAP2B, member of RAS oncogene family	4	6	579	5912	RAP2B	RAP2B_HUMAN	P61225	c	anchored to plasma membrane	183	0
72065	Rap2c	RAP2C_MOUSE	Q8BU31	RAP2C, member of RAS oncogene family	10	12	345	57826	RAP2C	RAP2C_HUMAN	Q9Y3L5	c	anchored to plasma membrane	183	0
104458	Rars	SYRC_MOUSE	Q9D0I9	arginyl-tRNA synthetase	11	14	312	5917	RARS	SYRC_HUMAN	P54136	c	cytoplasm	660	0
70727	Rasgef1a	Q3TYC0_MOUSE	Q3TYC0	RasGEF domain family, member 1A	2	3	939	221002	RASGEF1A	RGF1A_HUMAN	Q8N9B8	c	intracellular	344	0
19428	RasI2-9	RANT_MOUSE	Q61820	RAS-like, family 2, locus 9	2	2	998	-	-	-	-	c	nucleus	216	0
19660	Rbp2	RET2_MOUSE	Q08552	retinol binding protein 2, cellular	7	16	277	5948	RBP2	RET2_HUMAN	P50120	c	cytoplasm	134	0
54150	Rdh7	RDH7_MOUSE	O88451	retinol dehydrogenase 7	13	11	386	8608	RDH16	RDH16_HUMAN	O75452	c	integral to microsome membrane	316	0
19684	Rdx	RADI_MOUSE	P26043	radixin	2	2	999	5962	RDX	RADI_HUMAN	P35241	c	cytoskeleton	583	0
66532	Rep15	REP15_MOUSE	Q9D7T1	Rab15 effector protein	4	2	957	387849	REP15	REP15_HUMAN	Q6BDI9	c	associated to early endosome membrane	230	0
54391	Rfk	RIFK_MOUSE	Q8CFV9	riboflavin kinase	2	2	1098	55312	RFK	RIFK_HUMAN	Q969G6	c	cytoplasm	155	0
19744	Rheb	RHEB_MOUSE	Q921J2	RAS-homolog enriched in brain	2	2	955	6009	RHEB	RHEB_HUMAN	Q15382	c	cytoplasm	184	0
11848	Rhoa	RHOA_MOUSE	Q9QUI0	ras homolog gene family, member A	8	22	222	387	RHOA	RHOA_HUMAN	P61586	c	anchored to plasma membrane	193	0
11852	Rhob	RHOB_MOUSE	P62746	ras homolog gene family, member B	5	9	449	388	RHOB	RHOB_HUMAN	P62745	c	anchored to plasma membrane	196	0
11853	Rhoc	RHOC_MOUSE	Q62159	ras homolog gene family, member C	8	21	232	389	RHOC	RHOC_HUMAN	P08134	c	anchored to plasma membrane	193	0
11854	Rhod	RHOD_MOUSE	P97348	ras homolog gene family, member D	6	5	612	29984	RHOD	RHOD_HUMAN	O00212	c	plasma membrane	210	0
23912	Rhof	RHOF_MOUSE	Q8BYP3	ras homolog gene family, member f	4	4	764	54509	RHOF	RHOF_HUMAN	Q9HBH0	c	cytoskeleton	211	0
56212	Rhog	RHOG_MOUSE	P84096	ras homolog gene family, member G	10	17	262	391	RHOG	RHOG_HUMAN	P84095	c	anchored to plasma membrane	191	0
52428	Rhpn2	RHPN2_MOUSE	Q8BWR8	rhophilin, Rho GTPase binding protein 2	7	7	536	85415	RHPN2	RHPN2_HUMAN	Q8IUC4	c	cytoplasm	686	0
56532	Ripk3	RIPK3_MOUSE	Q9QZL0	receptor-interacting serine-threonine kinase 3	2	2	1150	11035	RIPK3	RIPK3_HUMAN	Q9Y572	c	cytoplasm	486	0
107702	Rnh1	RINI_MOUSE	Q91VI7	ribonuclease/angiogenin inhibitor 1	9	10	431	6050	RNH1	RINI_HUMAN	P13489	c	cytoplasm	456	0
215615	Rnpep	AMBP_MOUSE	Q8VCT3	arginyl aminopeptidase (aminopeptidase B)	13	11	395	6051	RNPEP	AMBP_HUMAN	Q9H4A4	c	plasma membrane / secreted	650	0
19889	Rp2h	XRP2_MOUSE	Q9EPK2	retinitis pigmentosa 2 homolog (human)	13	12	223	6102	RP2	XRP2_HUMAN	O75695	c	plasma membrane	350	0
245670	Rragb	RRAGB_MOUSE	Q6NTA4	Ras-related GTP binding B	2	2	1048	10325	RRAGB	RRAGB_HUMAN	Q5VZM2	c	cytoplasm	374	0
54170	Rragc	RRAGC_MOUSE	Q99K70	Ras-related GTP binding C	3	2	1017	64121	RRAGC	RRAGC_HUMAN	Q9HB90	c	cytoplasm	398	0
52187	Rragd	RRAGD_MOUSE	Q7TT45	Ras-related GTP binding D	3	3	1013	58528	RRAGD	RRAGD_HUMAN	Q9NQL2	c	cytoplasm	449	0
66922	Rras2	RRAS2_MOUSE	P62071	related RAS viral (r-ras) oncogene homolog 2	8	14	315	22800	RRAS2	RRAS2_HUMAN	P62070	c	plasma membrane	204	0
56505	Ruvbl1	RUVB1_MOUSE	P60122	RuvB-like protein 1	2	2	1099	8607	RUVBL1	RUVB1_HUMAN	Q9Y265	c	nucleus	456	0
56045	Samhd1	SAMH1_MOUSE	Q60710	SAM domain and HD domain, 1	10	8	471	25939	SAMHD1	SAMH1_HUMAN	Q9Y3Z3	c	nucleus	627	0
66397	Sar1b	SAR1B_MOUSE	Q9CQC9	SAR1 gene homolog B (S. cerevisiae)	5	7	539	51128	SAR1B	SAR1B_HUMAN	Q9Y6B6	c	associated to endoplasmic reticulum membrane	198	0
20226	Sars	SYSC_MOUSE	P26638	seryl-aminoacyl-tRNA synthetase	3	2	1000	6301	SARS	SYSC_HUMAN	P49591	c	cytoplasm	512	0
66711	Sbds	SBDS_MOUSE	P70122	Shwachman-Bodian-Diamond syndrome homolog (human)	3	2	1023	51119	SBDS	SBDS_HUMAN	Q9Y3A5	c	cytoplasm	250	0
20259	Scin	ADSV_MOUSE	Q60604	scinderin	15	16	275	85477	SCIN	ADSV_HUMAN	Q9Y6U3	c	cytoskeleton	715	0
20280	Scp2	NLTP_MOUSE	P32020	sterol carrier protein 2, liver	2	2	1001	6342	SCP2	NLTP_HUMAN	P22307	c	mitochondrion	547	0
13722	Scye1	MCA1_MOUSE	P31230	small inducible cytokine subfamily E, member 1	3	3	841	9255	SCYE1	MCA1_HUMAN	Q12904	c	extracellular space	310	0
228765	Sdcbp2	SDCB2_MOUSE	Q99JZ0	syndecan binding protein (syntenin) 2	2	1	1194	27111	SDCBP2	SDCB2_HUMAN	Q9H190	c	plasma membrane	292	0
66945	Sdha	DHSA_MOUSE	Q8K2B3	succinate dehydrogenase complex, subunit A, flavoprotein (Fp)	4	3	935	6389	SDHA	DHSA_HUMAN	P31040	c	associated to mitochondrial inner membrane	664	0
67680	Sdhb	DHSB_MOUSE	Q9CQA3	succinate dehydrogenase complex, subunit B, iron sulfur (Ip)	4	4	795	6390	SDHB	DHSB_HUMAN	P21912	c	mitochondrial inner membrane	282	0
56529	Sec11a	SC11A_MOUSE	Q9R0P6	SEC11 homolog A (S. cerevisiae)	2	2	1020	23478	SEC11A	SC11A_HUMAN	P67812	c	integral to endoplasmic reticulum membrane	179	1
110379	Sec13	SEC13_MOUSE	Q9D1M0	SEC13 homolog (S. cerevisiae)	6	8	477	6396	SEC13	SEC13_HUMAN	P55735	c	cytoplasm	322	0
67815	Sec14l2	S14L2_MOUSE	Q99J08	SEC14-like 2 (S. cerevisiae)	3	2	1027	23541	SEC14L2	S14L2_HUMAN	O76054	c	cytoplasm	403	0
20334	Sec23a	SC23A_MOUSE	Q01405	SEC23A (S. cerevisiae)	22	38	129	10484	SEC23A	SC23A_HUMAN	Q15436	c	COPII vesicle coat	765	0
218811	Sec24c	Q8CGF4_MOUSE	Q8CGF4	SEC24 related gene family, member C (S. cerevisiae)	8	6	600	9632	SEC24C	SC24C_HUMAN	P53992	c	COPII vesicle coat	105	0
69608	Sec24d	Q6NXL1_MOUSE	Q6NXL1	SEC24 related gene family, member D (S. cerevisiae)	5	5	640	9871	SEC24D	SC24D_HUMAN	O94855	c	COPII vesicle coat	1032	0
69162	Sec31a	SC31A_MOUSE	Q3UPL0	SEC31 homolog A (S. cerevisiae)	12	25	200	22872	SEC31A	SC31A_HUMAN	O94979	c	COPII vesicle coat	1230	0
20341	Selenbp1	SBP1_MOUSE	P17563	selenium binding protein 1	2	4	761	8991	SELENBP1	SBP1_HUMAN	Q13228	c	associated to membrane	472	0
320213	Senp5	SENP5_MOUSE	Q6NXL6	SUMO/sentrin specific peptidase 5	4	4	798	205564	SENP5	SENP5_HUMAN	Q96HI0	c	nucleus	749	0
68607	Serhl	SERHL_MOUSE	Q9EPB5	serine hydrolase-like	2	2	1104	253190	SERHL	SEHL2_HUMAN	Q9H4I8	c	cytoplasm	311	0
20719	Serpinb6a	SPB6_MOUSE	Q60854	serine (or cysteine) peptidase inhibitor, clade B, member 6a	6	10	423	5269	SERPINB6	SPB6_HUMAN	P35237	c	cytoplasm	378	0
55948	Sfn	1433S_MOUSE	O70456	stratifin	14	41	125	2810	SFN	1433S_HUMAN	P31947	c	cytoplasm	248	0
56726	Sh3bgrl	SH3L1_MOUSE	Q9JJU8	SH3-binding domain glutamic acid-rich protein like	9	10	405	6451	SH3BGRL	SH3L1_HUMAN	O75368	c	cytoplasm	114	0
20425	Shmt1	GLYC_MOUSE	P50431	serine hydroxymethyltransferase 1 (soluble)	7	10	410	6470	SHMT1	GLYC_HUMAN	P34896	c	cytoplasm	478	0
108037	Shmt2	Q3TFD0_MOUSE	Q3TFD0	serine hydroxymethyltransferase 2 (mitochondrial)	5	4	774	6472	SHMT2	GLYM_HUMAN	P34897	c	mitochondrial inner membrane	501	0
72171	Shq1	Q7TMX5_MOUSE	Q7TMX5	SHQ1 homolog (S. cerevisiae)	3	2	1037	55164	SHQ1	SHQ1_HUMAN	Q6PI26	c	-	260	0
21402	Skp1	SKP1_MOUSE	Q9WTX5	S-phase kinase-associated protein 1A	2	2	1094	6500	SKP1	SKP1_HUMAN	P63208	c	cytosol	163	0
18408	Slc25a15	ORNT1_MOUSE	Q9WVD5	solute carrier family 25 (mitochondrial carrier ornithine transporter), member 15	2	2	1137	10166	SLC25A15	ORNT1_HUMAN	Q9Y619	c	integral to mitochondrial inner membrane	301	0
11740	Slc25a5	ADT2_MOUSE	P51881	solute carrier family 25 (mitochondrial carrier, adenine nucleotide translocator), member 5	8	17	271	292	SLC25A5	ADT2_HUMAN	P05141	c	integral to plasma membrane	298	2
26941	Slc9a3r1	NHERF_MOUSE	P70441	solute carrier family 9 (sodium/hydrogen exchanger), member 3 regulator 1	23	98	55	9368	SLC9A3R1	NHERF_HUMAN	O14745	c	apical plasma membrane	355	0
17128	Smad4	SMAD4_MOUSE	P97471	MAD homolog 4 (Drosophila)	2	1	1189	4089	SMAD4	SMAD4_HUMAN	Q13485	c	cytoplasm	551	0
20619	Snap23	SNP23_MOUSE	O09044	synaptosomal-associated protein 23	9	38	128	8773	SNAP23	SNP23_HUMAN	O00161	c	plasma membrane	210	0
67474	Snap29	SNP29_MOUSE	Q9ERB0	synaptosomal-associated protein 29	4	3	817	9342	SNAP29	SNP29_HUMAN	O95721	c	associated to plasma membrane	260	0
56440	Snx1	SNX1_MOUSE	Q9WV80	sorting nexin 1	3	4	731	6642	SNX1	SNX1_HUMAN	Q13596	c	associated to endosome membrane	522	0
67804	Snx2	SNX2_MOUSE	Q9CWK8	sorting nexin 2	4	5	620	6643	SNX2	SNX2_HUMAN	O60749	c	cytoplasm	519	0
54198	Snx3	SNX3_MOUSE	O70492	sorting nexin 3	4	4	727	8724	SNX3	SNX3_HUMAN	O60493	c	cytoplasm	162	0
69150	Snx4	SNX4_MOUSE	Q91YJ2	sorting nexin 4	2	2	1105	8723	SNX4	SNX4_HUMAN	O95219	c	cytoplasm	450	0
69178	Snx5	SNX5_MOUSE	Q9D8U8	sorting nexin 5	6	4	736	27131	SNX5	SNX5_HUMAN	Q9Y5X3	c	-	978	0
72183	Snx6	SNX6_MOUSE	Q6P8X1	sorting nexin 6	5	4	697	58533	SNX6	SNX6_HUMAN	Q9UNH7	c	cytoplasmic vesicle	406	0
76561	Snx7	SNX7_MOUSE	Q9CYI8	sorting nexin 7	3	2	1039	51375	SNX7	SNX7_HUMAN	Q9UNH6	c	cytoplasmic vesicle	387	0
66616	Snx9	SNX9_MOUSE	Q91VH2	sorting nexin 9	2	2	1152	51429	SNX9	SNX9_HUMAN	Q9Y5X1	c	plasma membrane	595	0
20779	Src	SRC_MOUSE	P05480	Rous sarcoma oncogene	14	24	207	6714	SRC	SRC_HUMAN	P12931	c	plasma membrane	541	0
109552	Sri	SORCN_MOUSE	Q6P069	sorcin	4	5	642	6717	SRI	SORCN_HUMAN	P30626	c	cytoplasm	198	0
70356	St13	F10A1_MOUSE	Q99L47	suppression of tumorigenicity 13	3	3	910	6767	ST13	F10A1_HUMAN	P50502	c	cytoplasm	371	0
70527	Stambp	STABP_MOUSE	Q9CQ26	Stam binding protein	2	2	1156	10617	STAMBP	STABP_HUMAN	O95630	c	anchored to membrane	424	0
223255	Stk24	STK24_MOUSE	Q99KH8	serine/threonine kinase 24 (STE20 homolog, yeast)	11	11	396	8428	STK24	STK24_HUMAN	Q9Y6E0	c	cytoplasm	431	0
59041	Stk25	STK25_MOUSE	Q9Z2W1	serine/threonine kinase 25 (yeast)	11	14	314	10494	STK25	STK25_HUMAN	O00506	c	Golgi apparatus	426	0
106504	Stk38	STK38_MOUSE	Q91VJ4	serine/threonine kinase 38	5	3	822	11329	STK38	STK38_HUMAN	Q15208	c	cytoplasm	465	0
20911	Stxbp2	STXB2_MOUSE	Q64324	syntaxin binding protein 2	16	27	188	6813	STXBP2	STXB2_HUMAN	Q15833	c	cellular component	593	0
20912	Stxbp3a	STXB3_MOUSE	Q60770	syntaxin binding protein 3A	6	6	1002	6814	STXBP3	STXB3_HUMAN	O00186	c	cytoplasm	592	0
56362	Sult1b1	ST1B1_MOUSE	Q9QWG7	sulfotransferase family 1B, member 1	14	30	164	27284	SULT1B1	ST1B1_HUMAN	O43704	c	cytoplasm	299	0
54200	Sult2b1	ST2B1_MOUSE	O35400	sulfotransferase family, cytosolic, 2B, member 1	3	5	671	6820	SULT2B1	ST2B1_HUMAN	O00204	c	cytoplasm	338	0
21346	Tagln2	TAGL2_MOUSE	Q9WVA4	transgelin 2	9	18	253	8407	TAGLN2	TAGL2_HUMAN	P37802	c	plasma membrane	212	0
21351	Taldo1	TALDO_MOUSE	Q93092	transaldolase 1	6	11	393	6888	TALDO1	TALDO_HUMAN	P37837	c	cytoplasm	337	0
330177	Taok3	TAOK3_MOUSE	Q8BYC6	TAO kinase 3	2	2	1051	51347	TAOK3	TAOK3_HUMAN	Q9H2K8	c	associated to plasma membrane	898	0
110960	Tars	SYTC_MOUSE	Q9D0R2	threonyl-tRNA synthetase	16	14	313	6897	TARS	SYTC_HUMAN	P26639	c	cytoplasm	722	0
67673	Tceb2	ELOB_MOUSE	P62869	transcription elongation factor B (SIII), polypeptide 2	3	3	937	6923	TCEB2	ELOB_HUMAN	Q15370	c	cytoplasm	118	0
21454	Tcp1	TCPA1_MOUSE	P11984	t-complex protein 1	7	5	633	6950	TCP1	TCPA_HUMAN	P17987	c	cytoplasm	556	0
21753	Tes	TES_MOUSE	P47226	testis derived transcript	5	5	670	26136	TES	TES_HUMAN	Q9UGI8	c	apical plasma membrane	423	0
21787	Tfg	Q8C2C6_MOUSE	Q8C2C6	Trk-fused gene	3	2	934	10342	TFG	TFG_HUMAN	Q92734	c	cytoplasm	348	0
66834	Them2	THEM2_MOUSE	Q9CQR4	thioesterase superfamily member 2	2	2	1024	55856	THEM2	THEM2_HUMAN	Q9NPJ3	c	mitochondrion	140	0
232078	Thnsl2	THNS2_MOUSE	Q80W22	threonine synthase-like 2 (bacterial)	3	2	1117	55258	THNSL2	THNS2_HUMAN	Q86YJ6	c	-	507	0
50492	Thop1	THOP1_MOUSE	Q9R0W2	thimet oligopeptidase 1	5	5	634	7064	THOP1	THOP1_HUMAN	P52888	c	cytoplasm	687	0
21881	Tkt	TKT_MOUSE	P40142	transketolase	23	43	121	7086	TKT	TKT_HUMAN	P29401	c	cytosol	623	0
238799	Tnpo1	TNPO1_MOUSE	Q8BFY9	transportin 1	4	4	701	3842	TNPO1	TNPO1_HUMAN	Q92973	c	cytoplasm	890	0
54473	Tollip	TOLIP_MOUSE	Q9QZ06	toll interacting protein	2	2	1018	54472	TOLLIP	TOLIP_HUMAN	Q9H0E2	c	cytoplasm	274	0
66314	Tpd52l2	TPD54_MOUSE	Q9CYZ2	tumor protein D52-like 2	2	2	1151	7165	TPD52L2	TPD54_HUMAN	O43399	c	cytoplasm	220	0
21991	Tpi1	TPIS_MOUSE	P17751	triosephosphate isomerase 1	12	27	181	7167	TPI1	TPIS_HUMAN	P60174	c	cytosol	249	0
22003	Tpm1	TPM1_MOUSE	P58771	tropomyosin 1, alpha	5	3	1003	7168	TPM1	TPM1_HUMAN	P09493	c	cytoskeleton	284	0
71609	Tradd	TRADD_MOUSE	Q3U0V2	TNFRSF1A-associated via death domain	10	13	333	8717	TRADD	TRADD_HUMAN	Q15628	c	cytoskeleton	310	0
74735	Trim14	TRI14_MOUSE	Q8BVW3	tripartite motif-containing 14	9	8	475	9830	TRIM14	TRI14_HUMAN	Q14142	c	cytoplasm	440	0
224762	Trim31			tripartite motif-containing 31	9	10	433	11074	TRIM31	TRI31_HUMAN	Q9BZY9	c	cytoplasm	425	0
22088	Tsg101	TS101_MOUSE	Q61187	tumor susceptibility gene 101	2	2	1004	7251	TSG101	TS101_HUMAN	Q99816	c	cytoplasm	391	0
223723	Ttll12	TTL12_MOUSE	Q9JDE2	tubulin tyrosine ligase-like family, member 12	3	2	1116	23170	TTLL12	TTL12_HUMAN	Q14166	c	-	483	0
22142	Tuba1a	TBA1A_MOUSE	P68369	tubulin, alpha 1A	9	57	87	7846	TUBA1A	TBA1A_HUMAN	Q71U36	c	microtubule	451	0
22143	Tuba1b	TBA1B_MOUSE	P05213	tubulin, alpha 1B	16	115	41	10376	TUBA1B	TBA1B_HUMAN	P68363	c	microtubule	451	0
22144	Tuba3a	TBA3_MOUSE	P05214	tubulin, alpha 3A	2	4	762	113457	TUBA3D	Q8WU19_HUMAN	Q8WUY3	c	microtubule	450	0
22147	Tuba3b	TBA3_MOUSE	P05214	tubulin, alpha 3B	2	4	763	112714	TUBA3E	TBA3E_HUMAN	Q6PEY2	c	microtubule	450	0
238463	Tubal3	TBAL3_MOUSE	Q3UX10	tubulin, alpha-like 3	5	8	492	79861	TUBAL3	TBAL3_HUMAN	A6NHL2	c	microtubule	446	0

176

Gene ID (mouse)	Gene Symbol (mouse)	Swissprot protein ID (mouse)	Swissprot protein AC (mouse)	Gene Description	Max Diff peptide	Avg Pep count	Rank order	Ortho Gene ID	Ortho Gene Symbol	Swissprot protein ID	Swissprot protein AC	Topology	GO term	Protein length	Number of TM
22151	Tubb2a	TBB2A_MOUSE	Q7TMM9	tubulin, beta 2a	11	28	179	7280	TUBB2A	TBB2A_HUMAN	Q13885	c	microtubule	445	0
73710	Tubb2b	TBB2B_MOUSE	Q9CWF2	tubulin, beta 2b	11	28	180	347733	TUBB2B	TBB2B_HUMAN	Q9BVA1	c	microtubule	445	0
227613	Tubb2c	TBB2C_MOUSE	P68372	tubulin, beta 2c	19	87	61	10383	TUBB2C	TBB2C_HUMAN	P68371	c	microtubule	445	0
22153	Tubb4	TBB4_MOUSE	Q9D6F9	tubulin, beta 4	3	1	792	10382	TUBB4	TBB4_HUMAN	P04350	c	cytoskeleton	444	0
22154	Tubb5	TBB5_MOUSE	P99024	tubulin, beta 5	19	64	76	203068	TUBB	TBB5_HUMAN	P07437	c	cytoskeleton	444	0
67951	Tubb6	TBB6_MOUSE	Q922F4	tubulin, beta 6	9	11	387	84617	TUBB6	TBB6_HUMAN	Q9BUF5	c	microtubule	447	0
19230	Twf1	TWF1_MOUSE	Q91YR1	twinfilin, actin-binding protein, homolog 1 (Drosophila)	2	2	997	5756	TWF1	TWF1_HUMAN	Q12792	c	actin cytoskeleton	350	0
22166	Txn1	THIO_MOUSE	P10639	thioredoxin 1	5	14	300	7295	TXN	THIO_HUMAN	P10599	c	cytoplasm	105	0
53382	Txnl1	TXNL1_MOUSE	Q8CDN6	thioredoxin-like 1	2	2	1014	9352	TXNL1	TXNL1_HUMAN	O43396	c	cytoplasm	289	0
50493	Txnrd1	TRXR1_MOUSE	Q9JMH6	thioredoxin reductase 1	5	5	618	7296	TXNRD1	TRXR1_HUMAN	Q16881	c	cytoplasm	613	0
227620	Uap1	UAP1L_MOUSE	Q3TW96	UDP-N-acetylglucosamine pyrophosphorylase 1-like 1	6	4	740	91373	UAP1L1	UAP1L_HUMAN	Q3KQV9	c	-	401	0
22190	Ubc	UBIQ_MOUSE	P62991	ubiquitin C	8	115	42	7316	UB	UBIQ_HUMAN	P62988	c	cytoplasm	76	0
66105	Ube2d3	UB2D3_MOUSE	P61079	ubiquitin-conjugating enzyme E2D 3 (UBC4/5 homolog, yeast)	2	1	1192	7323	UBE2D3	UB2D3_HUMAN	P61077	c	-	527	0
22192	Ube2m	UBC12_MOUSE	P61082	ubiquitin-conjugating enzyme E2M (UBC12 homolog, yeast)	4	3	812	9040	UBE2M	UBC12_HUMAN	P61081	c	cellular_component	183	0
93765	Ube2n	UBE2N_MOUSE	P61089	ubiquitin-conjugating enzyme E2N	4	4	772	7334	UBE2N	UBE2N_HUMAN	P61088	c	cytoplasm	152	0
66589	Ube2v1	UB2V1_MOUSE	Q9CZY3	ubiquitin-conjugating enzyme E2 variant 1	3	2	1022	7335	UBE2V1	UB2V1_HUMAN	Q13404	c	cytoplasm	147	0
70620	Ube2v2	UB2V2_MOUSE	Q9D2M8	ubiquitin-conjugating enzyme E2 variant 2	3	2	1033	7336	UBE2V2	UB2V2_HUMAN	Q15819	c	cytoplasm	145	0
24109	Ubl3	UBL3_MOUSE	Q9Z2M6	ubiquitin-like 3	3	2	1007	5412	UBL3	UBL3_HUMAN	O95164	c	anchored to plasma membrane	117	0
56207	Uchl5	UCHL5_MOUSE	Q9WUP7	ubiquitin carboxyl-terminal esterase L5	2	3	852	51377	UCHL5	UCHL5_HUMAN	Q9Y5K5	c	cytosol	329	0
54122	Uevld	UEVLD_MOUSE	Q3U1V6	UEV and lactate/malate dehydrogenase domains	3	2	1016	55293	UEVLD	UEVLD_HUMAN	Q8IX04	c	-	474	0
216558	Ugp2	UGPA_MOUSE	Q91ZJ5	UDP-glucose pyrophosphorylase 2	3	2	965	7360	UGP2	UGPA_HUMAN	Q16851	c	cytoplasm	508	0
67003	Uqcrc2	QCR2_MOUSE	Q9DB77	ubiquinol cytochrome c reductase core protein 2	3	3	936	7385	UQCRC2	QCR2_HUMAN	P22695	c	mitochondrial inner membrane	453	0
72088	Ush1c	USH1C_MOUSE	Q5SE64	Usher syndrome 1C homolog (human)	22	77	67	10083	USH1C	USH1C_HUMAN	Q9Y6N9	c	apical part of cell	910	0
22217	Usp12	UBP12_MOUSE	Q9D9M2	ubiquitin specific peptidase 12	2	2	1141	219333	USP12	UBP12_HUMAN	O75317	c	cellular_component	370	0
22323	Vasp	VASP_MOUSE	P70460	vasodilator-stimulated phosphoprotein	2	2	1005	7408	VASP	VASP_HUMAN	P50552	c	actin cytoskeleton	375	0
26949	Vat1	VAT1_MOUSE	Q62465	vesicle amine transport protein 1 homolog (T californica)	7	11	385	10493	VAT1	VAT1_HUMAN	Q99536	c	integral to membrane	406	0
269523	Vcp	TERA_MOUSE	Q01853	valosin containing protein	6	13	322	7415	VCP	TERA_HUMAN	P55072	c	protein complex	806	0
22334	Vdac2	VDAC2_MOUSE	Q60930	voltage-dependent anion channel 2	6	8	469	7417	VDAC2	VDAC2_HUMAN	P45880	c	integral to mitochondrial outer membrane	296	0
22335	Vdac3	VDAC3_MOUSE	Q60931	voltage-dependent anion channel 3	5	6	570	7419	VDAC3	VDAC3_HUMAN	Q9Y277	c	integral to membrane	283	0
22337	Vdr	VDR_MOUSE	P48281	vitamin D receptor	2	2	1006	7421	VDR	VDR_HUMAN	P11473	c	nucleus	422	0
22349	Vil1	VILI_MOUSE	Q62468	villin 1	58	1566	2	7429	VIL1	VILI_HUMAN	P09327	c	cytoskeleton	827	0
30930	Vps26a	VP26A_MOUSE	P40336	vacuolar protein sorting 26 homolog A (yeast)	4	4	725	9559	VPS26A	VP26A_HUMAN	O75436	c	associated to endosome membrane	327	0
66914	Vps28	VPS28_MOUSE	Q9D1C8	vacuolar protein sorting 28 (yeast)	4	2	959	51160	VPS28	VPS28_HUMAN	Q9UK41	c	cytosol	221	0
56433	Vps29	VPS29_MOUSE	Q9QZ88	vacuolar protein sorting 29 (S. pombe)	5	4	765	51699	VPS29	VPS29_HUMAN	Q9UBQ0	c	associated to membrane	182	0
65114	Vps35	VPS35_MOUSE	Q9EQH3	vacuolar protein sorting 35	11	10	406	55737	VPS35	VPS35_HUMAN	Q96QK1	c	associated to membrane	796	0
116733	Vps4a	VPS4A_MOUSE	Q8VEJ9	vacuolar protein sorting 4a (yeast)	8	9	455	27183	VPS4A	VPS4A_HUMAN	Q9UN37	c	associated to late endosome membrane	437	0
20479	Vps4b	VPS4B_MOUSE	P46467	vacuolar protein sorting 4b (yeast)	8	10	422	9525	VPS4B	VPS4B_HUMAN	O75351	c	late endosome membrane	444	0
66201	Vta1	VTA1_MOUSE	Q9CR26	Vps20-associated 1 homolog (S. cerevisiae)	5	4	732	51534	VTA1	VTA1_HUMAN	Q9NP79	c	cytoplasm	309	0
22375	Wars	SYWC_MOUSE	P32921	tryptophanyl-tRNA synthetase	7	8	470	7453	WARS	SYWC_HUMAN	P23381	c	cytoplasm	481	0
22388	Wdr1	WDR1_MOUSE	O88342	WD repeat domain 1	12	16	281	9948	WDR1	WDR1_HUMAN	O75083	c	cytoskeleton	606	0
22436	Xdh	XDH_MOUSE	Q00519	xanthine dehydrogenase	9	6	571	7498	XDH	XDH_HUMAN	P47989	c	cytosol	1335	0
103573	Xpo1	XPO1_MOUSE	Q6P5F9	exportin 1, CRM1 homolog (yeast)	2	2	1111	7514	XPO1	XPO1_HUMAN	O14980	c	cytoplasm	1071	0
22612	Yes1	YES_MOUSE	Q04736	Yamaguchi sarcoma viral (v-yes) oncogene homolog 1	14	24	211	7525	YES1	YES_HUMAN	P07947	c	cytoplasm	541	0
56418	Ykt6	YKT6_MOUSE	Q9CQW1	YKT6 homolog (S. Cerevisiae)	4	4	730	10652	YKT6	YKT6_HUMAN	O15498	c	anchored to cytoplasmic vesicle membrane	198	0
54401	Ywhab	1433B_MOUSE	Q9CQV8	tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, beta polypeptide	16	57	88	7529	YWHAB	1433B_HUMAN	P31946	c	cytoplasm	246	0
22627	Ywhae	1433E_MOUSE	P62259	tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, epsilon polypeptide	18	54	92	7531	YWHAE	1433E_HUMAN	P62258	c	cytoplasm	255	0
22628	Ywhag	1433G_MOUSE	P61982	tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, gamma polypeptide	15	37	136	7532	YWHAG	1433G_HUMAN	P61981	c	cytoplasm	247	0
22629	Ywhah	1433F_MOUSE	P68510	tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, eta polypeptide	14	48	106	7533	YWHAH	1433F_HUMAN	Q04917	c	cytoplasm	246	0
22631	Ywhaz	1433Z_MOUSE	P63101	tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, zeta polypeptide	19	84	63	7534	YWHAZ	1433Z_HUMAN	P63104	c	cytoplasm	245	0
77219	Zadh1	ZADH1_MOUSE	Q8VDQ1	zinc binding alcohol dehydrogenase, domain containing	2	2	1040	145482	PTGR2	ZADH1_HUMAN	Q8N8N7	c	cytoplasm	351	0
69036	1810010M01Rik	ZG16_MOUSE	Q8K0C5	Zymogen granule membrane protein 16 [Precursor]	5	24	213	123887	ZG16	ZG16_HUMAN	O60844	s	extracellular region	167	0
70178	2210412D01Rik	F108C_MOUSE	Q8VCV1	Abhydrolase domain-containing protein FAM108C1	3	5	648	124872	FAM108C1	F108C_HUMAN	Q6PCB6	s		320	0
70564	5730469M10Rik	CJ058_MOUSE	Q9CYH2	RIKEN cDNA 5730469M10 gene	5	4	703	84293	C10orf58	CJ058_HUMAN	Q9BRX8	s	extracellular region	218	1
70025	Acot7	BACH_MOUSE	Q91V12	acyl-CoA thioesterase 7	3	2	1059	11332	ACOT7	BACH_HUMAN	O00154	s	cytoplasm	381	0
433256	Acsl5	ACSL5_MOUSE	Q8JZR0	acyl-CoA synthetase long-chain family member 5	23	34	150	51703	ACSL5	ACSL5_HUMAN	Q9ULC5	s	integral to endoplasmic reticulum membrane	683	1
11604	Agrp	AGRP_MOUSE	P56473	agouti related protein	4	6	582	181	AGRP	AGRP_HUMAN	O00253	s	extracellular region	114	0
11657	Alb	ALBU_MOUSE	P07724	albumin	28	35	142	213	ALB	ALBU_HUMAN	P02768	s	extracellular region	608	0
109960	Amy2-1	Q61297_MOUSE	Q61297	amylase 2-1, pancreatic	2	2	1120	276	AMY1A	AMY1_HUMAN	P04745	s	extracellular region	232	0
11806	Apoa1	APOA1_MOUSE	Q00623	apolipoprotein A-I	8	6	559	335	APOA1	APOA1_HUMAN	P02647	s	extracellular region	264	0
11808	Apoa4	APOA4_MOUSE	P06728	apolipoprotein A-IV	8	11	378	337	APOA4	APOA4_HUMAN	P06727	s	extracellular region	395	0
11816	Apoe	APOE_MOUSE	P08226	apolipoprotein E	3	2	1055	348	APOE	APOE_HUMAN	P02649	s	extracellular region	311	0
108147	Atic	PUR9_MOUSE	Q9CWJ9	5-aminoimidazole-4-carboxamide ribonucleotide formyltransferase/IMP cyclohydrolase	6	8	481	471	ATIC	PUR9_HUMAN	P31939	s		592	0
11950	Atp5f1	AT5F1_MOUSE	Q9CQQ7	ATP synthase, H+ transporting, mitochondrial F0 complex, subunit b, isoform 1	4	5	676	515	ATP5F1	AT5F1_HUMAN	P24539	s	mitochondrial inner membrane	256	0
12010	B2m	B2MG_MOUSE	P01887	beta-2 microglobulin	2	4	741	567	B2M	B2MG_HUMAN	P61769	s	extracellular region	119	1
14422	B4galnt2	B4GN2_MOUSE	Q09199	beta-1,4-N-acetyl-galactosaminyl transferase 2	6	5	677	124872	B4GALNT2	B4GN2_HUMAN	Q8NHY0	s	integral to Golgi membrane	510	1
12317	Calr	CALR_MOUSE	P14211	calreticulin	2	2	1119	811	CALR	CALR_HUMAN	P27797	s	extracellular region	416	0
12351	Car4	CAH4_MOUSE	Q64444	carbonic anhydrase 4	5	8	494	762	CA4	CAH4_HUMAN	P22748	s	anchored to plasma membrane	305	0
23844	Clca3	CLCA1_MOUSE	Q9DZ6	chloride channel calcium activated	29	49	104	1179	CLCA1	CLCA1_HUMAN	A8K7I4	s	integral to plasma membrane	913	0
12764	Cmas	NEUA_MOUSE	Q99KK2	cytidine monophospho-N-acetylneuraminic acid synthetase	4	4	778	55907	CMAS	NEUA_HUMAN	Q8NFW8	s	nucleus	432	0
66588	Cmpk1	KCY_MOUSE	Q9DBP5	cytidine monophosphate (UMP-CMP) kinase 1	5	8	648	51727	CMPK1	KCY_HUMAN	P30085	s	cytoplasm	196	0
76703	Cpb1	-	-	carboxypeptidase B1 (tissue)	3	3	878	1360	CPB1	CBPB1_HUMAN	P15086	s	extracellular region	500	2
68631	Cryl1	CRYL1_MOUSE	Q99KP3	crystallin, lambda 1	5	3	877	51084	CRYL1	CRYL1_HUMAN	Q9Y2S2	s	extracellular region	319	0
12974	Cs	CISY_MOUSE	Q9CZU6	citrate synthase	4	4	779	1431	CS	CISY_HUMAN	O75390	s	mitochondrial matrix	464	0
66473	Ctrb1	CTRB1_MOUSE	Q9CR35	chymotrypsinogen B1	3	2	1058	440387	CTRB2	CTRB1_HUMAN	P17538	s	extracellular region	263	0
13040	Ctss	CATS_MOUSE	O70370	cathepsin S	4	4	742	1520	CTSS	CATS_HUMAN	P25774	s	extracellular region	340	0
65969	Cubn	CUBN_MOUSE	Q9JLB4	cubilin (intrinsic factor-cobalamin receptor)	4	3	826	8029	CUBN	CUBN_HUMAN	O60494	s	extrinsic to external side of brush border membrane	3623	0
66445	Cyc1	CY1_MOUSE	Q9D0M3	cytochrome c-1	5	4	780	1537	CYC1	CY1_HUMAN	P08574	s	integral to mitochondrial inner membrane	325	1
13113	Cyp3a13	CP3AD_MOUSE	Q64464	cytochrome P450, family 3, subfamily a, polypeptide 13	9	12	346	1577	CYP3A5	CP3A5_HUMAN	P20815	s	associated to endoplasmic reticulum membrane	503	0
13479	Dpep1	DPEP1_MOUSE	P31428	dipeptidase 1 (renal)	19	113	43	1800	DPEP1	DPEP1_HUMAN	P16444	s	anchored to apical plasma membrane	410	0
276770	Eif5a	IF5A1_MOUSE	P63242	eukaryotic translation initiation factor 5A	2	3	916	1984	EIF5A	IF5A1_HUMAN	P63241	s	cytoplasm	154	0
109901	Ela1	Q91X79_MOUSE	Q91X79	elastase 1, pancreatic	3	2	1060	1990	ELA1	ELA1_HUMAN	Q9UNI1	s	extracellular region	266	0
238011	Enpp7	A2A5N7_MOUSE	A2A5N7	ectonucleotide pyrophosphatase/phosphodiesterase 7	10	53	94	339221	ENPP7	ENPP7_HUMAN	Q6UWV6	s	anchored to apical plasma membrane	425	1
13850	Ephx2	HYES_MOUSE	P34914	epoxide hydrolase 2, cytoplasmic	11	17	268	2053	EPHX2	HYES_HUMAN	P34913	s	cytoplasm	554	0
14302	Frk	FRK_MOUSE	Q922K9	fyn-related kinase	18	45	116	2444	FRK	FRK_HUMAN	P42685	s	cytoplasm	512	0
14433	Gapdh	G3P_MOUSE	P16858	glyceraldehyde-3-phosphate dehydrogenase	15	84	64	2597	GAPDH	G3P_HUMAN	P04406	s	membrane	333	0
353172	Gars	SYG_MOUSE	Q9CZD3	glycyl-tRNA synthetase	2	2	1061	2617	GARS	SYG_HUMAN	P41250	s	secretory granule	729	0
14544	Gda	GUAD_MOUSE	Q9R111	guanine deaminase	9	16	283	9615	GDA	GUAD_HUMAN	Q9Y2T3	s	cytosol	454	0
14683	Gnas	GNAS1_MOUSE	Q6R0H7	GNAS (guanine nucleotide binding protein, alpha stimulating) complex locus	9	21	237	2778	GNAS	GNAS1_HUMAN, GNAS2_HUMAN, GNAS3_HUMAN	Q5JWF2, P63092, O95467	s		1133	0
14694	Gnb2l1	GBLP_MOUSE	P68040	guanine nucleotide binding protein (G protein), beta polypeptide 2 like 1	8	9	438	10399	GNB2L1	GBLP_HUMAN	P63244	s	cytoplasm	317	0
14719	Got2	AATM_MOUSE	P05202	glutamate oxaloacetate transaminase 2, mitochondrial	4	3	941	2806	GOT2	AATM_HUMAN	P00505	s	mitochondrial inner membrane	430	0
114664	Hsd17b11	DHB11_MOUSE	Q9EQ06	hydroxysteroid (17-beta) dehydrogenase 11	7	6	583	51170	HSD17B11	DHB11_HUMAN	Q8NBQ5	s	extracellular region	298	0
56348	Hsd17b12	DHB12_MOUSE	O70503	hydroxysteroid (17-beta) dehydrogenase 12	3	3	942	51144	HSD17B12	DHB12_HUMAN	Q53GQ0	s	integral to endoplasmic reticulum membrane	312	1
22027	Hsp90b1	ENPL_MOUSE	P08113	heat shock protein 90, beta (Grp94), member 1	10	19	245	7184	HSP90B1	ENPL_HUMAN	P14625	s	endoplasmic reticulum membrane	802	0
14828	Hspa5	GRP78_MOUSE	P20029	heat shock protein 5	13	17	269	3309	HSPA5	GRP78_HUMAN	P11021	s	integral to endoplasmic reticulum membrane	655	0
212111	Inpp5a	Q3TZT4_MOUSE	Q3TZT4	inositol polyphosphate-5-phosphatase A	5	3	827	3632	INPP5A	I5P1_HUMAN	Q14642	s	anchored to membrane	420	0
66307	Isoc1	ISOC1_MOUSE	Q91V64	isochorismatase domain containing 1	4	3	915	51015	ISOC1	ISOC1_HUMAN	Q96CN7	s	peroxisome	297	0
16429	Itln1	ITL1A_MOUSE	O88310	intelectin 1 (galactofuranose binding)	3	3	876	55600	ITLN1	ITLN1_HUMAN	Q8WWA0	s	extracellular region	313	0
228550	Itpka	IP3KA_MOUSE	Q8R071	inositol 1,4,5-trisphosphate 3-kinase A	2	1	1196	3706	ITPKA	IP3KA_HUMAN	P23677	s	cytoskeleton	459	0
56735	Krt71	K2C71_MOUSE	Q9R0H5	keratin 71	3	4	799	112802	KRT71	K2C71_HUMAN	Q3SY84	s	cytoskeleton	524	0
17110	Lyz1	LYSC_MOUSE	P17897	lysozyme 1	3	6	603	4069	LYZ	LYSC_HUMAN	P61626	s	extracellular region	148	0
110119	Mpi	MPI_MOUSE	Q924M7	mannose phosphate isomerase	2	2	1121	4351	MPI	MPI_HUMAN	P34949	s	cytoplasm	423	0
110265	Msra	MSRA_MOUSE	Q9D6Y7	methionine sulfoxide reductase A	7	11	398	4482	MSRA	MSRA_HUMAN	Q9UJ68	s	extracellular region	233	0

Gene ID (mouse)	Gene Symbol (mouse)	Swissprot protein ID (mouse)	Swissprot protein AC (mouse)	Gene Description	Max Diff peptide	Avg Pep count	Rank order	Ortho Gene ID	Ortho Gene Symbol	Swissprot protein ID	Swissprot protein AC	Topology	GO term	Protein length	Number of TM
56428	Mtch2	MTCH2_MOUSE	Q791V5	mitochondrial carrier homolog 2 (C. elegans)	3	2	972	23788	MTCH2	MTCH2_HUMAN	Q9Y6C9	s	integral to mitochondrial inner membrane	303	0
17777	Mttp	MTP_MOUSE	O08601	microsomal triglyceride transfer protein	19	27	190	4547	MTTP	MTP_HUMAN	P55157	s	endoplasmic reticulum	894	0
52793	ORF9	FAM3B_MOUSE	Q9D309	open reading frame 9	4	2	970	54097	FAM3B	FAM3B_HUMAN	P58499	s	extracellular region	235	1
18453	P4hb	PDIA1_MOUSE	P09103	prolyl 4-hydroxylase, beta polypeptide	13	12	347	5034	P4HB	PDIA1_HUMAN	P07237	s	extracellular region	509	0
14827	Pdia3	PDIA3_MOUSE	P27773	protein disulfide isomerase associated 3	10	8	478	2923	PDIA3	PDIA3_HUMAN	P30101	s	endoplasmic reticulum lumen	505	0
94184	Pdxdc1	PDXD1_MOUSE	Q99K01	pyridoxal-dependent decarboxylase domain containing 1	4	5	678	23042	PDXDC1	PDXD1_HUMAN	Q6P996	s		787	0
66171	Pgls	6PGL_MOUSE	Q9CQ60	6-phospho-gluconolactonase	7	5	625	25796	PGLS	6PGL_HUMAN	O95336	s		257	0
53328	Pgrmc1	PGRC1_MOUSE	O55022	progesterone receptor membrane component 1	4	2	971	10857	PGRMC1	PGRC1_HUMAN	O00264	s	integral to plasma membrane	195	1
18673	Phb	PHB_MOUSE	P67778	prohibitin	6	9	457	5245	PHB	PHB_HUMAN	P35232	s	integral to plasma membrane	272	0
12034	Phb2	PHB2_MOUSE	O35129	prohibitin 2	8	7	520	11331	PHB2	PHB2_HUMAN	Q99623	s	mitochondrial inner membrane	299	0
18720	Pip5k1a	PI51A_MOUSE	P70182	phosphatidylinositol-4-phosphate 5-kinase, type 1 alpha	2	2	1169	8394	PIP5K1A	PI51A_HUMAN	Q99755	s	associated to plasma membrane	546	0
19325	Rab10	RAB10_MOUSE	P61027	RAB10, member RAS oncogene family	13	35	143	10890	RAB10	RAB10_HUMAN	P61026	s	anchored to plasma membrane	200	0
19329	Rab17	RAB17_MOUSE	P35292	RAB17, member RAS oncogene family	10	14	306	64284	RAB17	RAB17_HUMAN	Q9H0T7	s	anchored to plasma membrane	214	0
216344	Rab21	RAB21_MOUSE	P35282	RAB21, member RAS oncogene family	7	11	399	23011	RAB21	RAB21_HUMAN	Q9UL25	s	anchored to Golgi membrane	222	0
19344	Rab5b	RAB5B_MOUSE	P61021	RAB5B, member RAS oncogene family	8	18	257	5869	RAB5B	RAB5B_HUMAN	P61020	s	anchored to plasma membrane	215	0
19345	Rab5c	RAB5C_MOUSE	P35278	RAB5C, member RAS oncogene family	10	52	96	5878	RAB5C	RAB5C_HUMAN	P51148	s	anchored to plasma membrane	216	0
17252	Rdh11	RDH11_MOUSE	Q9QYF1	retinol dehydrogenase 11	4	3	914	51109	RDH11	RDH11_HUMAN	Q8TC12	s	integral to endoplasmic reticulum membrane	316	0
56632	Sphk2	SPHK2_MOUSE	Q9JIA7	sphingosine kinase 2	5	8	479	56848	SPHK2	SPHK2_HUMAN	Q9NRA0	s	membrane fraction	617	0
20751	Spr	SPRE_MOUSE	Q64105	sepiapterin reductase	2	2	1056	6697	SPR	SPRE_HUMAN	P35270	s	cytoplasm	261	0
20846	Stat1	STAT1_MOUSE	P42225	signal transducer and activator of transcription 1	11	11	397	6772	STAT1	STAT1_HUMAN	P42224	s	cytoplasm	749	0
66682	Trappc5	TPPC5_MOUSE	Q9CQA1	trafficking protein particle complex 5	2	1	1195	126003	TRAPPC5	TPPC5_HUMAN	Q8IUR0	s	endoplasmic reticulum	188	0
22122	Tsta3	FCL_MOUSE	P23591	tissue specific transplantation antigen P35B	6	6	560	7264	TSTA3	FCL_HUMAN	Q13630	s	cytoplasm	321	0
105245	Txndc5	TXND5_MOUSE	Q91W90	thioredoxin domain containing 5	5	4	743	81567	TXNDC5	TXND5_HUMAN	Q8NBS9	s	endoplasmic reticulum lumen	417	0
22273	Uqcrc1	QCR1_MOUSE	Q9CZ13	ubiquinol-cytochrome c reductase core protein 1	2	2	1057	7384	UQCRC1	QCR1_HUMAN	P31930	s	mitochondrial inner membrane	480	0
22333	Vdac1	VDAC1_MOUSE	Q60932	voltage-dependent anion channel 1	11	12	355	7416	VDAC1	VDAC1_HUMAN	P21796	s	integral to plasma membrane	296	0
170745	Xpnpep2	Q91Y31_MOUSE	Q91Y31	X-prolyl aminopeptidase (aminopeptidase P) 2, membrane-bound	18	37	132	7512	XPNPEP2	XPP2_HUMAN	O43895	s	anchored to plasma membrane	674	0
68617	1110012J17Rik	K0802_MOUSE	Q3UHU5	Uncharacterized protein KIAA0802	3	2	977	23255	KIAA0802	K0802_HUMAN	Q9Y4B5	tm	cytoskeleton	1945	0
435889	1810049H19Rik	Q3V2E0_MOUSE	Q3V2E0	Adult male stomach cDNA, RIKEN full-length enriched library, clone:2210415M03 product:TESP4 (0910001B19RIK protein) (Trypsinogen 9), full insert sequence	2	9	448	-	-	-	-	tm	-	255	0
69864	1810065E05Rik	Q5NC41_MOUSE	Q5NC41	RIKEN cDNA 1810065E05 gene	2	4	804	-	-	-	-	tm	-	235	0
625286	2010003H20Rik	A2ARJ3_MOUSE	A2ARJ3	FAM23A-like	hypothetical protein LOC625286	7	53	95	653567	FAM23A	FA23A_HUMAN	Q5W0B7	tm	integral to membrane	344	5
67715	2010106E10Rik	A2ANY3_MOUSE	A2ANY3	RIKEN cDNA 2010106E10 gene	11	56	89	-	-	-	-	tm	-	279	1
69983	2010204N08Rik	-	-	hypothetical protein LOC69983	sucrase-isomaltase (alpha-glucosidase)	93	1464	3	6476	SI	SUIS_HUMAN	P14410	tm	apical plasma membrane	1818	1
72273	2210404O07Rik	YS019_MOUSE	Q0VG18	Transmembrane protein HSPC323 homolog precursor.	7	100	52	284422	LOC284422	YS019_HUMAN	O75264	tm	integral to membrane	120	1
78354	2210407C18Rik	Q6YI28_MOUSE	Q6YI28	RIKEN cDNA 2210407C18 gene / EP1 protein (novel protein)	11	176	25	-	-	-	-	tm	-	220	0
69698	2310046K01Rik	CT054_MOUSE	Q9D6X5	Uncharacterized protein C20orf54 homolog precursor.	4	11	390	113278	C20orf54	CT054_HUMAN	Q9NQ40	tm	integral to membrane	460	11
71955	2400003C14Rik	K0174_MOUSE	Q9CX00	Uncharacterized protein KIAA0174.	5	13	328	9798	KIAA0174	K0174_HUMAN	P53990	tm	ER-Golgi intermediate compartment	362	0
66674	6330409N04Rik	CLLD6_MOUSE	Q3TFQ1	Chronic lymphocytic leukemia deletion region gene 6 protein homolog	3	2	1075	57213	C13orf1	CLLD6_HUMAN	Q5W111	tm		196	0
217830	9030617O03Rik	CN159_MOUSE	Q8BH86	UPF0317 protein C14orf159 homolog, mitochondrial precursor.	2	1	1206	80017	C14orf159	CN159_HUMAN	Q7Z3D6	tm	mitochondrion	617	0
67758	Aadac	AAAD_MOUSE	Q99PG0	Arylacetamide deacetylase (EC 3.1.1.-) (AADAC).	4	4	803	13	AADAC	AAAD_HUMAN	P22760	tm	integral to endoplasmic reticulum membrane	398	0
27413	Abcb11	ABCBB_MOUSE	Q9QY30	Bile salt export pump (ATP-binding cassette sub-family B member 11) (Sister of P-glycoprotein).	2	5	653	8647	ABCB11	ABCBB_HUMAN	O95342	tm	integral to plasma membrane	1321	9
18671	Abcb1a	MDR3_MOUSE	P21447	Multidrug resistance protein 3 (EC 3.6.3.44) (ATP-binding cassette sub-family B member 1A) (P-glycoprotein 3) (MDR1A).	51	311	11	5243	ABCB1	MDR1_HUMAN	P08183	tm	integral to plasma membrane	1276	11
18669	Abcb1b	MDR1_MOUSE	P06795	ATP-binding cassette, sub-family B (MDR/TAP), member 1B	27	19	246	-	-	-	-	tm	apical plasma membrane	1276	10
74104	Abcb6	ABCB6_MOUSE	Q9DC29	Mitochondrial ATP-binding cassette sub-family B member 6.	2	3	948	10058	ABCB6	ABCB6_HUMAN	Q9NP58	tm	integral to mitochondrial outer membrane	842	9
11306	Abcb7	ABCB7_MOUSE	Q61102	ATP-binding cassette sub-family B member 7, mitochondrial (ATP-binding cassette transporter 7) (ABC transporter 7 protein).	6	7	550	22	ABCB7	ABCB7_HUMAN	O75027	tm	integral to mitochondrial inner membrane	752	5
12780	Abcc2	MRP2_MOUSE	Q8VI47	Canalicular multispecific organic anion transporter 1 (ATP-binding cassette sub-family C member 2).	27	78	66	1244	ABCC2	MRP2_HUMAN	Q92887	tm	integral to plasma membrane	1543	16
26357	Abcg2	ABCG2_MOUSE	Q7TMS5	ATP-binding cassette sub-family G member 2 (Breast cancer resistance protein 1 homolog) (CD338 antigen).	20	63	78	9429	ABCG2	ABCG2_HUMAN	Q9UNQ0	tm	integral to plasma membrane	657	6
27409	Abcg5	ABCG5_MOUSE	Q99PE8	ATP-binding cassette sub-family G member 5 (Sterolin-1).	14	26	192	64240	ABCG5	ABCG5_HUMAN	Q9H222	tm	integral to membrane	652	6
67470	Abcg8	ABCG8_MOUSE	Q9DBM0	ATP-binding cassette sub-family G member 8 (Sterolin-2).	14	50	102	64241	ABCG8	ABCG8_HUMAN	Q9H221	tm	integral to membrane	673	6
66082	Abhd6	ABHD6_MOUSE	Q8R2Y0	Abhydrolase domain-containing protein 6 (EC 3.-.-.-).	2	2	1173	57406	ABHD6	ABHD6_HUMAN	Q9BV23	tm	integral to membrane	336	1
11363	Acadl	ACADL_MOUSE	P51174	Long-chain specific acyl-CoA dehydrogenase, mitochondrial precursor (EC 1.3.99.13) (LCAD).	4	3	917	33	ACADL	ACADL_HUMAN	P28330	tm	mitochondrial matrix	430	0
11421	Ace	ACET_MOUSE	P22967	Angiotensin-converting enzyme, somatic isoform precursor (EC 3.4.15.1) (Dipeptidyl carboxypeptidase I) (Kininase II)	51	89	59	1636	ACE	ACET_HUMAN	P22966	tm	integral to plasma membrane	732	1
70008	Ace2	ACE2_MOUSE	Q8R0I0	Angiotensin-converting enzyme 2 precursor (EC 3.4.17.-) (ACE-related carboxypeptidase).	34	174	27	59272	ACE2	ACE2_HUMAN	Q9BYF1	tm	integral to plasma membrane	805	1
11432	Acp2	PPAL_MOUSE	P24638	Lysosomal acid phosphatase precursor (EC 3.1.3.2) (LAP).	4	3	943	53	ACP2	PPAL_HUMAN	P11117	tm	integral to plasma membrane	423	2
11487	Adam10	ADA10_MOUSE	O35598	ADAM 10 precursor (EC 3.4.24.81) (A disintegrin and metalloproteinase domain 10) (Mammalian disintegrin-metalloprotease) (Kuzbanian protein homolog) (CD156c antigen).	3	2	1062	102	ADAM10	ADA10_HUMAN	O14672	tm	integral to plasma membrane	749	1
11522	Adh1	ADH1_MOUSE	P00329	Alcohol dehydrogenase 1 (EC 1.1.1.1) (Alcohol dehydrogenase A subunit) (ADH-A2).	10	24	208	126	ADH1C	ADH1G_HUMAN	P00326	tm	cytoplasm	375	0
11532	Adh5	ADHX_MOUSE	P28474	Alcohol dehydrogenase class-3 (EC 1.1.1.1)	6	4	704	128	ADH5	ADHX_HUMAN	P11766	tm	cytoplasm	374	0
69117	Adh6a	A1L3C0_MOUSE	A1L3C0	alcohol dehydrogenase 6A (class V)	11	12	364	-	-	-	-	tm	-	375	0
11529	Adh7	ADH7_MOUSE	Q64437	Alcohol dehydrogenase class 4 mu/sigma chain (EC 1.1.1.1).	3	2	973	131	ADH7	ADH7_HUMAN	P40394	tm	cytoplasm	374	0
23795	Agr2	AGR2_MOUSE	O88312	Anterior gradient protein 2 homolog precursor (mAG-2) (AG-2) (Secreted cement gland protein XAG-2 homolog) (Protein Gob-4).	5	8	503	10551	AGR2	AGR2_HUMAN	O95994	tm	extracellular region	175	1
11648	Akp3	PPBI_MOUSE	P24822	Intestinal alkaline phosphatase precursor (EC 3.1.3.1) (IAP).	28	137	32	248	ALPI	PPBI_HUMAN	P09923	tm	anchored to plasma membrane	559	1
11650	Akp5	PPBE_MOUSE	P24823	Embryonic alkaline phosphatase precursor (EC 3.1.3.1) (EAP).	5	8	495	250	ALPP	ALPPL2	PPB1_HUMAN	P05187	tm	anchored to plasma membrane	529	1
69748	Aldh16a1	A16A1_MOUSE	Q571I9	aldehyde dehydrogenase 16 family, member A1	10	17	270	126133	ALDH16A1	A16A1_HUMAN	Q8IZ83	tm	-	802	0
11671	Aldh3a2	AL3A2_MOUSE	P47740	Fatty aldehyde dehydrogenase (EC 1.2.1.3) (Aldehyde dehydrogenase, microsomal) (Aldehyde dehydrogenase family 3 member A2) (Aldehyde dehydrogenase 10).	3	3	918	224	ALDH3A2	AL3A2_HUMAN	P51648	tm	integral to endoplasmic reticulum membrane	484	1
67689	Aldh3b1	AL3B1_MOUSE	Q80VQ0	aldehyde dehydrogenase 3 family, member B1	17	23	218	221	ALDH3B1	AL3B1_HUMAN	P43353	tm	-	468	0
93835	Amn	AMNLS_MOUSE	Q99JB7	Amnionless protein precursor.	2	2	1127	81693	AMN	AMNLS_HUMAN	Q9BXJ7	tm	integral to membrane	458	1
16790	Anpep	AMPN_MOUSE	P97449	Aminopeptidase N (EC 3.4.11.2) (mAPN) (Alanyl aminopeptidase) (Microsomal aminopeptidase) (Aminopeptidase M) (Membrane protein p161) (CD13 antigen).	54	961	6	290	ANPEP	AMPN_HUMAN	P15144	tm	integral to plasma membrane	966	1
11767	Ap1m1	AP1M1_MOUSE	P35585	AP-1 complex subunit mu-1 (Adaptor-related protein complex 1 mu-1 subunit)	7	7	551	8907	AP1M1	AP1M1_HUMAN	Q9BXS5	tm	associated to membrane	423	0
71770	Ap2b1	AP2B1_MOUSE	Q9DBG3	AP-2 complex subunit beta-1 (Adapter-related protein complex 2 beta-1 subunit) (Beta2-adaptin) (Beta-adaptin)	6	6	591	163	AP2B1	AP2B1_HUMAN	P63010	tm	associated to plasma membrane	937	0
11826	Aqp1	AQP1_MOUSE	Q02013	Aquaporin-1 (AQP-1) (Aquaporin-CHIP)	3	20	239	358	AQP1	AQP1_HUMAN	P29972	tm	integral to plasma membrane	269	6

Gene ID (mouse)	Gene Symbol (mouse)	Swissprot protein ID (mouse)	Swissprot protein AC (mouse)	Gene Description	Max Diff peptide	Avg Pep count	Rank order	Ortho Gene ID	Ortho Gene Symbol	Swissprot protein ID	Swissprot protein AC	Topology	GO term	Protein length	Number of TM
11842	Arf3	ARF3_MOUSE	P61205	ADP-ribosylation factor 3.	8	30	165	377	ARF3	ARF3_HUMAN	P61204	tm	anchored to Golgi membrane	181	0
70497	Arhgap17	RHG17_MOUSE	Q3UIA2	Rho GTPase-activating protein 17 (Rho-type GTPase-activating protein 17)	2	3	890	55114	ARHGAP17	RHG17_HUMAN	Q68EM7	tm	associated to plasma membrane	846	0
54208	Arl6ip1	AR6P1_MOUSE	Q9JKW0	ADP-ribosylation factor-like protein 6-interacting protein 1 (ARL-6-interacting protein 1) (Aip-1) (Protein TBX2).	2	3	923	23204	ARL6IP1	AR6P1_HUMAN	Q15041	tm	integral to membrane	203	4
54447	Asah2	ASAH2_MOUSE	Q9JHE3	Neutral ceramidase (EC 3.5.1.23) (N-CDase)	17	63	77	56624	ASAH2	ASAH2_HUMAN	Q9NR71	tm	integral to plasma membrane	756	1
11928	Atp1a1	AT1A1_MOUSE	Q8VDN2	Sodium/potassium-transporting ATPase subunit alpha-1 precursor (EC 3.6.3.9)	36	244	18	476	ATP1A1	AT1A1_HUMAN	P05023	tm	integral to membrane	1023	10
98660	Atp1a2	AT1A2_MOUSE	Q6PIE5	Sodium/potassium-transporting ATPase subunit alpha-2 precursor (EC 3.6.3.9)	3	5	656	477	ATP1A2	AT1A2_HUMAN	P50993	tm	integral to plasma membrane	1020	8
232975	Atp1a3	AT1A3_MOUSE	Q6PIC6	Sodium/potassium-transporting ATPase subunit alpha-3 (EC 3.6.3.9)	3	5	659	478	ATP1A3	AT1A3_HUMAN	P13637	tm	integral to plasma membrane	1013	8
11931	Atp1b1	AT1B1_MOUSE	P14094	Sodium/potassium-transporting ATPase subunit beta-1	8	35	144	481	ATP1B1	AT1B1_HUMAN	P05026	tm	basolateral plasma membrane	304	1
67972	Atp2b1	Q05CJ5_MOUSE	Q05CJ5	ATPase, Ca++ transporting, plasma membrane 1	14	22	224	490	ATP2B1	AT2B1_HUMAN	P20020	tm	integral to membrane	313	0
11941	Atp2b2	AT2B2_MOUSE	Q9R0K7	Plasma membrane calcium-transporting ATPase 2 (EC 3.6.3.8)	2	3	828	491	ATP2B2	AT2B2_HUMAN	Q01814	tm	integral to membrane	1198	8
381290	Atp2b4	Q32ME1_MOUSE	Q32ME1	ATPase, Ca++ transporting, plasma membrane 4	2	3	893	493	ATP2B4	AT2B4_HUMAN	P23634	tm	integral to membrane	1120	7
50769	Atp8a2	AT8A2_MOUSE	P98200	Probable phospholipid-transporting ATPase IB (EC 3.6.3.1)	3	3	832	51761	ATP8A2	AT8A2_HUMAN	Q9NTI2	tm	integral to membrane	1148	8
54670	Atp8b1	Q6R964_MOUSE	Q6R964	ATPase, class I, type 8B, member 1	24	48	108	5205	ATP8B1	AT8B1_HUMAN	O43520	tm	apical plasma membrane	1251	10
12018	Bak1	BAK_MOUSE	O08734	Bcl-2 homologous antagonist/killer (Apoptosis regulator BAK).	2	2	1122	578	BAK1	BAK_HUMAN	Q16611	tm	mitochondrial outer membrane	208	3
192970	BC022224	DHR11_MOUSE	Q3U0B3	Dehydrogenase/reductase SDR family member 11 precursor (EC 1.-.-.-).	6	5	658	79154	MGC4172	DHR11_HUMAN	Q6UWP2	tm	extracellular region	260	1
230579	BC026682	F151A_MOUSE	Q8QZW3	Protein FAM151A.	3	5	627	338094	FAM151A	F151A_HUMAN	Q8WW52	tm	integral to membrane	608	1
27061	Bcap31	BAP31_MOUSE	Q61335	B-cell receptor-associated protein 31 (BCR-associated protein Bap31) (p28 Bap31).	4	3	922	10134	BCAP31	BAP31_HUMAN	P51572	tm	integral to membrane	245	2
12215	Bsg	BASI_MOUSE	P18572	Basigin precursor (Basic immunoglobulin superfamily) (Membrane glycoprotein gp42)	8	19	242	682	BSG	BASI_HUMAN	P35613	tm	integral to membrane	389	1
12182	Bst1	BST1_MOUSE	Q64277	ADP-ribosyl cyclase 2 precursor (EC 3.2.2.5) (Cyclic ADP-ribose hydrolase 2)	8	34	151	683	BST1	BST1_HUMAN	Q10588	tm	anchored to plasma membrane	311	1
108058	Camk2d	KCC2D_MOUSE	Q6PHZ2	Calcium/calmodulin-dependent protein kinase type II delta chain (EC 2.7.11.17) (CaM-kinase II delta chain)	5	5	684	817	CAMK2D	KCC2D_HUMAN	Q13557	tm	cytoplasm	499	0
12330	Canx	CALX_MOUSE	P35564	Calnexin precursor.	5	4	744	821	CANX	CALX_HUMAN	P27824	tm	integral to endoplasmic reticulum membrane	591	1
12389	Cav1	CAV1_MOUSE	P49817	Caveolin-1	3	3	919	857	CAV1	CAV1_HUMAN	Q03135	tm	integral to plasma membrane	178	1
12469	Cct8	TCPQ_MOUSE	P42932	T-complex protein 1 subunit theta (TCP-1-theta) (CCT-theta).	11	8	509	10694	CCT8	TCPQ_HUMAN	P50990	tm	cytoplasm	548	0
12476	Cd151	CD151_MOUSE	O35566	CD151 antigen (Platelet-endothelial tetraspan antigen 3)	2	2	1063	977	CD151	CD151_HUMAN	P48509	tm	integral to membrane	253	4
12491	Cd36	CD36_MOUSE	Q08857	Platelet glycoprotein 4 (Platelet glycoprotein IV)	9	13	335	948	CD36	CD36_HUMAN	P16671	tm	integral to membrane	472	2
12494	Cd38	CD38_MOUSE	P56528	ADP-ribosyl cyclase 1 (EC 3.2.2.5) (Cyclic ADP-ribose hydrolase 1)	4	6	584	952	CD38	CD38_HUMAN	P28907	tm	integral to membrane	304	1
16423	Cd47	CD47_MOUSE	Q61735	Leukocyte surface antigen CD47 precursor (Integrin-associated protein) (IAP).	2	2	975	961	CD47	CD47_HUMAN	Q08722	tm	integral to membrane	303	5
12520	Cd81	CD81_MOUSE	P35762	CD81 antigen (26 kDa cell surface protein TAPA-1)	2	2	1123	975	CD81	CD81_HUMAN	P60033	tm	integral to membrane	236	4
12521	Cd82	CD82_MOUSE	P40237	CD82 antigen (Inducible membrane protein R2) (C33 antigen) (IA4).	3	11	400	3732	CD82	CD82_HUMAN	P27701	tm	integral to membrane	266	4
12550	Cdh1	CADH1_MOUSE	P09803	Epithelial cadherin precursor (E-cadherin)	3	5	649	999	CDH1	CADH1_HUMAN	P12830	tm	integral to plasma membrane	884	1
12557	Cdh17	CAD17_MOUSE	Q9R100	Cadherin-17 precursor (Liver-intestine cadherin)	11	45	117	1015	CDH17	CAD17_HUMAN	Q12864	tm	basolateral plasma membrane	827	1
26365	Ceacam1	CEAM1_MOUSE	P31809	Carcinoembryonic antigen-related cell adhesion molecule 1 precursor (Biliary glycoprotein 1) (BGP-1)	4	32	153	634	CEACAM1	CEAM1_HUMAN	P13688	tm	integral to membrane	521	1
72431	Ceacam18	Q9D871_MOUSE	Q9D871	CEA-related cell adhesion molecule 1	7	5	626	-	-	-	-	tm	integral to membrane	376	1
71601	Ceacam20	Q80Y42_MOUSE	Q80Y42	CEA-related cell adhesion molecule 20	15	21	234	125931	CEACAM20	CEA20_HUMAN	Q6UY09	tm	integral to membrane	577	1
107934	Celsr3	CELR3_MOUSE	Q91ZI0	Cadherin EGF LAG seven-pass G-type receptor 3 precursor.	3	3	891	1951	CELSR3	CELR3_HUMAN	Q9NYQ7	tm	integral to membrane	3301	7
67006	Cisd2	CISD2_MOUSE	Q9CQ85	CDGSH iron sulfur domain-containing protein 2 (MitoNEET-related 1 protein)	2	2	1175	493856	CISD2	CISD2_HUMAN	Q8N5K1	tm	integral to endoplasmic reticulum membrane	135	1
12716	Ckmt1	KCRU_MOUSE	P30275	Creatine kinase, ubiquitous mitochondrial precursor (EC 2.7.3.2) (U- MtCK) (Acidic-type mitochondrial creatine kinase) (Mia-CK).	5	8	496	548596, 1159	CKMT1A, CKMT1B	KCRU_HUMAN	P12532	tm	anchored to mitochondrial inner membrane	418	0
99663	Clca6	Q6Q473_MOUSE	Q6Q473	chloride channel calcium activated 6	13	43	120	22802	CLCA4	Q14CN2_HUMAN	Q14CN2	tm	-	924	1
60363	Cldn15	CLD15_MOUSE	Q9Z0S5	Claudin-15.	2	3	924	24146	CLDN15	CLD15_HUMAN	P56746	tm	integral to plasma membrane	227	4
12739	Cldn3	CLD3_MOUSE	Q9Z0G9	Claudin-3 (Clostridium perfringens enterotoxin receptor 2) (CPE-receptor 2) (CPE-R 2).	3	9	458	1365	CLDN3	CLD3_HUMAN	O15551	tm	integral to plasma membrane	219	4
29876	Clic4	CLIC4_MOUSE	Q9QYB1	Chloride intracellular channel protein 4 (mc3s5/mtCLIC).	6	4	750	25932	CLIC4	CLIC4_HUMAN	Q9Y696	tm	integral to membrane	253	0
216705	Clint1	EPN4_MOUSE	Q99KN9	Clathrin interactor 1 (Epsin-4) (Epsin-related protein) (EpsinR) (Enthoprotin).	3	4	753	9685	CLINT1	EPN4_HUMAN	Q14677	tm	associated to Golgi membrane	631	0
212070	Clrn3	CLRN3_MOUSE	Q8BHH8	Clarin-3 (Transmembrane protein 12).	3	31	159	119467	CLRN3	CLRN3_HUMAN	Q8NCR9	tm	integral to membrane	226	4
94220	Cnnm4	CNNM4_MOUSE	Q69ZF7	Metal transporter CNNM4 (Cyclin-M4) (Ancient conserved domain-containing protein 4) (mACDP4).	2	1	1204	26504	CNNM4	CNNM4_HUMAN	Q6P4Q7	tm	integral to plasma membrane	771	4
23789	Coro1b	COR1B_MOUSE	Q9WUM3	Coronin-1B (Coronin-2).	8	9	439	57175	CORO1B	COR1B_HUMAN	Q9BR76	tm	cytoskeleton	484	0
23790	Coro1c	COR1C_MOUSE	Q9WUM4	Coronin-1C (Coronin-3).	11	15	296	23603	CORO1C	COR1C_HUMAN	Q9ULV4	tm	actin cytoskeleton	474	0
107684	Coro2a	COR2A_MOUSE	Q8C0P5	coronin, actin binding protein 2A	11	10	416	7464	CORO2A	COR2A_HUMAN	Q92828	tm	-	524	0
12857	Cox4i1	COX41_MOUSE	P19783	Cytochrome c oxidase subunit 4 isoform 1, mitochondrial precursor	8	4	482	1327	COX4I1	COX41_HUMAN	P13073	tm	integral to mitochondrial inner membrane	169	1
70568	Cpne3	CPNE3_MOUSE	Q8BT60	Copine-3 (Copine III).	7	9	461	8895	CPNE3	CPNE3_HUMAN	O75131	tm	cytoplasm	533	0
12946	Cr1l	CRRY_MOUSE	Q64735	Complement regulatory protein Crry precursor (Protein p65).	3	3	879	1378	CR1	CR1_HUMAN	P17927	tm	integral to plasma membrane	483	1
13052	Cxadr	CXAR_MOUSE	P97792	Coxsackievirus and adenovirus receptor homolog precursor (mCAR) (CAR).	4	10	414	1525	CXADR	CXAR_HUMAN	P78310	tm	basolateral plasma membrane	365	1
20430	Cyfip1	CYFP1_MOUSE	Q7TMB8	Cytoplasmic FMR1-interacting protein 1 (Specifically Rac1-associated protein 1) (Sra-1).	9	7	521	23191	CYFIP1	CYFP1_HUMAN	Q7L576	tm	cytoplasm	1253	0
13088	Cyp2b10	CP2BA_MOUSE	P12791	Cytochrome P450 2B10 (EC 1.14.14.1) (CYPIIB10)	10	9	445	1555	CYP2B6	CP2B6_HUMAN	P20813	tm	integral to endoplasmic reticulum membrane	500	1
72303	Cyp2c65	Q148B1_MOUSE	Q148B1	cytochrome P450, family 2, subfamily c, polypeptide 65	8	8	513	1558	CYP2C8	CP2C8_HUMAN	P10632	tm	integral to endoplasmic reticulum membrane	490	1
27999	D6Wsu176e	FAM3C_MOUSE	Q91VU0	Protein FAM3C precursor.	5	4	802	10447	FAM3C	FAM3C_HUMAN	Q92520	tm	extracellular region	227	1
13200	Ddost	OST48_MOUSE	O54734	Dolichyl-diphosphooligosaccharide--protein glycosyltransferase 48 kDa subunit precursor (EC 2.4.1.119)	6	3	880	1650	DDOST	OST48_HUMAN	P39656	tm	integral to endoplasmic reticulum membrane	441	2
13002	Dnajc5	DNJC5_MOUSE	P60904	DnaJ homolog subfamily C member 5 (Cysteine string protein) (CSP).	3	9	459	80331	DNAJC5	DNJC5_HUMAN	Q9H3Z4	tm	anchored to melanosome membrane	198	0
13482	Dpp4	DPP4_MOUSE	P28843	Dipeptidyl peptidase 4 (EC 3.4.14.5) (Dipeptidyl peptidase IV) (DPP IV)	33	107	45	1803	DPP4	DPP4_HUMAN	P27487	tm	integral to plasma membrane	760	1
214593	Duox2	A2AQ99_MOUSE	A2AQ99	dual oxidase 2	18	24	209	50506	DUOX2	DUOX2_HUMAN	Q9NRD8	tm	apical plasma membrane	1517	6
66811	Duoxa2	DOXA2_MOUSE	Q9D311	Dual oxidase maturation factor 2.	4	5	654	405753	DUOXA2	DOXA2_HUMAN	Q1HG44	tm	integral to endoplasmic reticulum membrane	320	5
13595	Ebp	EBP_MOUSE	P70245	3-beta-hydroxysteroid-Delta(8),Delta(7)-isomerase (EC 5.3.3.5) (Cholesterol Delta-isomerase)	2	2	1171	10682	EBP	EBP_HUMAN	Q15125	tm	integral to membrane	230	5
13641	Efnb1	EFNB1_MOUSE	P52795	Ephrin-B1 precursor (EPH-related receptor tyrosine kinase ligand 2) (LERK-2)	2	2	1064	1947	EFNB1	EFNB1_HUMAN	P98172	tm	integral to membrane	345	1
13809	Enpep	AMPE_MOUSE	P16406	Glutamyl aminopeptidase (EC 3.4.11.7) (EAP) (Aminopeptidase A)	53	431	10	2028	ENPEP	AMPE_HUMAN	Q07075	tm	apical plasma membrane	945	1
209558	Enpp3	ENPP3_MOUSE	Q6DYE8	Ectonucleotide pyrophosphatase/phosphodiesterase family member 3 (E- NPP 3) (Phosphodiesterase I/nucleotide pyrophosphatase 3)	26	112	44	5169	ENPP3	ENPP3_HUMAN	O14638	tm	integral to membrane	874	1
72090	Entpd8	ENTP8_MOUSE	Q8K0L2	Ectonucleoside triphosphate diphosphohydrolase 8 (EC 3.6.1.5) (E- NTPDase 8) (NTPDase 8) (NTPDase8).	5	14	310	377841	ENTPD8	ENTP8_HUMAN	Q5MY95	tm	integral to membrane	497	2
244373	Erlin2	ERLN2_MOUSE	Q8BFZ9	Erlin-2 (Endoplasmic reticulum lipid raft-associated protein 2)	4	5	685	11160	ERLIN2	ERLN2_HUMAN	O94905	tm	integral to endoplasmic reticulum membrane	340	0
16456	F11r	JAM1_MOUSE	O88792	Junctional adhesion molecule A precursor (JAM-A) (Junctional adhesion molecule 1) (JAM-1) (CD321 antigen).	4	8	498	50848	F11R	JAM1_HUMAN	Q9Y624	tm	integral to plasma membrane	300	1
14104	Fasn	FAS_MOUSE	P19096	Fatty acid synthase (EC 2.3.1.85)	30	31	158	2194	FASN	FAS_HUMAN	P49327	Topology	cytoplasm	2504	0
66437	Fis1	FIS1_MOUSE	Q9CQ92	Mitochondrial fission 1 protein (Fis1 homolog)	2	1	1201	51024	FIS1	FIS1_HUMAN	Q9Y3D6	tm	membrane	152	1

Gene ID (mouse)	Gene Symbol (mouse)	Swissprot protein ID (mouse)	Swissprot protein AC (mouse)	Gene Description	Max Diff peptide	Avg Pep count	Rank order	Ortho Gene ID	Ortho Gene Symbol	Swissprot protein ID	Swissprot protein AC	Topology	GO term	Protein length	Number of TM
14257	Flt4	VGFR3_MOUSE	P35917	Vascular endothelial growth factor receptor 3 precursor (EC 2.7.10.1) (VEGFR-3) (Tyrosine-protein kinase receptor FLT4).	2	2	1065	2324	FLT4	VGFR3_HUMAN	P35916	tm	integral to plasma membrane	1363	1
14263	Fmo5	FMO5_MOUSE	P97872	Dimethylaniline monooxygenase [N-oxide-forming] 5 (EC 1.14.13.8)	6	5	650	2330	FMO5	FMO5_HUMAN	P49326	tm	integral to endoplasmic reticulum membrane	533	1
69976	Galk2	GALK2_MOUSE	Q68FH4	N-acetylgalactosamine kinase (EC 2.7.1.157) (GalNAc kinase) (Galactokinase 2).	2	2	1176	2585	GALK2	GALK2_HUMAN	Q01415	tm	cytoplasm	458	0
66569	Gdpd1	GDPD1_MOUSE	Q9CRY7	Glycerophosphodiester phosphodiesterase domain-containing protein 1 (EC 3.1.-.-) (Glycerophosphodiester phosphodiesterase 4).	4	3	925	284161	GDPD1	GDPD1_HUMAN	Q8N9F7	tm	integral to membrane	314	2
14583	Gfpt1	GFPT1_MOUSE	P47856	Glucosamine--fructose-6-phosphate aminotransferase [isomerizing] 1 (EC 2.6.1.16) (Glutamine:fructose 6 phosphate amidotransferase 1)	14	14	307	2673	GFPT1	GFPT1_HUMAN	Q06210	tm	cytoplasm	697	0
14598	Ggt1	GGT1_MOUSE	Q60928	Gamma-glutamyltranspeptidase 1 precursor (EC 2.3.2.2) (Gamma-glutamyltransferase 1)	14	127	36	2678	GGT1	GGT1_HUMAN	P19440	tm	integral to membrane	568	1
14678	Gnai2	GNAI2_MOUSE	P08752	guanine nucleotide binding protein (G protein), alpha inhibiting 2	17	75	68	2771	GNAI2	GNAI2_HUMAN	P04899	tm	-	355	0
14679	Gnai3	GNAI3_MOUSE	Q9DC51	Guanine nucleotide-binding protein G(k) subunit alpha (G(i) alpha-3).	16	68	72	2773	GNAI3	GNAI3_HUMAN	P08754	tm	anchored to plasma membrane	354	0
59290	Gpa33	GPA33_MOUSE	Q9JKA5	Cell surface A33 antigen precursor (Glycoprotein A33) (mA33).	8	19	247	10223	GPA33	GPA33_HUMAN	Q99795	tm	integral to plasma membrane	319	1
14751	Gpi1	G6PI_MOUSE	P06745	Glucose-6-phosphate isomerase (EC 5.3.1.9) (GPI) (Phosphoglucose isomerase) (PGI) (Phosphohexose isomerase) (PHI) (Neuroleukin) (NLK).	12	14	308	2821	GPI	G6PI_HUMAN	P06744	tm	extracellular space	558	0
239853	Gpr128	GP128_MOUSE	Q8BM96	Probable G-protein coupled receptor 128 precursor.	11	59	83	84873	GPR128	GP128_HUMAN	Q96K78	tm	integral to plasma membrane	785	8
625249	Gpx4	GPX41_MOUSE	O70325	Phospholipid hydroperoxide glutathione peroxidase, mitochondrial precursor (EC 1.11.1.12)	7	8	515	2879	GPX4	GPX4_HUMAN	P36969	tm	mitochondrial inner membrane	197	0
66790	Grtp1	GRTP1_MOUSE	Q9D3N8	Growth hormone-regulated TBC protein 1 (TBC1 domain family member 6).	4	3	887	79774	GRTP1	GRTP1_HUMAN	Q5TC63	tm	cytoplasm	359	0
14782	Gsr	GSHR_MOUSE	P47791	Glutathione reductase, mitochondrial precursor (EC 1.8.1.7) (GRase) (GR).	4	3	829	2936	GSR	GSHR_HUMAN	P00390	tm	mitochondrion	500	0
14917	Gucy2c	GUC2C_MOUSE	Q3UWA6	Heat-stable enterotoxin receptor precursor (EC 4.6.1.2) (STA receptor) (Intestinal guanylate cyclase) (GC-C).	10	16	276	2984	GUCY2C	GUC2C_HUMAN	P25092	tm	apical plasma membrane	1072	1
14933	Gyk	GLPK_MOUSE	Q64516	Glycerol kinase (EC 2.7.1.30) (ATP:glycerol 3-phosphotransferase) (Glycerokinase) (GK).	3	3	920	2710	GK	GLPK_HUMAN	P32189	tm	mitochondrial outer membrane	524	0
14960	H2-Aa	HA2B_MOUSE	P14434	H-2 class II histocompatibility antigen, A-B alpha chain precursor (IAalpha).	3	6	561	3117	HLA-DQA1	HA22_HUMAN	P04226	tm	integral to plasma membrane	256	1
14961	H2-Ab1	HB2A_MOUSE	P14483	H-2 class II histocompatibility antigen, A beta chain precursor.	7	8	497	3120	HLA-DQB2	HB2X_HUMAN	P05538	tm	integral to plasma membrane	265	1
14964	H2-D1	HA11_MOUSE	P01899	H-2 class I histocompatibility antigen, D-B alpha chain precursor (H-2D(B)).	8	12	348	3107	HLA-C	1C01_HUMAN	P30499	tm	integral to plasma membrane	362	1
15000	H2-DMb2	2DMB_MOUSE	P35737	histocompatibility 2, class II, locus Mb2	4	5	651	3109	HLA-DMB	2DMB_HUMAN	P28068	tm	integral to late endosome membrane	261	1
14980	H2-L	Q31149_MOUSE	Q31149	MHC class I-alpha	8	12	349	-	-	-	-	tm	-	365	1
15006	H2-Q1	Q66JR0_MOUSE	Q66JR0	histocompatibility 2, Q region locus 1	2	3	881	-	-	-	-	tm	-	341	1
110557	H2-Q6	P79568_MOUSE	P79568	histocompatibility 2, Q region locus 6	2	3	892	-	-	-	-	tm	-	327	0
15018	H2-Q7	HA17_MOUSE	P14429	histocompatibility 2, Q region locus 7	2	3	882	-	-	-	-	tm	integral to membrane	334	1
15043	H2-T3	HA1U_MOUSE	P14433	histocompatibility 2, T region locus 3	5	4	705	-	-	-	-	tm	integral to membrane	361	1
27400	Hsd17b6	H17B6_MOUSE	Q9R092	Hydroxysteroid 17-beta dehydrogenase 6 precursor (EC 1.1.1.62)	10	9	460	8630	HSD17B6	H17B6_HUMAN	O14756	tm	endoplasmic reticulum	317	0
15519	Hsp90aa1	HS90A_MOUSE	P07901	Heat shock protein HSP 90-alpha (HSP 86) (Tumor-specific transplantation 86 kDa antigen) (TSTA).	24	46	113	3320	HSP90AA1	HS90A_HUMAN	P07900	tm	cytoplasm	733	0
15516	Hsp90ab1	HS90B_MOUSE	P11499	Heat shock protein HSP 90-beta (HSP 84) (Tumor-specific transplantation 84 kDa antigen) (TSTA).	33	136	34	3326	HSP90AB1	HS90B_HUMAN	P08238	tm	cytoplasm	724	0
16004	Igf2r	MPRI_MOUSE	Q07113	Cation-independent mannose-6-phosphate receptor precursor (CI Man-6-P receptor)	2	2	1066	3482	IGF2R	MPRI_HUMAN	P11717	tm	integral to plasma membrane	2483	1
71927	Itfg1	TIP_MOUSE	Q99KW9	T-cell immunomodulatory protein precursor (Protein TIP) (Integrin-alpha FG-GAP repeat-containing protein 1).	5	5	655	81533	ITFG1	TIP_HUMAN	Q8TB96	tm	integral to membrane	610	1
106581	Itfg3	ITFG3_MOUSE	Q8C0Z1	integrin alpha FG-GAP repeat containing 3	2	3	928	83986	ITFG3	ITFG3_HUMAN	Q9H0X4	tm	integral to membrane	555	1
16400	Itga3	ITA3_MOUSE	Q62470	Integrin alpha-3 precursor (Galactoprotein B3) (GAPB3) (VLA-3 alpha chain)	2	3	883	3675	ITGA3	ITA3_HUMAN	P26006	tm	integral to plasma membrane	1053	2
16403	Itga6	ITA6_MOUSE	Q61739	Integrin alpha-6 precursor (VLA-6) (CD49 antigen-like family member F) (CD49f antigen) [Contains: Integrin alpha-6 heavy chain; Integrin alpha-6 light chain].	3	4	745	3655	ITGA6	ITA6_HUMAN	P23229	tm	integral to plasma membrane	1091	1
16412	Itgb1	ITB1_MOUSE	P09055	Integrin beta-1 precursor (Fibronectin receptor subunit beta) (Integrin VLA-4 subunit beta) (CD29 antigen).	5	4	706	3688	ITGB1	ITB1_HUMAN	P05556	tm	integral to plasma membrane	798	1
545156	Kalrn	KALRN_MOUSE	A2CG49	Kalirin (EC 2.7.11.1) (Protein Duo) (Serine/threonine kinase with Dbl- and pleckstrin homology domain).	9	23	215	8997	KALRN	KALRN_HUMAN	O60229	tm	actin cytoskeleton	2964	0
16548	Khk	KHK_MOUSE	P97328	Ketohexokinase (EC 2.7.1.3) (Hepatic fructokinase).	6	8	510	3795	KHK	KHK_HUMAN	P50053	tm	cytoplasm	298	0
385643	Kng2	KNG1_MOUSE	O08677	kininogen 2	4	3	835	3827	KNG1	KNG1_HUMAN	P01042	tm	extracellular region	661	0
16783	Lamp1	LAMP1_MOUSE	P11438	Lysosome-associated membrane glycoprotein 1 precursor (LAMP-1) (Lysosomal membrane glycoprotein A)	4	9	446	3916	LAMP1	LAMP1_HUMAN	P11279	tm	integral to lysosomal membrane	406	1
16784	Lamp2	LAMP2_MOUSE	P17047	Lysosome-associated membrane glycoprotein 2 precursor (LAMP-2)	4	6	585	3920	LAMP2	LAMP2_HUMAN	P13473	tm	integral to plasma membrane	415	1
16818	Lck	LCK_MOUSE	P06240	Proto-oncogene tyrosine-protein kinase LCK (EC 2.7.10.2) (Lymphocyte cell-specific protein-tyrosine kinase) (p56-LCK) (LSK).	2	2	1067	3932	LCK	LCK_HUMAN	P06239	tm	anchored to plasma membrane	509	0
226413	Lct	Q3TZ78_MOUSE	Q3TZ78	lactase	47	251	15	3938	LCT	LPH_HUMAN	P09848	tm	apical plasma membrane	1220	0
16828	Ldha	LDHA_MOUSE	P06151	L-lactate dehydrogenase A chain (EC 1.1.1.27) (LDH-A)	16	52	98	3939	LDHA	LDHA_HUMAN	P00338	tm	cytoplasm	332	0
70361	Lman1	LMAN1_MOUSE	Q9D0F3	Protein ERGIC-53 precursor (ER-Golgi intermediate compartment 53 kDa protein) (Lectin mannose-binding 1) (p58).	4	3	889	3998	LMAN1	LMAN1_HUMAN	P49257	tm	integral to endoplasmic reticulum membrane	517	1
66890	Lman2	LMAN2_MOUSE	Q9DBH5	Vesicular integral-membrane protein VIP36 precursor (Lectin mannose- binding 2)	4	3	888	10960	LMAN2	LMAN2_HUMAN	Q12907	tm	integral to ER-Golgi intermediate compartment	358	1
231876	Lmtk2	LMTK2_MOUSE	Q3TYD6	Serine/threonine-protein kinase LMTK2 precursor (EC 2.7.11.1) (Lemur tyrosine kinase 2) (Brain-enriched kinase).	2	2	1182	22853	LMTK2	LMTK2_HUMAN	Q8IWU2	tm	integral to membrane	1471	2
16971	Lrp1	LRP1_MOUSE	Q91ZX7	Prolow-density lipoprotein receptor-related protein 1 precursor (LRP) (Alpha-2-macroglobulin receptor) (A2MR)	8	6	586	4035	LRP1	LRP1_HUMAN	Q07954	tm	integral to plasma membrane	4545	1
17113	M6pr	MPRD_MOUSE	P24668	Cation-dependent mannose-6-phosphate receptor precursor (CD Man-6-P receptor)	5	6	587	4074	M6PR	MPRD_HUMAN	P20645	tm	integral to plasma membrane	278	1
17161	Maoa	AOFA_MOUSE	Q64133	Amine oxidase [flavin-containing] A (EC 1.4.3.4) (Monoamine oxidase type A) (MAO-A).	10	14	317	4128	MAOA	AOFA_HUMAN	P21397	tm	integral to mitochondrial outer membrane	526	1
26396	Map2k2	MP2K2_MOUSE	Q63932	Dual specificity mitogen-activated protein kinase kinase 2 (EC 2.7.12.2)	6	8	504	5605	MAP2K2	MP2K2_HUMAN	P36507	tm	cytoplasm	401	0
26420	Mapk9	MK09_MOUSE	Q9WTU6	mitogen-activated protein kinase 9	2	3	921	5601	MAPK9	MK09_HUMAN	P45984	tm	-	423	0
17287	Mep1a	MEP1A_MOUSE	P28825	Meprin A subunit alpha precursor (EC 3.4.24.18) (Endopeptidase-2) (MEP-1).	11	35	145	4224	MEP1A	MEP1A_HUMAN	Q16819	tm	integral to plasma membrane	747	1
17288	Mep1b	MEP1B_MOUSE	Q61847	Meprin A subunit beta precursor (EC 3.4.24.18) (Endopeptidase-2).	23	204	23	4225	MEP1B	MEP1B_HUMAN	Q16820	tm	integral to plasma membrane	704	1
17380	Mme	NEP_MOUSE	Q61391	Neprilysin (EC 3.4.24.11) (Neutral endopeptidase 24.11)	39	175	26	4311	MME	NEP_HUMAN	P08473	tm	integral to plasma membrane	750	1
233549	Mogat2	MOGT2_MOUSE	Q80W94	2-acylglycerol O-acyltransferase 2 (EC 2.3.1.22) (Monoacylglycerol O-acyltransferase 2)	5	5	660	80168	MOGAT2	MOGT2_HUMAN	Q3SYC2	tm	integral to endoplasmic reticulum membrane	334	1
67247	Mosc2	MOSC2_MOUSE	Q922Q1	MOSC domain-containing protein 2, mitochondrial precursor (EC 1.-.-.-).	6	6	605	54996	MOSC2	MOSC2_HUMAN	Q969Z3	tm	associated to mitochondrial outer membrane	338	1

180

Gene ID (mouse)	Gene Symbol (mouse)	Swissprot protein ID (mouse)	Swissprot protein AC (mouse)	Gene Description	Max Diff peptide	Avg Pep count	Rank order	Ortho Gene ID	Ortho Gene Symbol	Swissprot protein ID	Swissprot protein AC	Topology	GO term	Protein length	Number of TM
69826	Ms4a10	M4A10_MOUSE	Q99N03	Membrane-spanning 4-domains subfamily A member 10.	5	14	318	341116	MS4A10	M4A10_HUMAN	Q96PG2	tm	integral to membrane	267	4
17709	mt-Co2	COX2_MOUSE	P00405	Cytochrome c oxidase subunit 2 (Cytochrome c oxidase polypeptide II).	2	1	1197	4513	MT-CO2	COX2_HUMAN	P00403	tm	integral to mitochondrial inner membrane	227	2
17063	Muc13	MUC13_MOUSE	P19467	Mucin-13 precursor (MUC-13) (Cell surface antigen 114/A10) (Lymphocyte antigen 64).	13	219	22	56667	MUC13	MUC13_HUMAN	Q9H3R2	tm	integral to plasma membrane	573	1
666339	Muc3	Q6PDD8_MOUSE	Q6PDD8	mucin 3, intestinal	8	16	284	140453	MUC17 | MUC3A	MUC17_HUMAN	Q685J3	tm	anchored to plasma membrane	275	1
72040	Mupcdh	MUCDL_MOUSE	Q8VHF2	Mucin and cadherin-like protein precursor (Mu-protocadherin).	13	100	51	53841	MUPCDH	MUCDL_HUMAN	Q9HBB8	tm	integral to plasma membrane	831	1
17922	Myo7b	MYO7B_MOUSE	Q99MZ6	Myosin-VIIb.	108	733	8	4648	MYO7B	MYO7B_HUMAN	Q6PIF6	tm	cytoskeleton	2113	0
381204	Naaladl1	NALDL_MOUSE	Q7M758	N-acetylated-alpha-linked acidic dipeptidase-like protein (EC 3.4.17.21) (NAALADase L).	26	101	50	10004	NAALADL1	NALDL_HUMAN	Q9UQQ1	tm	apical plasma membrane	745	1
94181	Nans	Q99J77_MOUSE	Q99J77	N-acetylneuraminic acid synthase (sialic acid synthase)	2	2	1083	54187	NANS	SIAS_HUMAN	Q9NR45	tm	cytoplasm	359	0
59287	Ncstn	NICA_MOUSE	P57716	Nicastrin precursor.	6	7	531	23385	NCSTN	NICA_HUMAN	Q92542	tm	integral to plasma membrane	708	1
101613	Nlrp6	NALP6_MOUSE	Q91WS2	NACHT, LRR and PYD domains-containing protein 6 (PYRIN-containing APAF1-like protein 5-like).	12	11	383	171389	NLRP6	NALP6_HUMAN	P59044	tm	cytoplasm	869	0
18145	Npc1	NPC1_MOUSE	O35604	Niemann-Pick C1 protein precursor.	2	6	562	4864	NPC1	NPC1_HUMAN	O15118	tm	integral to plasma membrane	1278	12
237636	Npc1l1	NPCL1_MOUSE	Q6T3U4	Niemann-Pick C1-like protein 1 precursor.	32	142	29	29881	NPC1L1	NPCL1_HUMAN	Q9UHC9	tm	apical plasma membrane	1333	13
20320	Nptn	NPTN_MOUSE	P97300	Neuroplastin precursor (Stromal cell derived receptor 1) (SDR-1).	3	4	708	27020	NPTN	NPTN_HUMAN	Q9Y639	tm	integral to plasma membrane	281	1
18300	Oit1	FAM3D_MOUSE	P97805	Protein FAM3D precursor (Oncoprotein-induced protein 1) (Protein EF-7).	2	4	800	131177	FAM3D	FAM3D_HUMAN	Q96BQ1	tm	extracellular region	223	1
330962	Ostb	OSTB_MOUSE	Q80WK2	Organic solute transporter subunit beta (OST-beta).	2	2	1183	123264	OSTbeta	OSTB_HUMAN	Q86UW2	tm	basolateral plasma membrane	128	1
69602	Otop3	OTOP3_MOUSE	Q80UF9	Otopetrin-3.	2	2	1079	347741	OTOP3	OTOP3_HUMAN	Q7RTS5	tm	integral to membrane	577	12
18438	P2rx4	P2RX4_MOUSE	Q9JJX6	P2X purinoceptor 4 (P2X4) (ATP receptor) (Purinergic receptor).	8	8	499	5025	P2RX4	P2RX4_HUMAN	Q99571	tm	integral to plasma membrane	388	2
70584	Pak4	PAK4_MOUSE	Q8BTW9	Serine/threonine-protein kinase PAK 4 (EC 2.7.11.1) (p21-activated kinase 4) (PAK-4).	2	2	1081	10298	PAK4	PAK4_HUMAN	O96013	tm	Golgi apparatus	593	0
268663	Pcdh24	Q3UN77_MOUSE	Q3UN77	protocadherin 24	28	241	19	54825	PCDH24	PCD24_HUMAN	Q9BYE9	tm	apical plasma membrane	1308	1
71853	Pdia6	PDIA6_MOUSE	Q922R8	Protein disulfide-isomerase A6 precursor (EC 5.3.4.1) (Thioredoxin domain-containing protein 7).	2	2	1126	10130	PDIA6	PDIA6_HUMAN	Q15084	tm	endoplasmic reticulum lumen	440	0
224020	Pi4ka	Q6DIC7_MOUSE	Q6DIC7	phosphatidylinositol 4-kinase, catalytic, alpha polypeptide	6	6	607	5297	PI4KA	PI4KA_HUMAN	P42356	tm	Golgi-associated vesicle	2044	0
18719	Pip5k1b	PI51B_MOUSE	P70181	Phosphatidylinositol-4-phosphate 5-kinase type-1 beta (EC 2.7.1.68)	2	2	1172	8395	PIP5K1B	PI51B_HUMAN	O14986	tm	associated to membrane	539	0
56460	Pkp3	PKP3_MOUSE	Q9QY23	Plakophilin-3.	3	4	785	11187	PKP3	PKP3_HUMAN	Q9Y446	tm	cytoplasm	797	0
665270	Plb1	PLB1_MOUSE	Q3TTY0	Phospholipase B1, membrane-associated precursor (Phospholipase B) (Phospholipase B/lipase) (PLB/LIP).	5	15	291	151056	PLB1	PLB1_HUMAN	Q6P1J6	tm	apical plasma membrane	1478	1
72287	Plekhf1	PKHF1_MOUSE	Q3TB82	Pleckstrin homology domain-containing family F member 1 (PH domain- containing family F member 1)	2	2	1082	79156	PLEKHF1	PKHF1_HUMAN	Q96S99	tm	cytoplasm	279	0
22038	Plscr1	PLS1_MOUSE	Q9JJ00	Phospholipid scramblase 1 (PL scramblase 1)	3	12	356	5359	PLSCR1	PLS1_HUMAN	O15162	tm	integral to plasma membrane	328	0
19012	Ppap2a	LPP1_MOUSE	Q61469	Lipid phosphate phosphohydrolase 1 (EC 3.1.3.4) (Phosphatidic acid phosphatase 2a)	5	6	500	8611	PPAP2A	LPP1_HUMAN	O14494	tm	integral to plasma membrane	283	6
50784	Ppap2c	LPP2_MOUSE	Q9DAX2	Lipid phosphate phosphohydrolase 2 (EC 3.1.3.4) (Phosphatidic acid phosphatase 2c) (PAP-2c)	4	6	590	8612	PPAP2C	LPP2_HUMAN	O43688	tm	integral to membrane	276	6
106564	Ppcs	PPCS_MOUSE	Q8VDG5	phosphopantothenoylcysteine synthetase	7	5	683	79717	PPCS	PPCS_HUMAN	Q9HAB8	tm	-	311	0
19035	Ppib	PPIB_MOUSE	P24369	Peptidyl-prolyl cis-trans isomerase B precursor (EC 5.2.1.8) (PPIase) (Rotamase)	2	1	1198	5479	PPIB	PPIB_HUMAN	P23284	tm	integral to endoplasmic reticulum membrane	208	1
110854	Ppp2r4	PTPA_MOUSE	P58389	Serine/threonine-protein phosphatase 2A regulatory subunit B'	5	8	514	5524	PPP2R4	PTPA_HUMAN	Q15257	tm	cytoplasm	323	0
108099	Prkag2	AAKG2_MOUSE	Q91WG5	5'-AMP-activated protein kinase subunit gamma-2 (AMPK gamma-2 chain) (AMPK gamma2).	5	6	592	51422	PRKAG2	AAKG2_HUMAN	Q9UGJ0	tm	cytoplasm	566	0
19146	Prss7	ENTK_MOUSE	P97435	Enteropeptidase (EC 3.4.21.9) (Enterokinase) (Serine protease 7)	4	11	379	5651	PRSS7	ENTK_HUMAN	P98073	tm	apical plasma membrane	1069	1
19164	Psen1	PSN1_MOUSE	P49769	Presenilin-1 (EC 3.4.23.-) (PS-1) (Protein S182)	2	5	652	5663	PSEN1	PSN1_HUMAN	P49768	tm	integral to plasma membrane	467	9
17463	Psmd7	PSD7_MOUSE	P26516	26S proteasome non-ATPase regulatory subunit 7 (26S proteasome regulatory subunit rpn8)	6	7	544	5713	PSMD7	PSD7_HUMAN	P51665	tm	cytosol	321	0
19243	Ptp4a1	TP4A1_MOUSE	Q63739	Protein tyrosine phosphatase type IVA 1 (EC 3.1.3.48)	5	4	707	7803	PTP4A1	TP4A1_HUMAN	Q93096	tm	anchored to plasma membrane	173	0
19264	Ptprc	CD45_MOUSE	P06800	Leukocyte common antigen precursor (EC 3.1.3.48) (L-CA) (Lymphocyte antigen 5) (Ly-5) (T200) (CD45 antigen)	2	3	945	5788	PTPRC	CD45_HUMAN	P08575	tm	integral to plasma membrane	1291	2
108705	Pttg1ip	PTTG_MOUSE	Q8R143	Pituitary tumor-transforming gene 1 protein-interacting protein precursor	2	4	752	754	PTTG1IP	PTTG_HUMAN	P53801	tm	integral to nuclear membrane	174	1
19330	Rab18	RAB18_MOUSE	P35293	Ras-related protein Rab-18.	7	11	401	22931	RAB18	RAB18_HUMAN	Q9NP72	tm	anchored to plasma membrane	206	0
235442	Rab8b	RAB8B_MOUSE	P61028	Ras-related protein Rab-8B.	10	18	251	51762	RAB8B	RAB8B_HUMAN	Q92930	tm	anchored to plasma membrane	207	0
19353	Rac1	RAC1_MOUSE	P63001	Ras-related C3 botulinum toxin substrate 1 precursor (p21-Rac1).	9	27	187	5879	RAC1	RAC1_HUMAN	P63000	tm	anchored to plasma membrane	192	0
13476	Reep5	REEP5_MOUSE	Q60870	Receptor expression-enhancing protein 5 (Polyposis locus protein 1 homolog) (TB2 protein homolog) (GP106).	2	2	1170	7905	REEP5	REEP5_HUMAN	Q00765	tm	integral to membrane	185	2
70335	Reep6	REEP6_MOUSE	Q9JM62	Receptor expression-enhancing protein 6 (Polyposis locus protein 1-like 1) (TB2 protein-like 1)	2	6	606	92840	REEP6	REEP6_HUMAN	Q96HR9	tm	integral to membrane	201	3
69581	Rhou	RHOU_MOUSE	Q9EQT3	Rho-related GTP-binding protein RhoU (Wnt-1 responsive Cdc42 homolog 1) (WRCH-1) (Rho GTPase-like protein ARHU).	3	2	1078	58480	RHOU	RHOU_HUMAN	Q7L0Q8	tm	anchored to plasma membrane	261	0
103963	Rpn1	RPN1_MOUSE	Q91YQ5	Dolichyl-diphosphooligosaccharide—protein glycosyltransferase subunit 1 precursor (EC 2.4.1.119)	4	3	834	6184	RPN1	RPN1_HUMAN	P04843	tm	integral to endoplasmic reticulum membrane	608	1
20014	Rpn2	RPN2_MOUSE	Q9DBG6	Dolichyl-diphosphooligosaccharide—protein glycosyltransferase subunit 2 precursor (EC 2.4.1.119)	3	2	1068	6185	RPN2	RPN2_HUMAN	P04844	tm	integral to endoplasmic reticulum membrane	631	3
20130	Rras	RRAS_MOUSE	P10833	Ras-related protein R-Ras precursor (p23).	6	13	336	6237	RRAS	RRAS_HUMAN	P10301	tm	anchored to plasma membrane	218	0
20168	Rtn3	RTN3_MOUSE	Q9ES97	Reticulon-3.	2	4	782	10313	RTN3	RTN3_HUMAN	O95197	tm	integral to endoplasmic reticulum membrane	964	3
83493	Sacm1l	SAC1_MOUSE	Q9EP69	Phosphatidylinositide phosphatase SAC1 (EC 3.1.3.-) (Suppressor of actin mutations 1-like protein).	2	2	1180	22908	SACM1L	SAC1_HUMAN	Q9NTJ5	tm	integral to endoplasmic reticulum membrane	587	2
107767	Scamp1	SCAM1_MOUSE	Q8K021	Secretory carrier-associated membrane protein 1 (Secretory carrier membrane protein 1).	2	2	1181	9522	SCAMP1	SCAM1_HUMAN	O15126	tm	integral to membrane	338	4
12492	Scarb2	SCRB2_MOUSE	O35114	Lysosome membrane protein 2 (Lysosome membrane protein II) (LIMP II) (Scavenger receptor class B member 2)	3	2	974	950	SCARB2	SCRB2_HUMAN	Q14108	tm	integral to lysosomal membrane	478	2
20333	Sec22b	SC22B_MOUSE	O08547	Vesicle-trafficking protein SEC22b (SEC22 vesicle-trafficking protein homolog B)	4	3	830	9554	SEC22B	SC22B_HUMAN	O75396	tm	integral to endoplasmic reticulum membrane	215	1
27054	Sec23b	SC23B_MOUSE	Q9D662	Protein transport protein Sec23B (SEC23-related protein B).	11	12	363	10483	SEC23B	SC23B_HUMAN	Q15437	tm	COPII vesicle coat	767	0
77371	Sec24a	SC24A_MOUSE	Q3U2P1	Protein transport protein Sec24A (SEC24-related protein A).	10	11	382	10802	SEC24A	SC24A_HUMAN	O95486	tm	COPII vesicle coat	1090	0
20703	Serpina1d	A1AT1_MOUSE	Q00897	Alpha-1-antitrypsin 1-4 precursor (Serine protease inhibitor 1-4) (Alpha-1 protease inhibitor 4).	2	2	1071	5265	SERPINA1	A1AT_HUMAN	P01009	tm	extracellular region	413	0
66222	Serpinb1a	ILEUA_MOUSE	Q9D154	Leukocyte elastase inhibitor A (Serine protease inhibitor EIA) (Serpin B1a).	15	25	202	1992	SERPINB1	ILEU_HUMAN	P30740	tm	cytoplasm	379	0
14057	Sfxn1	SFXN1_MOUSE	Q99JR1	Sideroflexin-1.	3	3	944	94081	SFXN1	SFXN1_HUMAN	Q9H9B4	tm	integral to mitochondrial inner membrane	322	4
20397	Sgpl1	SGPL1_MOUSE	Q8R0X7	Sphingosine-1-phosphate lyase 1 (EC 4.1.1.27) (SP-lyase) (mSPL) (Sphingosine-1-phosphate aldolase).	2	3	946	8879	SGPL1	SGPL1_HUMAN	O95470	tm	integral to endoplasmic reticulum membrane	568	1
20494	Slc10a2	NTCP2_MOUSE	P70172	Ileal sodium/bile acid cotransporter (Ileal Na(+)/bile acid cotransporter) (Na(+)-dependent ileal bile acid transporter)	5	11	389	6555	SLC10A2	NTCP2_HUMAN	Q12908	tm	apical plasma membrane	348	8
20496	Slc12a2	S12A2_MOUSE	P55012	Solute carrier family 12 member 2 (Bumetanide-sensitive sodium-(potassium)-chloride cotransporter 1) (Basolateral Na-K-Cl symporter)	4	4	746	6558	SLC12A2	S12A2_HUMAN	P55011	tm	integral to plasma membrane	1205	12
20500	Slc13a2	S13A2_MOUSE	Q9ES88	Solute carrier family 13 member 2 (Renal sodium/dicarboxylate cotransporter) (Na(+)/dicarboxylate cotransporter 1) (NaDC-1).	10	29	168	9058	SLC13A2	S13A2_HUMAN	Q13183	tm	integral to plasma membrane	586	13

Gene ID (mouse)	Gene Symbol (mouse)	Swissprot protein ID (mouse)	Swissprot protein AC (mouse)	Gene Description	Max Diff peptide	Avg Pep count	Rank order	Ortho Gene ID	Ortho Gene Symbol	Swissprot protein ID	Swissprot protein AC	Topology	GO term	Protein length	Number of TM
56643	Slc15a1	S15A1_MOUSE	Q9JIP7	Solute carrier family 15 member 1 (Peptide transporter 1) (Oligopeptide transporter, small intestine isoform) (Intestinal H(+)/peptide cotransporter) (Proton-coupled dipeptide cotransporter).	20	92	58	6564	SLC15A1	S15A1_HUMAN	P46059	tm	integral to plasma membrane	709	9
20501	Slc16a1	MOT1_MOUSE	P53986	Monocarboxylate transporter 1 (MCT 1) (Solute carrier family 16 member 1).	3	8	501	6566	SLC16A1	MOT1_HUMAN	P53985	tm	integral to plasma membrane	493	11
20510	Slc1a1	EAA3_MOUSE	P51906	Excitatory amino acid transporter 3 (Sodium-dependent glutamate/aspartate transporter 3) (Excitatory amino-acid carrier 1) (Solute carrier family 1 member 1).	2	2	1069	6505	SLC1A1	EAA3_HUMAN	P43005	tm	integral to plasma membrane	523	9
20514	Slc1a5	AAAT_MOUSE	P51912	Neutral amino acid transporter B(0) (ATB(0)) (Sodium-dependent neutral amino acid transporter type 2)	2	1	1199	6510	SLC1A5	AAAT_HUMAN	Q15758	tm	integral to plasma membrane	553	9
20520	Slc22a5	S22A5_MOUSE	Q9Z0E8	Solute carrier family 22 member 5 (Organic cation/carnitine transporter 2) (High-affinity sodium-dependent carnitine cotransporter).	5	5	681	6584	SLC22A5	S22A5_HUMAN	O76082	tm	integral to plasma membrane	557	11
20522	Slc23a1	S23A1_MOUSE	Q9Z2J0	Solute carrier family 23 member 1 (Sodium-dependent vitamin C transporter 1) (Na(+)/L-ascorbic acid transporter 1) (Yolk sac permease-like molecule 3).	5	4	747	9963	SLC23A1	S23A1_HUMAN	Q9UHI7	tm	integral to plasma membrane	605	11
13358	Slc25a1	Q3TDH6_MOUSE	Q3TDH6	Solute carrier family 25 (mitochondrial carrier, citrate transporter), member 1	3	4	781	6576	SLC25A1	TXTP_HUMAN	P53007	tm	integral to mitochondrial inner membrane	311	0
27376	Slc25a10	DIC_MOUSE	Q9QZD8	Mitochondrial dicarboxylate carrier (Solute carrier family 25 member 10).	2	2	1073	1468	SLC25A10	DIC_HUMAN	Q9UBX3	tm	integral to mitochondrial outer membrane	287	0
18674	Slc25a3	MPCP_MOUSE	Q8VEM8	Phosphate carrier protein, mitochondrial precursor (Phosphate transport protein) (PTP) (Solute carrier family 25 member 3).	6	6	604	5250	SLC25A3	MPCP_HUMAN	Q00325	tm	integral to plasma membrane	357	2
13521	Slc26a2	S26A2_MOUSE	Q62273	Sulfate transporter (Diastrophic dysplasia protein homolog) (Solute carrier family 26 member 2) (ST-OB).	2	5	680	1836	SLC26A2	S26A2_HUMAN	P50443	tm	integral to plasma membrane	739	9
13487	Slc26a3	S26A3_MOUSE	Q9WVC8	Chloride anion exchanger (Down-regulated in adenoma) (Protein DRA) (Solute carrier family 26 member 3).	2	5	679	1811	SLC26A3	S26A3_HUMAN	P40879	tm	integral to plasma membrane	757	11
171429	Slc26a6	Q812E2_MOUSE	Q812E2	solute carrier family 26, member 6	14	54	93	65010	SLC26A6	S26A6_HUMAN	Q9BXS9	tm	integral to membrane	735	9
26458	Slc27a2	S27A2_MOUSE	O35488	Very long-chain acyl-CoA synthetase (EC 6.2.1.-) (VLACS) (VLCS) (EC 6.2.1.3) (Fatty acid transport protein 2)	8	7	545	11001	SLC27A2	S27A2_HUMAN	O14975	tm	integral to endoplasmic reticulum membrane	620	2
26569	Slc27a4	S27A4_MOUSE	Q91VE0	Long-chain fatty acid transport protein 4 (EC 6.2.1.-) (Fatty acid transport protein 4) (FATP-4)	3	3	947	10999	SLC27A4	S27A4_HUMAN	Q6P1M0	tm	integral to plasma membrane	643	2
434203	Slc28a1	Q6P8I9_MOUSE	Q6P8I9	solute carrier family 28 (sodium-coupled nucleoside transporter), member 1	5	6	608	9154	SLC28A1	S28A1_HUMAN	O00337	tm	integral to plasma membrane	648	10
269346	Slc28a2	S28A2_MOUSE	O88627	Sodium/nucleoside cotransporter 2 (Na(+)/nucleoside cotransporter 2) (Sodium-coupled nucleoside transporter 2)	6	9	447	9153	SLC28A2	S28A2_HUMAN	O43868	tm	integral to plasma membrane	660	11
56485	Slc2a5	GTR5_MOUSE	Q9WV38	Solute carrier family 2, facilitated glucose transporter member 5 (Glucose transporter type 5, small intestine) (GLUT-5) (Fructose transporter).	5	13	337	6518	SLC2A5	GTR5_HUMAN	P22732	tm	integral to plasma membrane	501	12
435818	Slc2a7	GTR7_MOUSE	P0C6A1	Solute carrier family 2, facilitated glucose transporter member 7 (Glucose transporter type 7) (GLUT-7).	3	6	564	155184	SLC2A7	GTR7_HUMAN	Q6PXP3	tm	integral to membrane	513	10
117591	Slc2a9	Q99JJ2_MOUSE	Q99JJ2	solute carrier family 2 (facilitated glucose transporter), member 9	4	12	367	56606	SLC2A9	GTR9_HUMAN	Q9NRM0	tm	integral to plasma membrane	416	10
226781	Slc30a10	ZNT10_MOUSE	Q3UVU3	Zinc transporter 10 (ZnT-10) (Solute carrier family 30 member 10).	7	11	384	55532	SLC30A10	ZNT10_HUMAN	Q6XR72	tm	integral to membrane	470	6
20529	Slc31a1	COPT1_MOUSE	Q8K211	High affinity copper uptake protein 1 (Copper transporter 1) (CTR1) (Solute carrier family 31 member 1).	2	2	1070	1317	SLC31A1	COPT1_HUMAN	O15431	tm	integral to plasma membrane	196	3
20531	Slc34a2	NPT2B_MOUSE	Q9DBP0	Sodium-dependent phosphate transport protein 2B (Sodium-phosphate transport protein 2B).	11	35	146	10568	SLC34A2	NPT2B_HUMAN	O95436	tm	apical plasma membrane	697	9
72022	Slc35f2	S35F2_MOUSE	Q7TML3	Solute carrier family 35 member F2.	3	11	381	54733	SLC35F2	S35F2_HUMAN	Q8IXU6	tm	integral to membrane	375	9
215335	Slc36a1	S36A1_MOUSE	Q8K4D3	Proton-coupled amino acid transporter 1 (Proton/amino acid transporter 1) (Solute carrier family 36 member 1).	2	6	593	206358	SLC36A1	S36A1_HUMAN	Q7Z2H8	tm	integral to plasma membrane	475	11
56857	Slc37a2	SPX2_MOUSE	Q9WU81	Sugar phosphate exchanger 2 (cAMP-inducible protein 2) (Solute carrier family 37 member 2).	3	2	1074	219855	SLC37A2	SPX2_HUMAN	Q8TED4	tm	integral to membrane	501	12
72027	Slc39a4	S39A4_MOUSE	Q78IQ7	Zinc transporter ZIP4 precursor (Zrt- and Irt-like protein 4) (ZIP-4) (Solute carrier family 39 member 4) (Activated in W/Wv mouse stomach 2) (mAWMS2).	4	13	329	55630	SLC39A4	S39A4_HUMAN	Q6P5W5	tm	integral to plasma membrane	660	6
20532	Slc3a1	O55093_MOUSE	O55093	solute carrier family 3, member 1	23	62	79	6519	SLC3A1	SLC31_HUMAN	Q07837	tm	integral to plasma membrane	685	1
17254	Slc3a2	4F2_MOUSE	P10852	4F2 cell-surface antigen heavy chain (4F2hc).	17	27	186	6520	SLC3A2	4F2_HUMAN	P08195	tm	integral to plasma membrane	526	1
215113	Slc43a2	LAT4_MOUSE	Q8CGA3	Large neutral amino acids transporte	2	4	787	124935	SLC43A2	LAT4_HUMAN	Q8N370	tm	integral to membrane	568	12
100434	Slc44a1	CTL1_MOUSE	Q6X893	Choline transporter-like protein 1 (Solute carrier family 44 member 1) (CD92 antigen).	2	1	1205	23446	SLC44A1	CTL1_HUMAN	Q8WWI5	tm	integral to membrane	653	9
70129	Slc44a4	CTL4_MOUSE	Q91VA1	Choline transporter-like protein 4 (Solute carrier family 44 member 4).	6	11	380	80736	SLC44A4	CTL4_HUMAN	Q53GD3	tm	integral to membrane	707	10
52466	Slc46a1	PCFT_MOUSE	Q6PEM8	Proton-coupled folate transporter (Heme carrier protein 1) (PCFT/HCP1) (Solute carrier family 46 member 1).	3	14	309	113235	SLC46A1	PCFT_HUMAN	Q96NT5	tm	apical plasma membrane	459	10
20533	Slc4a1	B3AT_MOUSE	P04919	Band 3 anion transport protein (Anion exchange protein 1) (AE 1) (Solute carrier family 4 member 1) (MEB3) (CD233 antigen).	3	3	884	6521	SLC4A1	B3AT_HUMAN	P02730	tm	integral to plasma membrane	929	12
20537	Slc5a1	SC5A1_MOUSE	Q8C3K6	Sodium/glucose cotransporter 1 (Na(+)/glucose cotransporter 1) (High affinity sodium-glucose cotransporter) (Solute carrier family 5 member 1).	23	587	9	6523	SLC5A1	SC5A1_HUMAN	P13866	tm	apical plasma membrane	665	14
233836	Slc5a11	SC5AB_MOUSE	Q49B93	Sodium-coupled monocarboxylate transporter 2 (Electroneutral sodium monocarboxylate cotransporter)	7	31	162	159963	SLC5A11	SC5AB_HUMAN	Q8WWX8	tm	integral to membrane	673	14
241612	Slc5a12	SC5AC_MOUSE	Q9ET37	Low affinity sodium-glucose cotransporter (Sodium/glucose cotransporter 3) (Na(+)/glucose cotransporter 3) (Solute carrier family 5 member 4).	6	12	368	6527	SLC5A12	SC5AC_HUMAN	Q1EHB4	tm	apical plasma membrane	619	13
64452	Slc5a4a	SC5A4_MOUSE	Q5U4D8	Sodium-dependent multivitamin transporter (Na(+)-dependent multivitamin transporter) (Solute carrier family 5 member 6).	8	20	241	8884	SLC5A4	SC5A4_HUMAN	Q9NY91	tm	integral to membrane	656	14
64454	Slc5a4b	A0PJT9_MOUSE	A0PJT9	solute carrier family 5 (neutral amino acid transporters, system A), member 4b	12	61	80	-	-	-	-	tm		659	14
330064	Slc5a6	SC5A6_MOUSE	Q8BYF6	Sodium-coupled monocarboxylate transporter 1 (Electrogenic sodium monocarboxylate cotransporter) (Solute carrier family 5 member 8).	2	4	754	160728	SLC5A6	SC5A6_HUMAN	Q9Y289	tm	integral to plasma membrane	634	13
216225	Slc5a8	SC5A8_MOUSE	Q8VDT1	Sodium/glucose cotransporter 4 (Na(+)/glucose cotransporter 4) (mSGLT4) (Solute carrier family 5 member 9).	5	13	340	200010	SLC5A8	SC5A8_HUMAN	Q8N695	tm	apical plasma membrane	613	13
230612	Slc5a9	SC5A9_MOUSE	Q8K0E3	Sodium/myo-inositol cotransporter 2 (Na+/myo-inositol cotransporter 2) (Sodium/glucose cotransporter KST1)	7	12	350	115584	SLC5A9	SC5A9_HUMAN	Q2M3M2	tm	integral to membrane	685	14
56774	Slc6a14	S6A14_MOUSE	Q9JMA9	Sodium- and chloride-dependent neutral and basic amino acid transporter B(0+)	3	2	976	11254	SLC6A14	S6A14_HUMAN	Q9UN76	tm	integral to plasma membrane	638	12
74338	Slc6a19	S6A19_MOUSE	Q9D687	Sodium-dependent neutral amino acid transporter B(0) .	9	46	112	340024	SLC6A19	S6A19_HUMAN	Q695T7	tm	integral to plasma membrane	634	12
102680	Slc6a20a	S620A_MOUSE	Q8VDB9	solute carrier family 6 (neurotransmitter transporter), member 20A	5	12	366	54716	SLC6A20	S6A20_HUMAN	Q9NP91	tm	integral to plasma membrane	635	12
102857	Slc6a8	SC6A8_MOUSE	Q8VBW1	Sodium- and chloride-dependent creatine transporter 1 (Creatine transporter 1) (CT1) (Solute carrier family 6 member 8).	3	2	979	6535	SLC6A8	SC6A8_HUMAN	P48029	tm	integral to plasma membrane	640	12
328059	Slc7a15	Q50E62_MOUSE	Q50E62	solute carrier family 7 (cationic amino acid transporter, y+ system), member 15	2	2	1085	-	-	-	-	tm		488	12

Gene ID (mouse)	Gene Symbol (mouse)	Swissprot protein ID (mouse)	Swissprot protein AC (mouse)	Gene Description	Max Diff peptide	Avg Pep count	Rank order	Ortho Gene ID	Ortho Gene Symbol	Swissprot protein ID	Swissprot protein AC	Topology	GO term	Protein length	Number of TM
50934	Slc7a8	LAT2_MOUSE	Q9QXW9	Large neutral amino acids transporter small subunit 2 (L-type amino acid transporter 2) (mLAT2)	5	7	546	23428	SLC7A8	LAT2_HUMAN	Q9UHI5	tm	integral to plasma membrane	531	12
30962	Slc7a9	BAT1_MOUSE	Q9QXA6	B(0,+)-type amino acid transporter 1 (B(0,+)AT) (Glycoprotein-associated amino acid transporter b0,+AT1)	3	6	563	11136	SLC7A9	BAT1_HUMAN	P82251	tm	integral to plasma membrane	487	12
105243	Slc9a3	Q8BZU0_MOUSE	Q8BZU0	solute carrier family 9 (sodium/hydrogen exchanger), member 3	16	51	99	6550	SLC9A3	SL9A3_HUMAN	P48764	tm	apical plasma membrane	196	0
58994	Smpd3	NSMA2_MOUSE	Q9JJY3	Sphingomyelin phosphodiesterase 3 (EC 3.1.4.12) (Neutral sphingomyelinase II)	4	10	408	55512	SMPD3	NSMA2_HUMAN	Q9NY59	tm	integral to plasma membrane	655	2
100340	Smpdl3b	ASM3B_MOUSE	P58242	Acid sphingomyelinase-like phosphodiesterase 3b precursor (EC 3.1.4.-) (ASM-like phosphodiesterase 3b).	10	21	230	27293	SMPDL3B	ASM3B_HUMAN	Q92485	tm	extracellular region	456	0
20655	Sod1	SODC_MOUSE	P08228	Superoxide dismutase [Cu-Zn] (EC 1.15.1.1).	4	4	784	6647	SOD1	SODC_HUMAN	P00441	tm	cytoplasm	154	0
20322	Sord	DHSO_MOUSE	Q64442	sorbitol dehydrogenase	7	4	783	6652	SORD	DHSO_HUMAN	Q00796	tm	-	357	0
66624	Spcs2	SPCS2_MOUSE	Q9CYN2	Signal peptidase complex subunit 2 (EC 3.4.-.-) (Microsomal signal peptidase 25 kDa subunit) (SPase 25 kDa subunit).	2	2	1124	9789	SPCS2	SPCS2_HUMAN	Q15005	tm	integral to endoplasmic reticulum membrane	226	2
76650	Srxn1	SRXN1_MOUSE	Q9D975	Sulfiredoxin-1 (EC 1.8.98.2) (Neoplastic progression protein 3).	2	1	1202	140809	SRXN1	SRXN1_HUMAN	Q9BYN0	tm	cytoplasm	136	0
20832	Ssr4	SSRD_MOUSE	Q62186	Translocon-associated protein subunit delta precursor (TRAP-delta) (Signal sequence receptor subunit delta) (SSR-delta).	2	2	1072	6748	SSR4	SSRD_HUMAN	P51571	tm	integral to endoplasmic reticulum membrane	172	1
68991	Ssu72	SSU72_MOUSE	Q9CY97	RNA polymerase II subunit A C-terminal domain phosphatase SSU72 (EC 3.1.3.16) (CTD phosphatase SSU72).	3	2	1077	29101	SSU72	SSU72_HUMAN	Q9NP77	tm	cytoplasm	194	0
20443	St3gal4	SIA4C_MOUSE	Q91Y74	CMP-N-acetylneuraminate-beta-galactosamide-alpha-2,3-sialyltransferase (EC 2.4.99.-)	4	3	831	6484	ST3GAL4	SIA4C_HUMAN	Q11206	tm	integral to Golgi membrane	333	1
76205	Stard3nl	MENTO_MOUSE	Q9DCI3	MLN64 N-terminal domain homolog (STARD3 N-terminal-like protein).	2	2	1179	83930	STARD3NL	MENTO_HUMAN	O95772	tm	integral to late endosome membrane	235	4
13830	Stom	STOM_MOUSE	P54116	Erythrocyte band 7 integral membrane protein (Stomatin) (Protein 7.2b).	11	66	74	2040	STOM	STOM_HUMAN	P27105	tm	integral to plasma membrane	284	1
100226	Stx12	STX12_MOUSE	Q9ER00	Syntaxin-12.	6	7	548	23673	STX12	STX12_HUMAN	Q86Y82	tm	integral to Golgi membrane	274	1
20908	Stx3	STX3_MOUSE	Q64704	Syntaxin-3.	12	37	133	6809	STX3	STX3_HUMAN	Q13277	tm	integral to plasma membrane	289	1
20909	Stx4a	STX4_MOUSE	P70452	Syntaxin-4.	5	4	709	6810	STX4	STX4_HUMAN	Q12846	tm	basolateral plasma membrane	298	1
53331	Stx7	STX7_MOUSE	O70439	Syntaxin-7.	4	4	751	8417	STX7	STX7_HUMAN	O15400	tm	integral to plasma membrane	261	1
243659	Styk1	STYK1_MOUSE	Q6J9G1	Tyrosine protein-kinase STYK1 (EC 2.7.10.2) (Serine/threonine/tyrosine kinase 1) (Novel oncogene with kinase domain) (mNOK).	10	24	214	55359	STYK1	STYK1_HUMAN	Q6J9G0	tm	integral to plasma membrane	429	1
71733	Susd2	SUSD2_MOUSE	Q9DBX3	Sushi domain-containing protein 2 precursor.	16	42	123	56241	SUSD2	SUSD2_HUMAN	Q9UGT4	tm	integral to membrane	820	1
19027	Sypl	SYPL1_MOUSE	O09117	Synaptophysin-like protein 1 (Pantophysin).	3	8	511	6856	SYPL1	SYPL1_HUMAN	Q16563	tm	integral to plasma membrane	261	3
17075	Tacstd1	Q61512_MOUSE	Q61512	tumor-associated calcium signal transducer 1	9	55	91	4072	TACSTD1	TACD1_HUMAN	P16422	tm	integral to plasma membrane	314	1
21817	Tgm2	TGM2_MOUSE	P21981	Protein-glutamine gamma-glutamyltransferase 2 (EC 2.3.2.13)	5	4	748	7052	TGM2	TGM2_HUMAN	P21980	tm	plasma membrane	686	0
21838	Thy1	THY1_MOUSE	P01831	Thy-1 membrane glycoprotein precursor (Thy-1 antigen) (CD90 antigen).	3	6	588	7070	THY1	THY1_HUMAN	P04216	tm	anchored to plasma membrane	162	0
142980	Tlr3	TLR3_MOUSE	Q99MB1	Toll-like receptor 3 precursor (CD283 antigen).	3	2	1084	7098	TLR3	TLR3_HUMAN	O15455	tm	integral to plasma membrane	905	1
68059	Tm9sf2	TM9S2_MOUSE	P58021	Transmembrane 9 superfamily member 2 precursor.	4	7	532	9375	TM9SF2	TM9S2_HUMAN	Q99805	tm	integral to membrane	662	9
107358	Tm9sf3	TM9S3_MOUSE	Q9ET30	Transmembrane 9 superfamily member 3 precursor.	2	4	805	56889	TM9SF3	TM9S3_HUMAN	Q9HD45	tm	integral to membrane	587	9
353499	Tmc4	TMC4_MOUSE	Q7TQ65	Transmembrane channel-like protein 4.	9	13	330	147798	TMC4	TMC4_HUMAN	Q7Z404	tm	integral to membrane	694	8
74424	Tmc5	TMC5_MOUSE	Q32NZ6	Transmembrane channel-like protein 5.	6	12	365	79838	TMC5	TMC5_HUMAN	Q6UXY8	tm	integral to membrane	967	9
68581	Tmed10	TMEDA_MOUSE	Q9D1D4	Transmembrane emp24 domain-containing protein 10 precursor (21 kDa transmembrane-trafficking protein)	2	3	926	10972	TMED10	TMEDA_HUMAN	P49755	tm	integral to Golgi membrane	219	2
103694	Tmed4	TMED4_MOUSE	Q8R1V4	Transmembrane emp24 domain-containing protein 4 precursor (p26).	3	5	657	222068	TMED4	TMED4_HUMAN	Q7Z7H5	tm	integral to endoplasmic reticulum membrane	227	2
66676	Tmed7	-	-	transmembrane emp24 protein transport domain containing 7	4	4	710	51014	TMED7	TMED7_HUMAN	Q9Y3B3	tm	integral to endoplasmic reticulum membrane	224	1
67511	Tmed9	TMED9_MOUSE	Q99KF1	Transmembrane emp24 domain-containing protein 9 precursor (Glycoprotein 25L2).	3	5	682	54732	TMED9	TMED9_HUMAN	Q9BVK6	tm	integral to endoplasmic reticulum membrane	214	0
105722	Tmem16f	TM16F_MOUSE	Q6P9J9	Transmembrane protein 16F.	18	43	118	196527	TMEM16F	TM16F_HUMAN	Q4KMQ2	tm	integral to membrane	911	8
407243	Tmem189	TM189_MOUSE	Q99LQ7	Transmembrane protein 189.	3	2	1086	387521	TMEM189	TM189_HUMAN	A5PLL7	tm	integral to endoplasmic reticulum membrane	271	2
73067	Tmem192	TM192_MOUSE	Q9CXT7	Transmembrane protein 192.	2	2	1178	201931	TMEM192	TM192_HUMAN	Q8IY95	tm	integral to membrane	266	4
69981	Tmem30a	CC50A_MOUSE	Q8VEK0	Cell cycle control protein 50A (Transmembrane protein 30A).	3	3	833	55754	TMEM30A	CC50A_HUMAN	Q9NV96	tm	integral to membrane	364	2
238257	Tmem30b	CC50B_MOUSE	Q8BHG3	Cell cycle control protein 50B (Transmembrane protein 30B).	5	17	273	161291	TMEM30B	CC50B_HUMAN	Q3MIR4	tm	integral to membrane	353	2
67878	Tmem33	TMM33_MOUSE	Q9CR67	Transmembrane protein 33 (DB83 protein).	3	2	1076	55161	TMEM33	TMM33_HUMAN	P57088	tm	integral to membrane	247	3
77975	Tmem50b	TM50B_MOUSE	Q9D1X9	Transmembrane protein 50B.	2	1	1203	757	TMEM50B	TM50B_HUMAN	P56557	tm	integral to plasma membrane	158	4
72519	Tmem55a	TM55A_MOUSE	Q9CZX7	Transmembrane protein 55A (EC 3.1.3.-) (Type II phosphatidylinositol 4,5-bisphosphate 4-phosphatase)	2	2	1177	55529	TMEM55A	TM55A_HUMAN	Q8N4L2	tm	integral to late endosome membrane	257	2
219024	Tmem55b	TM55B_MOUSE	Q3TWL2	Transmembrane protein 55B (EC 3.1.3.-) (Type I phosphatidylinositol 4,5-bisphosphate 4-phosphatase) (PtdIns-4,5-P2 4-Ptase I).	3	2	980	90809	TMEM55B	TM55B_HUMAN	Q86T03	tm	integral to late endosome membrane	284	2
66601	Tmigd1	TMIG1_MOUSE	Q9D7L8	Transmembrane and immunoglobulin domain-containing protein 1 precursor.	4	8	512	388364	TMIGD1	TMIG1_HUMAN	Q6UXZ0	tm	integral to membrane	261	1
50528	Tmprss2	TMPS2_MOUSE	Q9JIQ8	Transmembrane protease, serine 2 (EC 3.4.21.-) (Epitheliasin) (Plasmic transmembrane protein X)	3	3	885	7113	TMPRSS2	TMPS2_HUMAN	O15393	tm	integral to membrane	490	1
214523	Tmprss4	TMPS4_MOUSE	Q8VCA5	Transmembrane protease, serine 4 (EC 3.4.21.-) (Channel-activating protease 2) (mCAP2).	2	4	786	56649	TMPRSS4	TMPS4_HUMAN	Q9NRS4	tm	integral to membrane	435	1
59069	Tpm3	TPM3_MOUSE	P21107	Tropomyosin alpha-3 chain (Tropomyosin-3) (Tropomyosin gamma).	Tropomyosin alpha-3 chain (Tropomyosin-3) (Tropomyosin gamma).	5	7	547	7170	TPM3	TPM3_HUMAN	P06753	tm	cytoskeleton	284	0
58866	Treh	TREA_MOUSE	Q9JLT2	Trehalase precursor (EC 3.2.1.28) (Alpha,alpha-trehalase) (Alpha,alpha-trehalose glucohydrolase).	26	103	49	11181	TREH	TREA_HUMAN	O43280	tm	anchored to plasma membrane	576	0
68667	Trpm4	TRPM4_MOUSE	Q7TN37	Transient receptor potential cation channel subfamily M member 4 (Long transient receptor potential channel 4) (LTrpC4)	2	2	1125	54795	TRPM4	TRPM4_HUMAN	Q8TD43	tm	integral to plasma membrane	1213	5
66109	Tspan13	TSN13_MOUSE	Q9D8C2	Tetraspanin-13 (Tspan-13) (Transmembrane 4 superfamily member 13).	2	3	886	27075	TSPAN13	TSN13_HUMAN	O95857	tm	integral to plasma membrane	204	4
70423	Tspan15	Q99L35_MOUSE	Q99L35	tetraspanin 15	2	2	978	23555	TSPAN15	TSN15_HUMAN	O95858	tm	integral to plasma membrane	206	3
216350	Tspan8	Q8R3G9_MOUSE	Q8R3G9	tetraspanin 8	6	41	124	7103	TSPAN8	TSN8_HUMAN	P19075	tm	integral to membrane	235	4
22146	Tuba1c	TBA1C_MOUSE	P68373	Tubulin alpha-1C chain (Tubulin alpha-6 chain) (Alpha-tubulin 6) (Alpha-tubulin isotype M-alpha-6).	16	105	47	84790	TUBA1C	TBA1C_HUMAN	Q9BQE3	tm	cytoskeleton	449	0
22152	Tubb3	TBB3_MOUSE	Q9ERD7	Tubulin beta-3 chain.	4	6	589	10381	TUBB3	TBB3_HUMAN	Q13509	tm	cytoskeleton	450	0
66958	Txndc14	TXD14_MOUSE	Q9D710	Thioredoxin domain-containing protein 14 precursor.	2	2	1174	51075	TXNDC14	TXD14_HUMAN	Q9Y320	tm	integral to membrane	295	2
22201	Uba1	A6H6S6_MOUSE	A6H6S6	ubiquitin-like modifier activating enzyme 1	4	4	801	7317	UBA1	UBA1_HUMAN	P22314	tm	-	1058	0
22235	Ugdh	UGDH_MOUSE	O70475	UDP-glucose dehydrogenase	16	22	220	7358	UGDH	UGDH_HUMAN	O60701	tm	-	493	0
394436	Ugt1a1	UD11_MOUSE	Q63886	UDP glucuronosyltransferase 1 family, polypeptide A1	4	4	788	54658	UGT1A1	UD11_HUMAN	P22309	tm	integral to plasma membrane	535	2
394432	Ugt1a7c	UD17C_MOUSE	Q6ZQM8	UDP glucuronosyltransferase 1 family, polypeptide A7C	8	9	462					tm	integral to endoplasmic reticulum membrane	531	1
100727	Ugt2b34	Q8K154_MOUSE	Q8K154	UDP glucuronosyltransferase 2 family, polypeptide B34	3	3	927	10720	UGT2B11	UDB11_HUMAN	O75310	tm	integral to endoplasmic reticulum membrane	532	1
76589	Unc5cl	UN5CL_MOUSE	Q6R653	UNC5C-like protein (Protein unc-5 homolog C-like) (ZU5 and death domain-containing protein).	5	13	338	222643	UNC5CL	UN5CL_HUMAN	Q8IV45	tm	integral to membrane	518	1
22271	Upp1	UPP1_MOUSE	P52624	Uridine phosphorylase 1 (EC 2.4.2.3) (UrdPase) (UPase 1).	6	10	415	7378	UPP1	UPP1_HUMAN	Q16831	tm	cytoplasm	311	0
22319	Vamp3	VAMP3_MOUSE	P63024	Vesicle-associated membrane protein 3 (VAMP-3) (Synaptobrevin-3) (Cellubrevin) (CEB).	3	4	749	9341	VAMP3	VAMP3_HUMAN	Q15836	tm	integral to plasma membrane	103	1
20955	Vamp7	VAMP7_MOUSE	P70280	Vesicle-associated membrane protein 7 (VAMP-7) (Synaptobrevin-like protein 1).	2	1	1200	6845	VAMP7	VAMP7_HUMAN	P51809	tm	integral to late endosome membrane	220	1
22320	Vamp8	VAMP8_MOUSE	O70404	Vesicle-associated membrane protein 8 (VAMP-8) (Endobrevin) (Edb).	5	8	502	8673	VAMP8	VAMP8_HUMAN	Q9BV40	tm	integral to early endosome membrane	101	1
22361	Vnn1	VNN1_MOUSE	Q9Z0K8	Pantetheinase precursor (EC 3.5.1.92) (Pantetheine hydrolase) (Vascular non-inflammatory molecule 1) (Vanin-1).	9	37	130	8876	VNN1	VNN1_HUMAN	O95497	tm	anchored to membrane	512	0

183

Gene ID (mouse)	Gene Symbol (mouse)	Swissprot protein ID (mouse)	Swissprot protein AC (mouse)	Gene Description	Max Diff peptide	Avg Pep count	Rank order	Ortho Gene ID	Ortho Gene Symbol	Swissprot protein ID	Swissprot protein AC	Topology	GO term	Protein length	Number of TM
66700	Vps24	CHMP3_MOUSE	Q9CQ10	Charged multivesicular body protein 3 (Chromatin-modifying protein 3) (Vacuolar protein-sorting-associated protein 24).	4	8	505	51652	VPS24	CHMP3_HUMAN	Q9Y3E7	tm	integral to mitochondrial outer membrane	224	0
53612	Vti1b	VTI1B_MOUSE	O88384	Vesicle transport through interaction with t-SNAREs homolog 1B (Vesicle transport v-SNARE protein Vti1-like 1) (Vti1-rp1).	6	7	552	10490	VTI1B	VTI1B_HUMAN	Q9UEU0	tm	integral to Golgi membrane	232	1
70465	Wdr77	MEP50_MOUSE	Q99J09	Methylosome protein 50 (MEP-50) (WD repeat-containing protein 77).	2	2	1080	79084	WDR77	MEP50_HUMAN	Q9BQA1	tm	cytoplasm	342	0
170750	Xpnpep1	XPP1_MOUSE	Q6P1B1	Xaa-Pro aminopeptidase 1 (EC 3.4.11.9) (X-Pro aminopeptidase 1) (X- prolyl aminopeptidase 1, soluble)	11	13	339	7511	XPNPEP1	XPP1_HUMAN	Q9NQW7	tm	cytoplasm	623	0
22630	Ywhaq	1433T_MOUSE	P68254	14-3-3 protein theta (14-3-3 protein tau)	19	50	101	10971	YWHAQ	1433T_HUMAN	P27348	tm	cytoplasm	245	0

184

A2. BBM protein identification from small intestine sections

Total list of the 1625 identified BBM proteins from sections or from whole mucosa. The protein must have at least a diff. peptide count of two in any of the 4 analysis to be included in the table. The proteins are listed by decreasing maximum number of different peptides.

Gene Symbol: associated gene symbol

Gene Description: description of gene product as provided by the internal Roche database entry

Sum of Sum_Pep_count Pep count: sum of peptide counts for all the LC-MS analysis of a given protein for a section or the whole mucosa analysis.

Max of Max_Diff_peptide: maximum number of different peptides observed in any of the LC-MS analysis from a given section or whole mucosa analysis

Proximal, central, distal: small intestine sections in contrast to whole mucosa

Gene Symbol	Gene Description	Data	proximal	central	distal	total mucosa
Myo7b	myosin VIIb	Sum of Sum_pept_count	599	779	593	917
		Max of Max_Dif_pept	108	114	80	118
2010204N08Rik	RIKEN cDNA 2010204N08 gene	Sum of Sum_pept_count	1508	1354	1408	1068
		Max of Max_Dif_pept	96	91	86	83
Mgam	maltase-glucoamylase	Sum of Sum_pept_count	1358	1253	1822	901
		Max of Max_Dif_pept	83	87	85	76
Cltc	clathrin, heavy polypeptide (Hc)	Sum of Sum_pept_count	0	0	0	464
		Max of Max_Dif_pept	0	0	0	87
Dync1h1	dynein cytoplasmic 1 heavy chain 1	Sum of Sum_pept_count	0	0	2	98
		Max of Max_Dif_pept	0	0	1	83
Myo1a	myosin IA	Sum of Sum_pept_count	856	986	990	984
		Max of Max_Dif_pept	73	80	75	78
Vil1	villin 1	Sum of Sum_pept_count	620	1041	817	1091
		Max of Max_Dif_pept	56	66	60	59
Anpep	alanyl (membrane) aminopeptidase	Sum of Sum_pept_count	1616	1294	1593	1088
		Max of Max_Dif_pept	62	55	57	60
Enpep	glutamyl aminopeptidase	Sum of Sum_pept_count	472	413	550	445
		Max of Max_Dif_pept	56	45	51	53
Ace	angiotensin I converting enzyme (peptidyl-dipeptidase A) 1	Sum of Sum_pept_count	229	133	76	133
		Max of Max_Dif_pept	56	36	29	47
Abcb1a	ATP-binding cassette, sub-family B (MDR/TAP), member 1A	Sum of Sum_pept_count	662	322	592	320
		Max of Max_Dif_pept	54	43	45	48
Lct	lactase	Sum of Sum_pept_count	621	485	164	337
		Max of Max_Dif_pept	53	45	40	51
Myo1d	myosin ID	Sum of Sum_pept_count	190	158	246	201
		Max of Max_Dif_pept	40	42	43	47
Abcc2	ATP-binding cassette, sub-family C (CFTR/MRP), member 2	Sum of Sum_pept_count	205	81	61	88
		Max of Max_Dif_pept	46	28	17	29
Nos2	nitric oxide synthase 2, inducible, macrophage	Sum of Sum_pept_count	0	20	171	29
		Max of Max_Dif_pept	0	17	43	14
Pdzk1	PDZ domain containing 1	Sum of Sum_pept_count	205	356	218	297
		Max of Max_Dif_pept	29	38	28	42
Mme	membrane metallo endopeptidase	Sum of Sum_pept_count	439	214	151	215
		Max of Max_Dif_pept	41	37	32	38
AI427122	expressed sequence AI427122	Sum of Sum_pept_count	98	200	142	198
		Max of Max_Dif_pept	32	41	38	38
Dpp4	dipeptidylpeptidase 4	Sum of Sum_pept_count	169	159	198	122
		Max of Max_Dif_pept	32	33	37	32
Papss2	3'-phosphoadenosine 5'-phosphosulfate synthase 2	Sum of Sum_pept_count	21	7	3	236
		Max of Max_Dif_pept	14	6	3	36
Npc1l1	NPC1-like 1	Sum of Sum_pept_count	200	179	215	148
		Max of Max_Dif_pept	36	31	33	29
Ezr	ezrin	Sum of Sum_pept_count	58	294	145	180
		Max of Max_Dif_pept	16	36	23	34
Hspa8	heat shock protein 8	Sum of Sum_pept_count	80	118	144	113
		Max of Max_Dif_pept	32	33	35	30
Duox2	dual oxidase 2	Sum of Sum_pept_count	1	18	191	41
		Max of Max_Dif_pept	1	6	35	18
Atp1a1	ATPase, Na+/K+ transporting, alpha 1 polypeptide	Sum of Sum_pept_count	204	97	361	261
		Max of Max_Dif_pept	22	18	29	34
Anxa2	annexin A2	Sum of Sum_pept_count	103	204	218	138
		Max of Max_Dif_pept	25	34	34	32
Alb	albumin	Sum of Sum_pept_count	21	50	64	70
		Max of Max_Dif_pept	19	33	34	25
Ace2	angiotensin I converting enzyme (peptidyl-dipeptidase A) 2	Sum of Sum_pept_count	432	335	330	234
		Max of Max_Dif_pept	31	29	27	33
Akp3	alkaline phosphatase 3, intestine, not Mn requiring	Sum of Sum_pept_count	390	26	1	147
		Max of Max_Dif_pept	32	10	1	26
Gna11	guanine nucleotide binding protein, alpha 11	Sum of Sum_pept_count	365	286	298	207
		Max of Max_Dif_pept	30	31	31	27
Enpp3	ectonucleotide pyrophosphatase/phosphodiesterase 3	Sum of Sum_pept_count	246	75	23	113
		Max of Max_Dif_pept	31	17	11	23
Slc3a1	solute carrier family 3, member 1	Sum of Sum_pept_count	138	117	189	75
		Max of Max_Dif_pept	25	30	28	24
Pcdh24	protocadherin 24	Sum of Sum_pept_count	340	232	242	283
		Max of Max_Dif_pept	18	16	14	30

Gene Symbol	Gene Description	Data	proximal	central	distal	total mucosa
Eps8l3	ESP8-like 3	Sum of Sum_pept_count	224	404	299	272
		Max of Max_Dif_pept	18	30	25	24
Treh	trehalase (brush-border membrane glycoprotein)	Sum of Sum_pept_count	186	110	79	92
		Max of Max_Dif_pept	29	26	21	23
Anxa4	annexin A4	Sum of Sum_pept_count	98	271	208	118
		Max of Max_Dif_pept	23	29	29	23
Prkg2	protein kinase, cGMP-dependent, type II	Sum of Sum_pept_count	104	94	62	79
		Max of Max_Dif_pept	22	28	19	26
Pkm2	pyruvate kinase, muscle	Sum of Sum_pept_count	37	74	49	142
		Max of Max_Dif_pept	18	28	18	28
Naaladl1	N-acetylated alpha-linked acidic dipeptidase-like 1	Sum of Sum_pept_count	93	163	312	148
		Max of Max_Dif_pept	22	28	28	25
Alpi	alkaline phosphatase, intestinal	Sum of Sum_pept_count	432	192	181	276
		Max of Max_Dif_pept	28	25	22	23
Actg1	actin, gamma, cytoplasmic 1	Sum of Sum_pept_count	430	1225	802	1252
		Max of Max_Dif_pept	25	28	23	27
Actb	actin, beta	Sum of Sum_pept_count	430	1225	802	1252
		Max of Max_Dif_pept	25	28	23	27
Lta4h	leukotriene A4 hydrolase	Sum of Sum_pept_count	1	13	6	51
		Max of Max_Dif_pept	1	9	4	27
Myo7a	myosin VIIa	Sum of Sum_pept_count	45	63	42	72
		Max of Max_Dif_pept	8	16	8	26
Clca3	chloride channel calcium activated 3	Sum of Sum_pept_count	33	52	51	56
		Max of Max_Dif_pept	14	18	20	26
Prkcd	protein kinase C, delta	Sum of Sum_pept_count	50	73	56	36
		Max of Max_Dif_pept	14	25	16	17
Pdcd6ip	programmed cell death 6 interacting protein	Sum of Sum_pept_count	6	51	32	24
		Max of Max_Dif_pept	3	25	16	18
Mep1b	meprin 1 beta	Sum of Sum_pept_count	267	223	309	196
		Max of Max_Dif_pept	24	24	25	22
Gnaq	guanine nucleotide binding protein, alpha q polypeptide	Sum of Sum_pept_count	185	116	147	65
		Max of Max_Dif_pept	25	23	24	19
Fasn	fatty acid synthase	Sum of Sum_pept_count	0	0	0	35
		Max of Max_Dif_pept	0	0	0	25
Xpnpep2	X-prolyl aminopeptidase (aminopeptidase P) 2, membrane-bound	Sum of Sum_pept_count	40	83	124	50
		Max of Max_Dif_pept	14	22	24	18
Slc5a1	solute carrier family 5 (sodium/glucose cotransporter), member 1	Sum of Sum_pept_count	407	353	391	287
		Max of Max_Dif_pept	22	24	21	24
Hsp90ab1	heat shock protein 90kDa alpha (cytosolic), class B member 1	Sum of Sum_pept_count	14	22	45	149
		Max of Max_Dif_pept	4	8	14	24
Anxa13	annexin A13	Sum of Sum_pept_count	106	133	161	84
		Max of Max_Dif_pept	22	21	23	24
Abcg2	ATP-binding cassette, sub-family G (WHITE), member 2	Sum of Sum_pept_count	95	54	97	60
		Max of Max_Dif_pept	24	20	21	17
Ush1c	Usher syndrome 1C homolog (human)	Sum of Sum_pept_count	44	138	60	103
		Max of Max_Dif_pept	13	19	17	23
Slc9a3r1	solute carrier family 9 (sodium/hydrogen exchanger), member 3 regulator 1	Sum of Sum_pept_count	44	140	78	127
		Max of Max_Dif_pept	8	23	10	23
Iqgap2	IQ motif containing GTPase activating protein 2	Sum of Sum_pept_count	7	29	20	54
		Max of Max_Dif_pept	3	6	6	23
Frk	fyn-related kinase	Sum of Sum_pept_count	62	55	47	68
		Max of Max_Dif_pept	20	23	17	19
Dak	dihydroxyacetone kinase 2 homolog (yeast)	Sum of Sum_pept_count	18	31	18	59
		Max of Max_Dif_pept	14	20	15	23
Atp8b1	ATPase, class I, type 8B, member 1	Sum of Sum_pept_count	87	63	113	52
		Max of Max_Dif_pept	22	23	23	16
Abcb1b	ATP-binding cassette, sub-family B (MDR/TAP), member 1B	Sum of Sum_pept_count	28	15	74	14
		Max of Max_Dif_pept	23	5	21	4
Stxbp2	syntaxin binding protein 2	Sum of Sum_pept_count	63	80	81	67
		Max of Max_Dif_pept	12	19	13	22
Slc15a1	solute carrier family 15 (oligopeptide transporter), member 1	Sum of Sum_pept_count	113	115	144	93
		Max of Max_Dif_pept	20	22	22	19

Gene Symbol	Gene Description	Data	proximal	central	distal	total mucosa
Hsp90ab1	heat shock protein 90kDa alpha (cytosolic), class B member 1	Sum of Sum_pept_count	14	22	45	149
		Max of Max_Dif_pept	4	8	14	24
Anxa13	annexin A13	Sum of Sum_pept_count	106	133	161	84
		Max of Max_Dif_pept	22	21	23	24
Abcg2	ATP-binding cassette, sub-family G (WHITE), member 2	Sum of Sum_pept_count	95	54	97	60
		Max of Max_Dif_pept	24	20	21	17
Ush1c	Usher syndrome 1C homolog (human)	Sum of Sum_pept_count	44	138	60	103
		Max of Max_Dif_pept	13	19	17	23
Slc9a3r1	solute carrier family 9 (sodium/hydrogen exchanger), member 3 regulator 1	Sum of Sum_pept_count	44	140	78	127
		Max of Max_Dif_pept	8	23	10	23
Iqgap2	IQ motif containing GTPase activating protein 2	Sum of Sum_pept_count	7	29	20	54
		Max of Max_Dif_pept	3	6	6	23
Frk	fyn-related kinase	Sum of Sum_pept_count	62	55	47	68
		Max of Max_Dif_pept	20	23	17	19
Dak	dihydroxyacetone kinase 2 homolog (yeast)	Sum of Sum_pept_count	18	31	18	59
		Max of Max_Dif_pept	14	20	15	23
Atp8b1	ATPase, class I, type 8B, member 1	Sum of Sum_pept_count	87	63	113	52
		Max of Max_Dif_pept	22	23	23	16
Abcb1b	ATP-binding cassette, sub-family B (MDR/TAP), member 1B	Sum of Sum_pept_count	28	15	74	14
		Max of Max_Dif_pept	23	5	21	4
Stxbp2	syntaxin binding protein 2	Sum of Sum_pept_count	63	80	81	67
		Max of Max_Dif_pept	12	19	13	22
Slc15a1	solute carrier family 15 (oligopeptide transporter), member 1	Sum of Sum_pept_count	113	115	144	93
		Max of Max_Dif_pept	20	22	22	19
Prkar2a	protein kinase, cAMP dependent regulatory, type II alpha	Sum of Sum_pept_count	171	103	119	98
		Max of Max_Dif_pept	22	21	18	22
Plcb3	phospholipase C, beta 3	Sum of Sum_pept_count	32	120	87	21
		Max of Max_Dif_pept	11	22	15	6
Ehd1	EH-domain containing 1	Sum of Sum_pept_count	37	41	29	41
		Max of Max_Dif_pept	14	22	17	19
Eef2	eukaryotic translation elongation factor 2	Sum of Sum_pept_count	20	44	30	70
		Max of Max_Dif_pept	4	12	7	22
Ywhaz	tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, zeta polypeptide	Sum of Sum_pept_count	32	51	70	81
		Max of Max_Dif_pept	13	16	17	21
Tubb5	tubulin, beta 5	Sum of Sum_pept_count	67	95	67	128
		Max of Max_Dif_pept	16	19	16	21
Tubb2c	tubulin, beta 2c	Sum of Sum_pept_count	69	59	75	179
		Max of Max_Dif_pept	16	18	14	21
Prkca	protein kinase C, alpha	Sum of Sum_pept_count	31	36	26	21
		Max of Max_Dif_pept	14	21	15	12
Idh1	isocitrate dehydrogenase 1 (NADP+), soluble	Sum of Sum_pept_count	27	19	23	34
		Max of Max_Dif_pept	14	13	11	21
Gdi2	guanosine diphosphate (GDP) dissociation inhibitor 2	Sum of Sum_pept_count	28	50	39	26
		Max of Max_Dif_pept	14	21	14	17
Eno1	enolase 1, alpha non-neuron	Sum of Sum_pept_count	14	33	29	57
		Max of Max_Dif_pept	9	17	13	21
Dpep1	dipeptidase 1 (renal)	Sum of Sum_pept_count	135	134	155	108
		Max of Max_Dif_pept	19	20	21	17
Actn4	actinin alpha 4	Sum of Sum_pept_count	0	11	9	28
		Max of Max_Dif_pept	0	6	2	21
Tkt	transketolase	Sum of Sum_pept_count	10	24	26	68
		Max of Max_Dif_pept	8	17	17	20
Susd2	sushi domain containing 2	Sum of Sum_pept_count	44	41	54	43
		Max of Max_Dif_pept	18	17	20	19
Smpdl3b	sphingomyelin phosphodiesterase, acid-like 3B	Sum of Sum_pept_count	54	37	39	43
		Max of Max_Dif_pept	20	17	16	13
Pld1	phospholipase D1	Sum of Sum_pept_count	100	57	85	51
		Max of Max_Dif_pept	20	18	17	13
Iigp2	interferon inducible GTPase 2	Sum of Sum_pept_count	59	24	41	19
		Max of Max_Dif_pept	20	11	8	11
Gna13	guanine nucleotide binding protein, alpha 13	Sum of Sum_pept_count	44	43	60	51
		Max of Max_Dif_pept	20	16	15	12

Gene Symbol	Gene Description	Data	proximal	central	distal	total mucosa
Ahcyl1	S-adenosylhomocysteine hydrolase-like 1	Sum of Sum_pept_count	3	25	4	17
		Max of Max_Dif_pept	3	17	4	13
Gapdh	glyceraldehyde-3-phosphate dehydrogenase	Sum of Sum_pept_count	47	66	58	118
		Max of Max_Dif_pept	12	14	13	17
2010002M12Rik	RIKEN cDNA 2010002M12 gene	Sum of Sum_pept_count	40	24	20	34
		Max of Max_Dif_pept	17	14	10	14
Acta1	actin, alpha 1, skeletal muscle	Sum of Sum_pept_count	154	360	376	379
		Max of Max_Dif_pept	11	14	17	14
Cndp2	CNDP dipeptidase 2 (metallopeptidase M20 family)	Sum of Sum_pept_count	6	46	21	32
		Max of Max_Dif_pept	6	17	11	15
Baiap2l1	BAI1-associated protein 2-like 1	Sum of Sum_pept_count	25	123	88	60
		Max of Max_Dif_pept	5	16	13	17
Asah2	N-acylsphingosine amidohydrolase 2	Sum of Sum_pept_count	102	93	72	71
		Max of Max_Dif_pept	17	17	16	16
Tuba4a	tubulin, alpha 4A	Sum of Sum_pept_count	5	6	43	14
		Max of Max_Dif_pept	4	3	16	9
Slc9a3	solute carrier family 9 (sodium/hydrogen exchanger), member 3	Sum of Sum_pept_count	129	83	46	62
		Max of Max_Dif_pept	13	16	8	13
Slc27a2	solute carrier family 27 (fatty acid transporter), member 2	Sum of Sum_pept_count	32	3	1	11
		Max of Max_Dif_pept	16	3	1	10
Slc26a6	solute carrier family 26, member 6	Sum of Sum_pept_count	152	89	79	72
		Max of Max_Dif_pept	16	14	12	15
Rab8a	RAB8A, member RAS oncogene family	Sum of Sum_pept_count	72	33	53	44
		Max of Max_Dif_pept	16	16	15	14
Rab1b	RAB1B, member RAS oncogene family	Sum of Sum_pept_count	134	45	38	52
		Max of Max_Dif_pept	16	13	11	14
Rab14	RAB14, member RAS oncogene family	Sum of Sum_pept_count	45	28	36	41
		Max of Max_Dif_pept	15	16	13	13
Psmc5	protease (prosome, macropain) 26S subunit, ATPase 5	Sum of Sum_pept_count	0	0	1	21
		Max of Max_Dif_pept	0	0	1	16
Abcg5	ATP-binding cassette, sub-family G (WHITE), member 5	Sum of Sum_pept_count	55	34	41	29
		Max of Max_Dif_pept	15	16	14	12
Ppp2r1a	protein phosphatase 2 (formerly 2A), regulatory subunit A (PR 65), alpha isoform	Sum of Sum_pept_count	3	13	5	29
		Max of Max_Dif_pept	3	8	2	16
Aldh1a1	aldehyde dehydrogenase family 1, subfamily A1	Sum of Sum_pept_count	3	12	3	21
		Max of Max_Dif_pept	3	10	3	16
Pfkp	phosphofructokinase, platelet	Sum of Sum_pept_count	7	16	18	29
		Max of Max_Dif_pept	4	11	12	16
Nars	asparaginyl-tRNA synthetase	Sum of Sum_pept_count	2	13	14	22
		Max of Max_Dif_pept	2	12	13	16
Mttp	microsomal triglyceride transfer protein	Sum of Sum_pept_count	0	1	0	39
		Max of Max_Dif_pept	0	1	0	16
Mep1a	meprin 1 alpha	Sum of Sum_pept_count	23	47	101	32
		Max of Max_Dif_pept	8	16	16	10
Lgals4	lectin, galactose binding, soluble 4	Sum of Sum_pept_count	88	91	106	121
		Max of Max_Dif_pept	11	13	11	16
Hsp90aa1	heat shock protein 90, alpha (cytosolic), class A member 1	Sum of Sum_pept_count	8	12	12	30
		Max of Max_Dif_pept	2	6	5	16
Gpx2	glutathione peroxidase 2	Sum of Sum_pept_count	15	34	35	24
		Max of Max_Dif_pept	11	16	16	15
Gnb1	guanine nucleotide binding protein (G protein), beta 1	Sum of Sum_pept_count	125	138	119	114
		Max of Max_Dif_pept	15	16	16	16
Gnas	GNAS (guanine nucleotide binding protein, alpha stimulating) complex locus	Sum of Sum_pept_count	52	33	57	29
		Max of Max_Dif_pept	16	16	14	11
Enpp7	ectonucleotide pyrophosphatase/phosphodiesterase 7	Sum of Sum_pept_count	42	84	83	71
		Max of Max_Dif_pept	10	12	16	11
EG638904	predicted gene, EG638904	Sum of Sum_pept_count	133	93	199	84
		Max of Max_Dif_pept	15	14	16	14

189

Gene Symbol	Gene Description	Data	proximal	central	distal	total mucosa
Dsp	desmoplakin	Sum of Sum_pept_count	21	10	43	14
		Max of Max_Dif_pept	8	7	16	4
Casp1	caspase 1	Sum of Sum_pept_count	21	18	15	26
		Max of Max_Dif_pept	16	12	10	13
Akr1a4	aldo-keto reductase family 1, member A4 (aldehyde reductase)	Sum of Sum_pept_count	7	15	14	51
		Max of Max_Dif_pept	6	12	11	16
Ywhab	tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, beta polypeptide	Sum of Sum_pept_count	20	23	27	52
		Max of Max_Dif_pept	8	10	9	15
Tpi1	triosephosphate isomerase 1	Sum of Sum_pept_count	13	16	16	50
		Max of Max_Dif_pept	10	7	10	15
Tmem16f	transmembrane protein 16F	Sum of Sum_pept_count	61	51	55	41
		Max of Max_Dif_pept	15	14	14	14
Styk1	serine/threonine/tyrosine kinase 1	Sum of Sum_pept_count	87	69	108	44
		Max of Max_Dif_pept	15	15	15	13
Slc30a10	solute carrier family 30, member 10	Sum of Sum_pept_count	65	9	18	17
		Max of Max_Dif_pept	15	2	3	5
Scin	scinderin	Sum of Sum_pept_count	2	14	10	17
		Max of Max_Dif_pept	2	13	7	15
Rp2h	retinitis pigmentosa 2 homolog (human)	Sum of Sum_pept_count	36	25	38	30
		Max of Max_Dif_pept	15	12	10	11
Rnpep	arginyl aminopeptidase (aminopeptidase B)	Sum of Sum_pept_count	2	13	7	19
		Max of Max_Dif_pept	2	13	6	15
Rnh1	ribonuclease/angiogenin inhibitor 1	Sum of Sum_pept_count	7	17	12	17
		Max of Max_Dif_pept	7	15	10	11
Rab1	RAB1, member RAS oncogene family	Sum of Sum_pept_count	264	76	102	117
		Max of Max_Dif_pept	15	14	12	15
Prdx1	peroxiredoxin 1	Sum of Sum_pept_count	19	29	24	30
		Max of Max_Dif_pept	10	15	10	13
Ppp1ca	protein phosphatase 1, catalytic subunit, alpha isoform	Sum of Sum_pept_count	8	23	17	24
		Max of Max_Dif_pept	7	15	11	14
Anxa7	annexin A7	Sum of Sum_pept_count	3	40	24	12
		Max of Max_Dif_pept	2	15	8	7
Pklr	pyruvate kinase liver and red blood cell	Sum of Sum_pept_count	3	12	8	28
		Max of Max_Dif_pept	3	10	7	15
OTTMUSG0000000	predicted gene, OTTMUSG00000005723	Sum of Sum_pept_count	14	11	8	19
		Max of Max_Dif_pept	10	10	6	15
Mupcdh	mucin-like protocadherin	Sum of Sum_pept_count	142	93	104	92
		Max of Max_Dif_pept	15	14	12	14
Lyn	Yamaguchi sarcoma viral (v-yes-1) oncogene homolog	Sum of Sum_pept_count	35	41	34	32
		Max of Max_Dif_pept	13	15	11	12
Lap3	leucine aminopeptidase 3	Sum of Sum_pept_count	7	20	11	32
		Max of Max_Dif_pept	4	14	10	15
Acsl5	acyl-CoA synthetase long-chain family member 5	Sum of Sum_pept_count	16	6	17	34
		Max of Max_Dif_pept	13	5	8	15
Eps8l2	EPS8-like 2	Sum of Sum_pept_count	36	113	61	91
		Max of Max_Dif_pept	5	12	6	15
Eif4a1	eukaryotic translation initiation factor 4A1	Sum of Sum_pept_count	19	34	28	41
		Max of Max_Dif_pept	10	12	9	15
Coro1c	coronin, actin binding protein 1C	Sum of Sum_pept_count	14	25	27	14
		Max of Max_Dif_pept	13	15	14	12
Clic5	chloride intracellular channel 5	Sum of Sum_pept_count	90	143	99	169
		Max of Max_Dif_pept	11	15	12	14
Cfl1	cofilin 1, non-muscle	Sum of Sum_pept_count	18	52	34	56
		Max of Max_Dif_pept	8	15	11	15
Calml4	calmodulin-like 4	Sum of Sum_pept_count	9	45	13	31
		Max of Max_Dif_pept	5	15	7	15
Ywhag	tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, gamma polypeptide	Sum of Sum_pept_count	2	18	22	40
		Max of Max_Dif_pept	2	12	9	14
Tmc4	transmembrane channel-like gene family 4	Sum of Sum_pept_count	33	32	50	33
		Max of Max_Dif_pept	13	11	14	12

Gene Symbol	Gene Description	Data	proximal	central	distal	total mucosa
Src	Rous sarcoma oncogene	Sum of Sum_pept_count	24	25	24	25
		Max of Max_Dif_pept	11	14	10	10
Slc5a4b	solute carrier family 5 (neutral amino acid transporters, system A), member 4b	Sum of Sum_pept_count	88	85	100	52
		Max of Max_Dif_pept	11	14	11	13
Slc13a2	solute carrier family 13 (sodium-dependent dicarboxylate transporter), member 2	Sum of Sum_pept_count	52	43	72	37
		Max of Max_Dif_pept	13	10	14	11
Rap1b	RAS related protein 1b	Sum of Sum_pept_count	70	40	50	33
		Max of Max_Dif_pept	14	13	11	12
Rap1a	RAS-related protein-1a	Sum of Sum_pept_count	56	35	23	16
		Max of Max_Dif_pept	14	13	6	7
Rab8b	RAB8B, member RAS oncogene family	Sum of Sum_pept_count	45	22	32	27
		Max of Max_Dif_pept	14	12	12	12
Rab35	RAB35, member RAS oncogene family	Sum of Sum_pept_count	73	29	58	34
		Max of Max_Dif_pept	14	12	14	10
Psmc2	proteasome (prosome, macropain) 26S subunit, ATPase 2	Sum of Sum_pept_count	0	0	1	22
		Max of Max_Dif_pept	0	0	1	14
Prdx6	peroxiredoxin 6	Sum of Sum_pept_count	17	9	14	29
		Max of Max_Dif_pept	10	5	10	14
Aldh1b1	aldehyde dehydrogenase 1 family, member B1	Sum of Sum_pept_count	0	0	0	53
		Max of Max_Dif_pept	0	0	0	14
Pafah1b1	platelet-activating factor acetylhydrolase, isoform 1b, beta1 subunit	Sum of Sum_pept_count	0	0	0	16
		Max of Max_Dif_pept	0	0	0	14
Muc13	mucin 13, epithelial transmembrane	Sum of Sum_pept_count	366	336	390	345
		Max of Max_Dif_pept	13	13	14	12
Mapk3	mitogen-activated protein kinase 3	Sum of Sum_pept_count	8	16	9	44
		Max of Max_Dif_pept	4	7	5	14
H2-K1	histocompatibility 2, K1, K region	Sum of Sum_pept_count	24	12	36	31
		Max of Max_Dif_pept	10	7	14	14
Gpr128	G protein-coupled receptor 128	Sum of Sum_pept_count	124	75	110	81
		Max of Max_Dif_pept	13	12	12	14
Gnb2	guanine nucleotide binding protein (G protein), beta 2	Sum of Sum_pept_count	76	90	88	97
		Max of Max_Dif_pept	11	14	14	13
Aldob	aldolase B, fructose-bisphosphate	Sum of Sum_pept_count	21	42	35	51
		Max of Max_Dif_pept	10	14	12	11
Gfpt1	glutamine fructose-6-phosphate transaminase 1	Sum of Sum_pept_count	0	6	18	16
		Max of Max_Dif_pept	0	5	11	14
Ephx2	epoxide hydrolase 2, cytoplasmic	Sum of Sum_pept_count	11	11	1	24
		Max of Max_Dif_pept	10	9	1	14
Eef1a1	eukaryotic translation elongation factor 1 alpha 1	Sum of Sum_pept_count	53	61	70	98
		Max of Max_Dif_pept	9	14	11	13
Dstn	destrin	Sum of Sum_pept_count	34	62	44	53
		Max of Max_Dif_pept	12	14	14	14
Abce1	ATP-binding cassette, sub-family E (OABP), member 1	Sum of Sum_pept_count	0	0	1	14
		Max of Max_Dif_pept	0	0	1	13
Trim14	tripartite motif-containing 14	Sum of Sum_pept_count	12	9	17	8
		Max of Max_Dif_pept	11	8	13	8
Stk25	serine/threonine kinase 25 (yeast)	Sum of Sum_pept_count	17	28	19	22
		Max of Max_Dif_pept	8	13	6	9
Anxa11	annexin A11	Sum of Sum_pept_count	31	68	55	31
		Max of Max_Dif_pept	11	13	12	10
Sfn	stratifin	Sum of Sum_pept_count	12	16	18	23
		Max of Max_Dif_pept	5	10	9	13
Rps6ka1	ribosomal protein S6 kinase polypeptide 1	Sum of Sum_pept_count	5	8	6	21
		Max of Max_Dif_pept	4	8	6	13
Rab5c	RAB5C, member RAS oncogene family	Sum of Sum_pept_count	71	42	91	53
		Max of Max_Dif_pept	13	9	8	11
Rab2a	RAB2A, member RAS oncogene family	Sum of Sum_pept_count	64	51	75	58
		Max of Max_Dif_pept	13	13	12	12

Gene Symbol	Gene Description	Data	proximal	central	distal	total mucosa
Rab11b	RAB11B, member RAS oncogene family	Sum of Sum_pept_count	63	49	84	47
		Max of Max_Dif_pept	13	11	11	13
Rab10	RAB10, member RAS oncogene family	Sum of Sum_pept_count	55	22	50	41
		Max of Max_Dif_pept	13	12	13	13
Psmd11	proteasome (prosome, macropain) 26S subunit, non-ATPase, 11	Sum of Sum_pept_count	1	1	2	21
		Max of Max_Dif_pept	1	1	1	13
Ppia	peptidylprolyl isomerase A	Sum of Sum_pept_count	22	40	47	53
		Max of Max_Dif_pept	11	11	12	13
Nlrp6	NLR family, pyrin domain containing 6	Sum of Sum_pept_count	55	18	19	17
		Max of Max_Dif_pept	13	7	7	8
Ap1m2	adaptor protein complex AP-1, mu 2 subunit	Sum of Sum_pept_count	0	2	2	18
		Max of Max_Dif_pept	0	2	2	13
Lgals2	lectin, galactose-binding, soluble 2	Sum of Sum_pept_count	39	27	37	59
		Max of Max_Dif_pept	9	9	10	13
Lcp1	lymphocyte cytosolic protein 1	Sum of Sum_pept_count	0	20	16	21
		Max of Max_Dif_pept	0	13	11	9
Hspa1b	heat shock protein 1B	Sum of Sum_pept_count	3	9	15	15
		Max of Max_Dif_pept	1	8	12	13
Hspa1a	heat shock protein 1A	Sum of Sum_pept_count	3	9	15	15
		Max of Max_Dif_pept	1	8	12	13
Gucy2c	guanylate cyclase 2c	Sum of Sum_pept_count	39	20	29	19
		Max of Max_Dif_pept	13	7	12	6
Ahcy	S-adenosylhomocysteine hydrolase	Sum of Sum_pept_count	7	14	9	13
		Max of Max_Dif_pept	7	13	8	12
Cpne2	copine II	Sum of Sum_pept_count	7	20	13	7
		Max of Max_Dif_pept	6	13	7	6
Cbr1	carbonyl reductase 1	Sum of Sum_pept_count	8	11	5	24
		Max of Max_Dif_pept	5	6	4	13
Bpnt1	bisphosphate 3'-nucleotidase 1	Sum of Sum_pept_count	37	12	14	25
		Max of Max_Dif_pept	12	9	7	13
Baiap2l2	BAI1-associated protein 2-like 2	Sum of Sum_pept_count	5	48	27	15
		Max of Max_Dif_pept	2	13	7	7
Unc5cl	unc-5 homolog C (C. elegans)-like	Sum of Sum_pept_count	43	18	30	25
		Max of Max_Dif_pept	12	7	7	8
Tagln2	transgelin 2	Sum of Sum_pept_count	12	35	19	26
		Max of Max_Dif_pept	7	12	8	10
Stx3	syntaxin 3	Sum of Sum_pept_count	59	41	46	37
		Max of Max_Dif_pept	12	9	9	12
Stom	stomatin	Sum of Sum_pept_count	75	76	142	92
		Max of Max_Dif_pept	9	9	12	12
Sh3bgrl	SH3-binding domain glutamic acid-rich protein like	Sum of Sum_pept_count	14	52	16	22
		Max of Max_Dif_pept	6	12	7	10
Actg2	actin, gamma 2, smooth muscle, enteric	Sum of Sum_pept_count	65	51	70	113
		Max of Max_Dif_pept	10	10	10	12
Rap2c	RAP2C, member of RAS oncogene family	Sum of Sum_pept_count	29	19	35	14
		Max of Max_Dif_pept	12	10	10	9
Rab11a	RAB11a, member RAS oncogene family	Sum of Sum_pept_count	52	29	71	50
		Max of Max_Dif_pept	12	10	10	11
Psmc3	proteasome (prosome, macropain) 26S subunit, ATPase 3	Sum of Sum_pept_count	1	3	2	28
		Max of Max_Dif_pept	1	3	2	12
Prkaca	protein kinase, cAMP dependent, catalytic, alpha	Sum of Sum_pept_count	15	18	18	16
		Max of Max_Dif_pept	6	7	8	12
Arf4	ADP-ribosylation factor 4	Sum of Sum_pept_count	84	68	73	46
		Max of Max_Dif_pept	12	10	9	7
P4hb	prolyl 4-hydroxylase, beta polypeptide	Sum of Sum_pept_count	0	1	11	16
		Max of Max_Dif_pept	0	1	4	12
Myo15b	myosin XVB	Sum of Sum_pept_count	33	80	61	68
		Max of Max_Dif_pept	10	12	12	11
Mtm1	X-linked myotubular myopathy gene 1	Sum of Sum_pept_count	1	14	8	16
		Max of Max_Dif_pept	1	12	8	8
Aldh9a1	aldehyde dehydrogenase 9, subfamily A1	Sum of Sum_pept_count	2	13	7	16
		Max of Max_Dif_pept	2	12	7	12

Gene Symbol	Gene Description	Data	proximal	central	distal	total mucosa
Krt5	keratin 5	Sum of Sum_pept_count	13	7	29	22
		Max of Max_Dif_pept	8	7	12	12
Kalrn	kalirin, RhoGEF kinase	Sum of Sum_pept_count	45	53	44	42
		Max of Max_Dif_pept	10	12	11	9
Inadl	InaD-like (Drosophila)	Sum of Sum_pept_count	58	74	68	23
		Max of Max_Dif_pept	12	12	11	6
Aldoa	aldolase A, fructose-bisphosphate	Sum of Sum_pept_count	6	19	21	15
		Max of Max_Dif_pept	6	11	12	8
Gpi1	glucose phosphate isomerase 1	Sum of Sum_pept_count	0	20	9	19
		Max of Max_Dif_pept	0	7	6	12
Gipc2	GIPC PDZ domain containing family, member 2	Sum of Sum_pept_count	2	18	6	15
		Max of Max_Dif_pept	2	12	5	10
Acta2	actin, alpha 2, smooth muscle, aorta	Sum of Sum_pept_count	65	51	70	113
		Max of Max_Dif_pept	10	10	10	12
Fabp2	fatty acid binding protein 2, intestinal	Sum of Sum_pept_count	12	208	32	36
		Max of Max_Dif_pept	4	12	5	10
Ap2m1	adaptor protein complex AP-2, mu1	Sum of Sum_pept_count	5	12	8	12
		Max of Max_Dif_pept	5	11	7	12
Cpne3	copine III	Sum of Sum_pept_count	4	17	16	11
		Max of Max_Dif_pept	4	12	10	9
Cobl	cordon-bleu	Sum of Sum_pept_count	45	90	34	94
		Max of Max_Dif_pept	8	11	7	12
Ceacam20	carcinoembryonic antigen-related cell adhesion molecule 20	Sum of Sum_pept_count	40	31	43	27
		Max of Max_Dif_pept	12	11	11	10
Calm3	calmodulin 3	Sum of Sum_pept_count	22	69	65	48
		Max of Max_Dif_pept	6	12	9	7
Calm2	calmodulin 2	Sum of Sum_pept_count	22	69	65	48
		Max of Max_Dif_pept	6	12	9	7
Calm1	calmodulin 1	Sum of Sum_pept_count	22	69	65	48
		Max of Max_Dif_pept	6	12	9	7
Cab39	calcium binding protein 39	Sum of Sum_pept_count	6	16	11	9
		Max of Max_Dif_pept	6	12	8	8
Adh1	alcohol dehydrogenase 1 (class I)	Sum of Sum_pept_count	4	12	0	22
		Max of Max_Dif_pept	4	7	0	11
Yes1	Yamaguchi sarcoma viral (v-yes) oncogene homolog 1	Sum of Sum_pept_count	30	28	26	25
		Max of Max_Dif_pept	11	11	10	9
Vps4b	vacuolar protein sorting 4b (yeast)	Sum of Sum_pept_count	5	19	9	11
		Max of Max_Dif_pept	5	11	8	7
Vnn1	vanin 1	Sum of Sum_pept_count	60	56	83	48
		Max of Max_Dif_pept	11	8	10	9
Sult1b1	sulfotransferase family 1B, member 1	Sum of Sum_pept_count	16	18	10	25
		Max of Max_Dif_pept	6	8	4	11
Snap23	synaptosomal-associated protein 23	Sum of Sum_pept_count	74	44	54	55
		Max of Max_Dif_pept	11	7	10	10
Slc5a4a	solute carrier family 5, member 4a	Sum of Sum_pept_count	42	41	32	32
		Max of Max_Dif_pept	9	11	9	10
Slc44a4	solute carrier family 44, member 4	Sum of Sum_pept_count	13	20	35	18
		Max of Max_Dif_pept	3	7	11	5
Akr1c12	aldo-keto reductase family 1, member C12	Sum of Sum_pept_count	3	0	3	14
		Max of Max_Dif_pept	3	0	3	11
Rpn1	ribophorin I	Sum of Sum_pept_count	0	0	9	11
		Max of Max_Dif_pept	0	0	8	11
Rac1	RAS-related C3 botulinum substrate 1	Sum of Sum_pept_count	80	50	64	38
		Max of Max_Dif_pept	10	10	9	11
Arf1	ADP-ribosylation factor 1	Sum of Sum_pept_count	81	83	95	52
		Max of Max_Dif_pept	11	10	9	8
Anxa5	annexin A5	Sum of Sum_pept_count	5	13	13	8
		Max of Max_Dif_pept	5	11	9	8
Psmc1	protease (prosome, macropain) 26S subunit, ATPase 1	Sum of Sum_pept_count	0	0	0	20
		Max of Max_Dif_pept	0	0	0	11
Prkag1	protein kinase, AMP-activated, gamma 1 non-catalytic subunit	Sum of Sum_pept_count	2	9	14	11
		Max of Max_Dif_pept	2	9	9	11

Gene Symbol	Gene Description	Data	proximal	central	distal	total mucosa
Arf3	ADP-ribosylation factor 3	Sum of Sum_pept_count	81	83	95	52
		Max of Max_Dif_pept	11	10	9	8
Ap1b1	adaptor protein complex AP-1, beta 1 subunit	Sum of Sum_pept_count	7	13	8	41
		Max of Max_Dif_pept	5	6	4	11
Pi4ka	phosphatidylinositol 4-kinase, catalytic, alpha polypeptide	Sum of Sum_pept_count	11	10	4	13
		Max of Max_Dif_pept	7	6	3	11
Arf5	ADP-ribosylation factor 5	Sum of Sum_pept_count	77	64	65	47
		Max of Max_Dif_pept	11	10	10	9
Myl6	myosin, light polypeptide 6, alkali, smooth muscle and non-muscle	Sum of Sum_pept_count	50	51	37	90
		Max of Max_Dif_pept	9	10	8	11
Mpp1	membrane protein, palmitoylated	Sum of Sum_pept_count	8	16	8	23
		Max of Max_Dif_pept	5	11	4	11
Lasp1	LIM and SH3 protein 1	Sum of Sum_pept_count	10	36	5	62
		Max of Max_Dif_pept	4	6	2	11
Lad1	ladinin	Sum of Sum_pept_count	34	35	11	66
		Max of Max_Dif_pept	7	11	5	10
Jup	junction plakoglobin	Sum of Sum_pept_count	16	7	25	14
		Max of Max_Dif_pept	11	7	9	5
Idh2	isocitrate dehydrogenase 2 (NADP+), mitochondrial	Sum of Sum_pept_count	0	1	0	17
		Max of Max_Dif_pept	0	1	0	11
Hspa5	heat shock protein 5	Sum of Sum_pept_count	0	0	12	8
		Max of Max_Dif_pept	0	0	11	7
H2-Ab1	histocompatibility 2, class II antigen A, beta 1	Sum of Sum_pept_count	9	8	31	15
		Max of Max_Dif_pept	7	7	11	8
Ap2a2	adaptor protein complex AP-2, alpha 2 subunit	Sum of Sum_pept_count	0	18	10	9
		Max of Max_Dif_pept	0	11	8	8
Gda	guanine deaminase	Sum of Sum_pept_count	12	19	8	21
		Max of Max_Dif_pept	10	8	6	11
Fabp6	fatty acid binding protein 6, ileal (gastrotropin)	Sum of Sum_pept_count	0	1	46	15
		Max of Max_Dif_pept	0	1	11	7
Dnaja1	DnaJ (Hsp40) homolog, subfamily A, member 1	Sum of Sum_pept_count	10	12	9	12
		Max of Max_Dif_pept	9	11	9	11
Ckb	creatine kinase, brain	Sum of Sum_pept_count	11	10	2	59
		Max of Max_Dif_pept	6	5	2	11
Cib1	calcium and integrin binding 1 (calmyrin)	Sum of Sum_pept_count	19	30	38	18
		Max of Max_Dif_pept	11	9	10	8
Chmp4b	chromatin modifying protein 4B	Sum of Sum_pept_count	26	27	30	43
		Max of Max_Dif_pept	8	9	9	11
Capzb	capping protein (actin filament) muscle Z-line, beta	Sum of Sum_pept_count	8	16	12	14
		Max of Max_Dif_pept	5	10	7	11
Xpnpep1	X-prolyl aminopeptidase (aminopeptidase P) 1, soluble	Sum of Sum_pept_count	3	27	23	26
		Max of Max_Dif_pept	1	10	8	9
Tubb3	tubulin, beta 3	Sum of Sum_pept_count	3	7	16	27
		Max of Max_Dif_pept	1	4	9	10
Adh6a	alcohol dehydrogenase 6A (class V)	Sum of Sum_pept_count	8	15	6	21
		Max of Max_Dif_pept	7	10	4	8
Tmc5	transmembrane channel-like gene family 5	Sum of Sum_pept_count	19	20	29	14
		Max of Max_Dif_pept	8	8	10	6
Tlr3	toll-like receptor 3	Sum of Sum_pept_count	13	8	20	7
		Max of Max_Dif_pept	5	4	10	4
Atp2b1	ATPase, Ca++ transporting, plasma membrane 1	Sum of Sum_pept_count	64	19	40	20
		Max of Max_Dif_pept	10	9	9	8
Tacstd1	tumor-associated calcium signal transducer 1	Sum of Sum_pept_count	48	36	82	81
		Max of Max_Dif_pept	8	9	9	10
Stk24	serine/threonine kinase 24 (STE20 homolog, yeast)	Sum of Sum_pept_count	2	27	11	18
		Max of Max_Dif_pept	1	10	4	8
2210407C18Rik	RIKEN cDNA 2210407C18 gene	Sum of Sum_pept_count	439	137	105	270
		Max of Max_Dif_pept	10	9	8	9
Slc6a19	solute carrier family 6 (neurotransmitter transporter), member 19	Sum of Sum_pept_count	57	45	50	38
		Max of Max_Dif_pept	10	9	8	9
Slc5a11	solute carrier family 5 (sodium/glucose cotransporter), member 11	Sum of Sum_pept_count	85	54	34	46
		Max of Max_Dif_pept	10	9	8	10

Gene Symbol	Gene Description	Data	proximal	central	distal	total mucosa
Sec31a	SEC31 homolog A (S. cerevisiae)	Sum of Sum_pept_count	0	1	0	61
		Max of Max_Dif_pept	0	1	0	10
Rhoa	ras homolog gene family, member A	Sum of Sum_pept_count	73	53	52	38
		Max of Max_Dif_pept	10	8	8	7
Bst1	bone marrow stromal cell antigen 1	Sum of Sum_pept_count	80	69	59	53
		Max of Max_Dif_pept	10	10	8	10
Rab21	RAB21, member RAS oncogene family	Sum of Sum_pept_count	20	9	14	13
		Max of Max_Dif_pept	10	6	7	8
Psmd13	proteasome (prosome, macropain) 26S subunit, non-ATPase, 13	Sum of Sum_pept_count	0	0	1	11
		Max of Max_Dif_pept	0	0	1	10
Psmd12	proteasome (prosome, macropain) 26S subunit, non-ATPase, 12	Sum of Sum_pept_count	0	0	1	11
		Max of Max_Dif_pept	0	0	1	10
Arpc2	actin related protein 2/3 complex, subunit 2	Sum of Sum_pept_count	6	17	14	18
		Max of Max_Dif_pept	5	10	9	9
Anxa6	annexin A6	Sum of Sum_pept_count	1	10	8	1
		Max of Max_Dif_pept	1	10	8	1
Ppp1cc	protein phosphatase 1, catalytic subunit, gamma isoform	Sum of Sum_pept_count	7	1	17	2
		Max of Max_Dif_pept	6	1	10	1
Ppp1cb	protein phosphatase 1, catalytic subunit, beta isoform	Sum of Sum_pept_count	7	21	17	20
		Max of Max_Dif_pept	6	10	10	10
Pepd	peptidase D	Sum of Sum_pept_count	6	11	8	14
		Max of Max_Dif_pept	5	8	7	10
Pafah2	platelet-activating factor acetylhydrolase 2	Sum of Sum_pept_count	0	3	0	10
		Max of Max_Dif_pept	0	3	0	10
Napa	N-ethylmaleimide sensitive fusion protein attachment protein alpha	Sum of Sum_pept_count	10	13	11	13
		Max of Max_Dif_pept	6	9	6	10
Atp6v1a	ATPase, H+ transporting, lysosomal V1 subunit A	Sum of Sum_pept_count	0	0	0	10
		Max of Max_Dif_pept	0	0	0	10
Mpp5	membrane protein, palmitoylated 5 (MAGUK p55 subfamily member 5)	Sum of Sum_pept_count	23	23	20	18
		Max of Max_Dif_pept	7	8	7	10
Mdh2	malate dehydrogenase 2, NAD (mitochondrial)	Sum of Sum_pept_count	0	10	6	11
		Max of Max_Dif_pept	0	10	6	7
Mapk1	mitogen-activated protein kinase 1	Sum of Sum_pept_count	0	9	4	20
		Max of Max_Dif_pept	0	6	4	10
Lgals3	lectin, galactose binding, soluble 3	Sum of Sum_pept_count	15	13	25	8
		Max of Max_Dif_pept	10	9	10	4
Krt31	keratin 31	Sum of Sum_pept_count	34	10	6	0
		Max of Max_Dif_pept	10	7	4	0
Cdh17	cadherin 17	Sum of Sum_pept_count	19	6	50	50
		Max of Max_Dif_pept	6	3	10	9
Ifit1	interferon-induced protein with tetratricopeptide repeats 1	Sum of Sum_pept_count	16	7	5	14
		Max of Max_Dif_pept	10	5	3	9
Gsdmd	gasdermin D	Sum of Sum_pept_count	17	22	13	21
		Max of Max_Dif_pept	7	10	5	6
Csnk1g3	casein kinase 1, gamma 3	Sum of Sum_pept_count	31	27	24	17
		Max of Max_Dif_pept	10	8	8	10
Celsr3	cadherin, EGF LAG seven-pass G-type receptor 3 (flamingo homolog, Drosophila)	Sum of Sum_pept_count	13	29	33	12
		Max of Max_Dif_pept	3	6	10	5
Vps35	vacuolar protein sorting 35	Sum of Sum_pept_count	2	7	9	12
		Max of Max_Dif_pept	2	6	8	9
1500003O03Rik	RIKEN cDNA 1500003O03 gene	Sum of Sum_pept_count	14	5	6	7
		Max of Max_Dif_pept	9	4	4	5
Actbl2	actin, beta-like 2	Sum of Sum_pept_count	0	8	34	4
		Max of Max_Dif_pept	0	7	9	4
Car4	carbonic anhydrase 4	Sum of Sum_pept_count	14	11	23	16
		Max of Max_Dif_pept	8	7	9	7
Tuba1a	tubulin, alpha 1A	Sum of Sum_pept_count	34	20	37	48
		Max of Max_Dif_pept	6	9	9	9
Trim31	tripartite motif-containing 31	Sum of Sum_pept_count	1	8	4	10
		Max of Max_Dif_pept	1	7	3	9
Sri	sorcin	Sum of Sum_pept_count	10	32	22	10
		Max of Max_Dif_pept	5	9	6	6

Gene Symbol	Gene Description	Data	proximal	central	distal	total mucosa
Sphk2	sphingosine kinase 2	Sum of Sum_pept_count	36	26	16	18
		Max of Max_Dif_pept	8	9	5	7
Capza2	capping protein (actin filament) muscle Z-line, alpha 2	Sum of Sum_pept_count	5	16	9	17
		Max of Max_Dif_pept	3	7	4	9
Slc27a4	solute carrier family 27 (fatty acid transporter), member 4	Sum of Sum_pept_count	1	0	2	10
		Max of Max_Dif_pept	1	0	1	9
Sec24a	SEC24 related gene family, member A (S. cerevisiae)	Sum of Sum_pept_count	0	5	1	21
		Max of Max_Dif_pept	0	4	1	9
Rhog	ras homolog gene family, member G	Sum of Sum_pept_count	25	22	27	19
		Max of Max_Dif_pept	8	9	8	7
Rhoc	ras homolog gene family, member C	Sum of Sum_pept_count	26	45	35	29
		Max of Max_Dif_pept	6	9	7	8
Rars	arginyl-tRNA synthetase	Sum of Sum_pept_count	0	0	0	14
		Max of Max_Dif_pept	0	0	0	9
Rala	v-ral simian leukemia viral oncogene homolog A (ras related)	Sum of Sum_pept_count	54	21	34	22
		Max of Max_Dif_pept	9	6	8	8
Rac2	RAS-related C3 botulinum substrate 2	Sum of Sum_pept_count	17	4	3	4
		Max of Max_Dif_pept	9	1	1	2
Rab4a	RAB4A, member RAS oncogene family	Sum of Sum_pept_count	16	8	15	11
		Max of Max_Dif_pept	8	7	9	6
Prdx5	peroxiredoxin 5	Sum of Sum_pept_count	8	11	10	15
		Max of Max_Dif_pept	6	8	6	9
Pgam1	phosphoglycerate mutase 1	Sum of Sum_pept_count	2	8	7	27
		Max of Max_Dif_pept	1	6	4	9
Pfn1	profilin 1	Sum of Sum_pept_count	28	61	44	37
		Max of Max_Dif_pept	6	9	7	9
Pdcd10	programmed cell death 10	Sum of Sum_pept_count	10	14	10	11
		Max of Max_Dif_pept	8	5	5	9
2010106E10Rik	RIKEN cDNA 2010106E10 gene	Sum of Sum_pept_count	57	49	48	62
		Max of Max_Dif_pept	8	9	7	8
Oas1g	2'-5' oligoadenylate synthetase 1G	Sum of Sum_pept_count	9	9	6	14
		Max of Max_Dif_pept	4	9	6	8
Oas1a	2'-5' oligoadenylate synthetase 1A	Sum of Sum_pept_count	9	9	6	14
		Max of Max_Dif_pept	4	9	6	8
Nme2	non-metastatic cells 2, protein (NM23B) expressed in	Sum of Sum_pept_count	0	19	10	9
		Max of Max_Dif_pept	0	9	7	6
Clic1	chloride intracellular channel 1	Sum of Sum_pept_count	15	33	25	14
		Max of Max_Dif_pept	7	8	8	9
Atp6v0d1	ATPase, H+ transporting, lysosomal V0 subunit D1	Sum of Sum_pept_count	16	11	19	11
		Max of Max_Dif_pept	9	7	9	7
1200009I06Rik	RIKEN cDNA 1200009I06 gene	Sum of Sum_pept_count	16	20	10	15
		Max of Max_Dif_pept	5	7	2	9
Myh9	myosin, heavy polypeptide 9, non-muscle	Sum of Sum_pept_count	0	0	0	10
		Max of Max_Dif_pept	0	0	0	9
Ap1m1	adaptor-related protein complex AP-1, mu subunit 1	Sum of Sum_pept_count	0	0	0	11
		Max of Max_Dif_pept	0	0	0	9
Cdc42	cell division cycle 42 homolog (S. cerevisiae)	Sum of Sum_pept_count	51	49	46	47
		Max of Max_Dif_pept	9	8	7	8
Map2k1	mitogen-activated protein kinase kinase 1	Sum of Sum_pept_count	6	2	6	13
		Max of Max_Dif_pept	6	2	6	9
Actr2	ARP2 actin-related protein 2 homolog (yeast)	Sum of Sum_pept_count	3	14	12	10
		Max of Max_Dif_pept	3	7	9	5
Kras	v-Ki-ras2 Kirsten rat sarcoma viral oncogene homolog	Sum of Sum_pept_count	36	26	21	15
		Max of Max_Dif_pept	9	8	8	7
Inpp5a	inositol polyphosphate-5-phosphatase A	Sum of Sum_pept_count	6	8	10	5
		Max of Max_Dif_pept	6	7	9	5
Hbb-b2	hemoglobin, beta adult minor chain	Sum of Sum_pept_count	14	21	19	25
		Max of Max_Dif_pept	6	9	7	9
Hbb-b1	hemoglobin, beta adult major chain	Sum of Sum_pept_count	14	21	19	25
		Max of Max_Dif_pept	6	9	7	9
Cmpk2	cytidine monophosphate (UMP-CMP) kinase 2, mitochondrial	Sum of Sum_pept_count	0	3	0	9
		Max of Max_Dif_pept	0	3	0	9

Gene Symbol	Gene Description	Data	proximal	central	distal	total mucosa
H2-L	histocompatibility 2, D region	Sum of Sum_pept_count	7	2	17	15
		Max of Max_Dif_pept	5	2	9	6
H2-D1	histocompatibility 2, D region locus 1	Sum of Sum_pept_count	7	2	17	15
		Max of Max_Dif_pept	5	2	9	6
Gpx1	glutathione peroxidase 1	Sum of Sum_pept_count	8	14	10	13
		Max of Max_Dif_pept	5	9	6	8
B4galnt2	beta-1,4-N-acetyl-galactosaminyl transferase 2	Sum of Sum_pept_count	1	1	11	4
		Max of Max_Dif_pept	1	1	9	4
Capn1	calpain 1	Sum of Sum_pept_count	0	4	9	5
		Max of Max_Dif_pept	0	4	9	5
Ap2b1	adaptor-related protein complex 2, beta 1 subunit	Sum of Sum_pept_count	6	20	14	9
		Max of Max_Dif_pept	5	9	7	6
Ehd4	EH-domain containing 4	Sum of Sum_pept_count	0	11	8	12
		Max of Max_Dif_pept	0	9	5	9
Ehd3	EH-domain containing 3	Sum of Sum_pept_count	0	12	0	0
		Max of Max_Dif_pept	0	9	0	0
Dnpep	aspartyl aminopeptidase	Sum of Sum_pept_count	2	9	4	15
		Max of Max_Dif_pept	2	7	4	9
Dnaja2	DnaJ (Hsp40) homolog, subfamily A, member 2	Sum of Sum_pept_count	9	10	9	13
		Max of Max_Dif_pept	6	9	7	8
Dars	aspartyl-tRNA synthetase	Sum of Sum_pept_count	0	0	0	9
		Max of Max_Dif_pept	0	0	0	9
Atic	5-aminoimidazole-4-carboxamide ribonucleotide formyltransferase/IMP cyclohydrolase	Sum of Sum_pept_count	0	7	4	19
		Max of Max_Dif_pept	0	7	4	9
Vcp	valosin containing protein	Sum of Sum_pept_count	2	13	10	13
		Max of Max_Dif_pept	2	8	4	5
Coro1b	coronin, actin binding protein 1B	Sum of Sum_pept_count	9	24	14	14
		Max of Max_Dif_pept	4	8	4	6
Vat1	vesicle amine transport protein 1 homolog (T californica)	Sum of Sum_pept_count	13	16	5	15
		Max of Max_Dif_pept	7	7	3	8
Atp1b1	ATPase, Na+/K+ transporting, beta 1 polypeptide	Sum of Sum_pept_count	32	17	33	31
		Max of Max_Dif_pept	7	4	8	7
Ugt1a7c	UDP glucuronosyltransferase 1 family, polypeptide A7C	Sum of Sum_pept_count	5	1	4	8
		Max of Max_Dif_pept	5	1	4	8
Tubb6	tubulin, beta 6	Sum of Sum_pept_count	6	7	4	28
		Max of Max_Dif_pept	3	4	1	8
Adam10	a disintegrin and metallopeptidase domain 10	Sum of Sum_pept_count	1	0	8	6
		Max of Max_Dif_pept	1	0	8	6
Tradd	TNFRSF1A-associated via death domain	Sum of Sum_pept_count	11	6	7	11
		Max of Max_Dif_pept	8	4	5	5
Taldo1	transaldolase 1	Sum of Sum_pept_count	2	16	12	11
		Max of Max_Dif_pept	1	8	5	6
Smpd3	sphingomyelin phosphodiesterase 3, neutral	Sum of Sum_pept_count	37	11	26	17
		Max of Max_Dif_pept	8	4	7	4
Slc28a2	solute carrier family 28 (sodium-coupled nucleoside transporter), member 2	Sum of Sum_pept_count	19	11	8	9
		Max of Max_Dif_pept	8	6	5	5
Slc28a1	solute carrier family 28 (sodium-coupled nucleoside transporter), member 1	Sum of Sum_pept_count	26	4	0	12
		Max of Max_Dif_pept	8	3	0	5
Slc26a3	solute carrier family 26, member 3	Sum of Sum_pept_count	12	0	30	10
		Max of Max_Dif_pept	6	0	8	5
Atp5a1	ATP synthase, H+ transporting, mitochondrial F1 complex, alpha subunit, isoform 1	Sum of Sum_pept_count	0	0	1	17
		Max of Max_Dif_pept	0	0	1	8
Serpinb6a	serine (or cysteine) peptidase inhibitor, clade B, member 6a	Sum of Sum_pept_count	7	13	11	17
		Max of Max_Dif_pept	5	8	6	8
Sec23b	SEC23B (S. cerevisiae)	Sum of Sum_pept_count	0	3	2	16
		Max of Max_Dif_pept	0	2	2	8
9130017N09Rik	RIKEN cDNA 9130017N09 gene	Sum of Sum_pept_count	19	28	38	53
		Max of Max_Dif_pept	6	6	8	6
Rras2	related RAS viral (r-ras) oncogene homolog 2	Sum of Sum_pept_count	19	11	19	12
		Max of Max_Dif_pept	8	6	7	7
Rpn2	ribophorin II	Sum of Sum_pept_count	0	0	9	1
		Max of Max_Dif_pept	0	0	8	1

Gene Symbol	Gene Description	Data	proximal	central	distal	total mucosa
Rhpn2	rhophilin, Rho GTPase binding protein 2	Sum of Sum_pept_count	3	10	10	8
		Max of Max_Dif_pept	3	8	7	7
Rdh7	retinol dehydrogenase 7	Sum of Sum_pept_count	0	0	0	9
		Max of Max_Dif_pept	0	0	0	8
Atp5f1	ATP synthase, H+ transporting, mitochondrial F0 complex, subunit b, isoform 1	Sum of Sum_pept_count	0	0	4	15
		Max of Max_Dif_pept	0	0	3	8
Ralb	v-ral simian leukemia viral oncogene homolog B (ras related)	Sum of Sum_pept_count	38	14	24	14
		Max of Max_Dif_pept	8	6	8	8
Bsg	basigin	Sum of Sum_pept_count	27	10	29	22
		Max of Max_Dif_pept	8	4	8	8
Rab5b	RAB5B, member RAS oncogene family	Sum of Sum_pept_count	16	7	18	12
		Max of Max_Dif_pept	8	4	7	7
Rab5a	RAB5A, member RAS oncogene family	Sum of Sum_pept_count	28	5	33	24
		Max of Max_Dif_pept	7	4	6	8
Rab3d	RAB3D, member RAS oncogene family	Sum of Sum_pept_count	11	8	15	16
		Max of Max_Dif_pept	4	4	7	8
Cct4	chaperonin containing Tcp1, subunit 4 (delta)	Sum of Sum_pept_count	0	3	5	11
		Max of Max_Dif_pept	0	3	5	8
Ctnnd1	catenin (cadherin associated protein), delta 1	Sum of Sum_pept_count	4	2	17	7
		Max of Max_Dif_pept	3	1	8	2
Rab25	RAB25, member RAS oncogene family	Sum of Sum_pept_count	14	3	20	14
		Max of Max_Dif_pept	7	3	8	7
Rab22a	RAB22A, member RAS oncogene family	Sum of Sum_pept_count	17	8	9	10
		Max of Max_Dif_pept	8	6	4	7
Rab18	RAB18, member RAS oncogene family	Sum of Sum_pept_count	16	9	15	15
		Max of Max_Dif_pept	8	6	7	6
Psmc6	proteasome (prosome, macropain) 26S subunit, ATPase, 6	Sum of Sum_pept_count	1	1	3	13
		Max of Max_Dif_pept	1	1	2	8
Psmc4	proteasome (prosome, macropain) 26S subunit, ATPase, 4	Sum of Sum_pept_count	0	0	0	13
		Max of Max_Dif_pept	0	0	0	8
Prss7	protease, serine, 7 (enterokinase)	Sum of Sum_pept_count	48	1	0	12
		Max of Max_Dif_pept	8	1	0	4
Aldh16a1	aldehyde dehydrogenase 16 family, member A1	Sum of Sum_pept_count	0	2	2	17
		Max of Max_Dif_pept	0	2	2	8
Prkag2	protein kinase, AMP-activated, gamma 2 non-catalytic subunit	Sum of Sum_pept_count	1	9	3	10
		Max of Max_Dif_pept	1	8	3	7
Cd36	CD36 antigen	Sum of Sum_pept_count	14	10	6	12
		Max of Max_Dif_pept	8	8	5	7
Ppap2a	phosphatidic acid phosphatase 2a	Sum of Sum_pept_count	112	42	24	34
		Max of Max_Dif_pept	8	5	3	5
Ap1g1	adaptor protein complex AP-1, gamma 1 subunit	Sum of Sum_pept_count	0	0	1	15
		Max of Max_Dif_pept	0	0	1	8
Pdia3	protein disulfide isomerase associated 3	Sum of Sum_pept_count	0	0	0	8
		Max of Max_Dif_pept	0	0	0	8
Pck1	phosphoenolpyruvate carboxykinase 1, cytosolic	Sum of Sum_pept_count	1	0	1	8
		Max of Max_Dif_pept	1	0	1	8
Papss1	3'-phosphoadenosine 5'-phosphosulfate synthase 1	Sum of Sum_pept_count	3	2	1	14
		Max of Max_Dif_pept	1	2	1	8
Nsf	N-ethylmaleimide sensitive fusion protein	Sum of Sum_pept_count	3	0	5	8
		Max of Max_Dif_pept	2	0	3	8
Cyp3a13	cytochrome P450, family 3, subfamily a, polypeptide 13	Sum of Sum_pept_count	12	1	8	19
		Max of Max_Dif_pept	5	1	6	8
Nras	neuroblastoma ras oncogene	Sum of Sum_pept_count	7	21	7	10
		Max of Max_Dif_pept	3	8	4	8

Gene Symbol	Gene Description	Data	proximal	central	distal	total mucosa
Clic4	chloride intracellular channel 4 (mitochondrial)	Sum of Sum_pept_count	2	10	4	7
		Max of Max_Dif_pept	2	8	4	6
D1Pas1	DNA segment, Chr 1, Pasteur Institute 1	Sum of Sum_pept_count	8	11	32	10
		Max of Max_Dif_pept	3	2	8	4
Ncstn	nicastrin	Sum of Sum_pept_count	8	7	9	9
		Max of Max_Dif_pept	7	7	8	5
Nat2	N-acetyltransferase 2 (arylamine N-acetyltransferase)	Sum of Sum_pept_count	2	2	1	8
		Max of Max_Dif_pept	2	2	1	8
Mvp	major vault protein	Sum of Sum_pept_count	0	0	2	20
		Max of Max_Dif_pept	0	0	2	8
Muc3	mucin 3, intestinal	Sum of Sum_pept_count	14	13	32	20
		Max of Max_Dif_pept	7	5	8	6
Arf6	ADP-ribosylation factor 6	Sum of Sum_pept_count	36	27	42	33
		Max of Max_Dif_pept	8	7	8	8
Msra	methionine sulfoxide reductase A	Sum of Sum_pept_count	3	2	3	14
		Max of Max_Dif_pept	3	1	2	8
Mical1	microtubule associated monoxygenase, calponin and LIM domain containing 1	Sum of Sum_pept_count	3	16	5	9
		Max of Max_Dif_pept	2	8	3	3
Mdh1	malate dehydrogenase 1, NAD (soluble)	Sum of Sum_pept_count	8	18	13	19
		Max of Max_Dif_pept	4	8	6	7
Mapk13	mitogen-activated protein kinase 13	Sum of Sum_pept_count	6	14	12	13
		Max of Max_Dif_pept	3	6	5	8
LOC628409	similar to cytoplasmic beta-actin	Sum of Sum_pept_count	15	0	22	0
		Max of Max_Dif_pept	8	0	7	0
Lgals9	lectin, galactose binding, soluble 9	Sum of Sum_pept_count	44	27	41	29
		Max of Max_Dif_pept	8	8	8	6
Krt8	keratin 8	Sum of Sum_pept_count	3	12	27	60
		Max of Max_Dif_pept	2	4	8	8
Krt42	keratin 42	Sum of Sum_pept_count	4	5	20	12
		Max of Max_Dif_pept	2	2	8	5
Igf2r	insulin-like growth factor 2 receptor	Sum of Sum_pept_count	0	0	0	8
		Max of Max_Dif_pept	0	0	0	8
Ceacam1	carcinoembryonic antigen-related cell adhesion molecule 1	Sum of Sum_pept_count	27	25	67	30
		Max of Max_Dif_pept	5	4	8	4
Cmpk1	cytidine monophosphate (UMP-CMP) kinase 1	Sum of Sum_pept_count	8	9	10	10
		Max of Max_Dif_pept	8	8	7	7
Gstm3	glutathione S-transferase, mu 3	Sum of Sum_pept_count	6	1	1	8
		Max of Max_Dif_pept	6	1	1	8
Dera	2-deoxyribose-5-phosphate aldolase homolog (C. elegans)	Sum of Sum_pept_count	1	3	2	11
		Max of Max_Dif_pept	1	3	2	8
Gsr	glutathione reductase	Sum of Sum_pept_count	0	7	7	8
		Max of Max_Dif_pept	0	7	7	8
Golph3	golgi phosphoprotein 3	Sum of Sum_pept_count	13	15	10	7
		Max of Max_Dif_pept	6	8	5	6
Gna14	guanine nucleotide binding protein, alpha 14	Sum of Sum_pept_count	26	31	25	26
		Max of Max_Dif_pept	7	7	8	6
Gale	galactose-4-epimerase, UDP	Sum of Sum_pept_count	1	1	2	13
		Max of Max_Dif_pept	1	1	2	8
Dhrs1	dehydrogenase/reductase (SDR family) member 1	Sum of Sum_pept_count	9	4	1	9
		Max of Max_Dif_pept	8	4	1	8
Flnb	filamin, beta	Sum of Sum_pept_count	0	1	4	11
		Max of Max_Dif_pept	0	1	3	8
Fkbp4	FK506 binding protein 4	Sum of Sum_pept_count	0	1	2	15
		Max of Max_Dif_pept	0	1	2	8
Fabp1	fatty acid binding protein 1, liver	Sum of Sum_pept_count	18	39	5	26
		Max of Max_Dif_pept	5	8	2	8
EG546101	predicted gene, EG546101	Sum of Sum_pept_count	33	3	37	5
		Max of Max_Dif_pept	7	3	8	5
Efr3a	EFR3 homolog A (S. cerevisiae)	Sum of Sum_pept_count	11	9	13	2
		Max of Max_Dif_pept	6	8	7	1
Efhd2	EF hand domain containing 2	Sum of Sum_pept_count	1	25	12	21
		Max of Max_Dif_pept	1	8	4	6
Sept7	septin 7	Sum of Sum_pept_count	0	3	3	12
		Max of Max_Dif_pept	0	2	2	7
Acox1	acyl-Coenzyme A oxidase 1, palmitoyl	Sum of Sum_pept_count	0	0	0	13
		Max of Max_Dif_pept	0	0	0	7

Gene Symbol	Gene Description	Data	proximal	central	distal	total mucosa
Vti1b	vesicle transport through interaction with t-SNAREs 1B homolog	Sum of Sum_pept_count	8	4	10	9
		Max of Max_Dif_pept	4	3	7	3
Apoa1	apolipoprotein A-I	Sum of Sum_pept_count	3	0	4	9
		Max of Max_Dif_pept	3	0	4	7
Vdac2	voltage-dependent anion channel 2	Sum of Sum_pept_count	3	1	9	11
		Max of Max_Dif_pept	3	1	6	7
Vdac1	voltage-dependent anion channel 1	Sum of Sum_pept_count	4	2	6	9
		Max of Max_Dif_pept	4	2	6	7
Coro2a	coronin, actin binding protein 2A	Sum of Sum_pept_count	0	8	5	13
		Max of Max_Dif_pept	0	5	4	7
Vasp	vasodilator-stimulated phosphoprotein	Sum of Sum_pept_count	3	7	1	5
		Max of Max_Dif_pept	3	7	1	4
Upp1	uridine phosphorylase 1	Sum of Sum_pept_count	3	10	12	7
		Max of Max_Dif_pept	2	7	7	4
2010003H20Rik	RIKEN cDNA 2010003H20 gene	Sum of Sum_pept_count	127	81	130	89
		Max of Max_Dif_pept	7	7	7	7
Txn1	thioredoxin 1	Sum of Sum_pept_count	9	23	14	15
		Max of Max_Dif_pept	3	7	5	7
Arl14	ADP-ribosylation factor-like 14	Sum of Sum_pept_count	8	4	1	3
		Max of Max_Dif_pept	7	4	1	3
Tubb2b	tubulin, beta 2b	Sum of Sum_pept_count	14	4	12	30
		Max of Max_Dif_pept	5	1	3	7
Tsta3	tissue specific transplantation antigen P35B	Sum of Sum_pept_count	0	1	2	10
		Max of Max_Dif_pept	0	1	1	7
Arl1	ADP-ribosylation factor-like 1	Sum of Sum_pept_count	3	8	6	15
		Max of Max_Dif_pept	3	6	5	7
Tspan8	tetraspanin 8	Sum of Sum_pept_count	38	33	59	41
		Max of Max_Dif_pept	5	5	6	7
2400003C14Rik	RIKEN cDNA 2400003C14 gene	Sum of Sum_pept_count	16	23	17	21
		Max of Max_Dif_pept	5	7	6	4
Tpm3	tropomyosin 3, gamma	Sum of Sum_pept_count	1	2	3	8
		Max of Max_Dif_pept	1	2	2	7
Tmigd1	transmembrane and immunoglobulin domain containing 1	Sum of Sum_pept_count	0	6	24	5
		Max of Max_Dif_pept	0	4	7	3
Tmem30b	transmembrane protein 30B	Sum of Sum_pept_count	38	34	40	27
		Max of Max_Dif_pept	6	6	7	6
Tmem30a	transmembrane protein 30A	Sum of Sum_pept_count	6	7	17	5
		Max of Max_Dif_pept	3	4	7	3
Casp6	caspase 6	Sum of Sum_pept_count	4	6	1	18
		Max of Max_Dif_pept	2	3	1	7
Tcp1	t-complex protein 1	Sum of Sum_pept_count	0	2	2	8
		Max of Max_Dif_pept	0	1	1	7
Coro1a	coronin, actin binding protein 1A	Sum of Sum_pept_count	1	7	5	3
		Max of Max_Dif_pept	1	7	4	2
Stx12	syntaxin 12	Sum of Sum_pept_count	0	0	4	13
		Max of Max_Dif_pept	0	0	3	7
Stat1	signal transducer and activator of transcription 1	Sum of Sum_pept_count	2	2	5	8
		Max of Max_Dif_pept	2	2	3	7
Akr1b7	aldo-keto reductase family 1, member B7	Sum of Sum_pept_count	0	4	0	9
		Max of Max_Dif_pept	0	4	0	7
Slc6a14	solute carrier family 6 (neurotransmitter transporter), member 14	Sum of Sum_pept_count	0	0	9	2
		Max of Max_Dif_pept	0	0	7	2
Slc5a12	solute carrier family 5 (sodium/glucose cotransporter), member 12	Sum of Sum_pept_count	18	16	7	15
		Max of Max_Dif_pept	6	6	4	7
Slc2a5	solute carrier family 2 (facilitated glucose transporter), member 5	Sum of Sum_pept_count	28	8	3	13
		Max of Max_Dif_pept	7	4	2	5
Akr1c13	aldo-keto reductase family 1, member C13	Sum of Sum_pept_count	3	7	3	0
		Max of Max_Dif_pept	3	7	3	0
Slc25a5	solute carrier family 25 (mitochondrial carrier, adenine nucleotide translocator), member 5	Sum of Sum_pept_count	2	2	13	25
		Max of Max_Dif_pept	1	1	4	7
1810010M01Rik	RIKEN cDNA 1810010M01 gene	Sum of Sum_pept_count	13	10	33	28
		Max of Max_Dif_pept	6	5	7	7
Shmt1	serine hydroxymethyltransferase 1 (soluble)	Sum of Sum_pept_count	1	2	2	10
		Max of Max_Dif_pept	1	1	1	7

Gene Symbol	Gene Description	Data	proximal	central	distal	total mucosa
Sec13	SEC13 homolog (S. cerevisiae)	Sum of Sum_pept_count	1	1	1	10
		Max of Max_Dif_pept	1	1	1	7
Rps8	ribosomal protein S8	Sum of Sum_pept_count	6	0	7	8
		Max of Max_Dif_pept	6	0	7	7
Rhod	ras homolog gene family, member D	Sum of Sum_pept_count	15	7	13	10
		Max of Max_Dif_pept	7	4	7	6
Cct3	chaperonin containing Tcp1, subunit 3 (gamma)	Sum of Sum_pept_count	0	0	3	7
		Max of Max_Dif_pept	0	0	2	7
Rdh9	retinol dehydrogenase 9	Sum of Sum_pept_count	0	0	0	9
		Max of Max_Dif_pept	0	0	0	7
Rbp2	retinol binding protein 2, cellular	Sum of Sum_pept_count	12	21	6	19
		Max of Max_Dif_pept	5	7	2	7
Rap2b	RAP2B, member of RAS oncogene family	Sum of Sum_pept_count	0	11	17	10
		Max of Max_Dif_pept	0	7	7	7
Araf	v-raf murine sarcoma 3611 viral oncogene homolog	Sum of Sum_pept_count	15	4	7	9
		Max of Max_Dif_pept	5	2	3	7
Ran	RAN, member RAS oncogene family	Sum of Sum_pept_count	6	9	10	20
		Max of Max_Dif_pept	5	4	4	7
Rab6	RAB6, member RAS oncogene family	Sum of Sum_pept_count	38	17	35	40
		Max of Max_Dif_pept	7	6	7	7
Arl8a	ADP-ribosylation factor-like 8A	Sum of Sum_pept_count	9	6	8	7
		Max of Max_Dif_pept	7	5	5	6
Arl8b	ADP-ribosylation factor-like 8B	Sum of Sum_pept_count	8	8	11	8
		Max of Max_Dif_pept	5	6	7	6
Rab27b	RAB27b, member RAS oncogene family	Sum of Sum_pept_count	6	1	7	6
		Max of Max_Dif_pept	5	1	7	6
Rab17	RAB17, member RAS oncogene family	Sum of Sum_pept_count	9	8	12	10
		Max of Max_Dif_pept	7	7	7	7
Pycard	PYD and CARD domain containing	Sum of Sum_pept_count	2	8	4	8
		Max of Max_Dif_pept	2	7	3	6
2210404O07Rik	RIKEN cDNA 2210404O07 gene	Sum of Sum_pept_count	187	155	150	161
		Max of Max_Dif_pept	6	6	7	7
Cyb5r3	cytochrome b5 reductase 3	Sum of Sum_pept_count	4	3	4	10
		Max of Max_Dif_pept	4	3	4	7
Prkacb	protein kinase, cAMP dependent, catalytic, beta	Sum of Sum_pept_count	0	0	0	8
		Max of Max_Dif_pept	0	0	0	7
Prkaa1	protein kinase, AMP-activated, alpha 1 catalytic subunit	Sum of Sum_pept_count	0	4	2	8
		Max of Max_Dif_pept	0	4	2	7
Pnp1	purine-nucleoside phosphorylase 1	Sum of Sum_pept_count	3	13	11	13
		Max of Max_Dif_pept	2	6	5	7
Clca6	chloride channel calcium activated 6	Sum of Sum_pept_count	33	31	46	42
		Max of Max_Dif_pept	6	5	6	7
Pdcd6	programmed cell death 6	Sum of Sum_pept_count	3	11	4	6
		Max of Max_Dif_pept	2	7	4	3
P2rx4	purinergic receptor P2X, ligand-gated ion channel 4	Sum of Sum_pept_count	7	6	13	8
		Max of Max_Dif_pept	7	6	7	7
Numb	numb gene homolog (Drosophila)	Sum of Sum_pept_count	16	13	20	7
		Max of Max_Dif_pept	5	7	7	2
D10Bwg1364e	DNA segment, Chr 10, Brigham & Women's Genetics 1364 expressed	Sum of Sum_pept_count	28	11	10	9
		Max of Max_Dif_pept	7	4	4	6
Nme1	non-metastatic cells 1, protein (NM23A) expressed in	Sum of Sum_pept_count	5	10	9	6
		Max of Max_Dif_pept	5	7	6	6
Aldh2	aldehyde dehydrogenase 2, mitochondrial	Sum of Sum_pept_count	0	0	0	13
		Max of Max_Dif_pept	0	0	0	7
Naprt1	nicotinate phosphoribosyltransferase domain containing 1	Sum of Sum_pept_count	0	3	1	7
		Max of Max_Dif_pept	0	3	1	7
Nampt	nicotinamide phosphoribosyltransferase	Sum of Sum_pept_count	1	2	0	16
		Max of Max_Dif_pept	1	2	0	7
Map2k2	mitogen-activated protein kinase kinase 2	Sum of Sum_pept_count	9	7	7	13
		Max of Max_Dif_pept	6	4	7	7

Gene Symbol	Gene Description	Data	proximal	central	distal	total mucosa
Maoa	monoamine oxidase A	Sum of Sum_pept_count	1	0	4	10
		Max of Max_Dif_pept	1	0	4	7
Lrrc57	leucine rich repeat containing 57	Sum of Sum_pept_count	6	4	8	4
		Max of Max_Dif_pept	5	4	7	4
Lman2	lectin, mannose-binding 2	Sum of Sum_pept_count	1	0	5	7
		Max of Max_Dif_pept	1	0	5	7
Akp5	alkaline phosphatase 5	Sum of Sum_pept_count	8	10	14	14
		Max of Max_Dif_pept	7	5	6	6
Krt86	keratin 86	Sum of Sum_pept_count	5	8	4	0
		Max of Max_Dif_pept	4	7	4	0
Krt33a	keratin 33A	Sum of Sum_pept_count	25	13	2	0
		Max of Max_Dif_pept	7	5	2	0
Krt14	keratin 14	Sum of Sum_pept_count	9	10	9	13
		Max of Max_Dif_pept	7	6	3	5
Khk	ketohexokinase	Sum of Sum_pept_count	2	10	4	11
		Max of Max_Dif_pept	2	7	4	5
Iqgap1	IQ motif containing GTPase activating protein 1	Sum of Sum_pept_count	1	14	8	8
		Max of Max_Dif_pept	1	5	5	7
Ifi47	interferon gamma inducible protein 47	Sum of Sum_pept_count	6	5	7	8
		Max of Max_Dif_pept	6	4	6	7
Hras1	Harvey rat sarcoma virus oncogene 1	Sum of Sum_pept_count	13	25	15	8
		Max of Max_Dif_pept	3	7	4	2
Gstp1	glutathione S-transferase, pi 1	Sum of Sum_pept_count	4	12	7	10
		Max of Max_Dif_pept	3	6	5	7
Gsto1	glutathione S-transferase omega 1	Sum of Sum_pept_count	5	9	11	12
		Max of Max_Dif_pept	3	3	4	7
Gpt	glutamic pyruvic transaminase, soluble	Sum of Sum_pept_count	4	8	3	8
		Max of Max_Dif_pept	4	7	3	7
Gpa33	glycoprotein A33 (transmembrane)	Sum of Sum_pept_count	16	11	22	27
		Max of Max_Dif_pept	7	5	7	7
Golga7	golgi autoantigen, golgin subfamily a, 7	Sum of Sum_pept_count	12	7	10	9
		Max of Max_Dif_pept	4	5	7	4
Gnb2l1	guanine nucleotide binding protein (G protein), beta polypeptide 2 like 1	Sum of Sum_pept_count	1	0	0	11
		Max of Max_Dif_pept	1	0	0	7
2810405K02Rik	RIKEN cDNA 2810405K02 gene	Sum of Sum_pept_count	5	11	8	23
		Max of Max_Dif_pept	5	7	5	7
2210417D09Rik	RIKEN cDNA 2210417D09 gene	Sum of Sum_pept_count	3	4	12	6
		Max of Max_Dif_pept	2	3	7	5
Fdps	farnesyl diphosphate synthetase	Sum of Sum_pept_count	9	16	13	8
		Max of Max_Dif_pept	1	4	7	5
4931406C07Rik	RIKEN cDNA 4931406C07 gene	Sum of Sum_pept_count	4	8	5	5
		Max of Max_Dif_pept	3	7	4	4
Adh5	alcohol dehydrogenase 5 (class III), chi polypeptide	Sum of Sum_pept_count	0	0	0	6
		Max of Max_Dif_pept	0	0	0	6
Wars	tryptophanyl-tRNA synthetase	Sum of Sum_pept_count	0	3	3	14
		Max of Max_Dif_pept	0	3	2	6
Cotl1	coactosin-like 1 (Dictyostelium)	Sum of Sum_pept_count	3	8	9	8
		Max of Max_Dif_pept	2	3	5	6
Ugp2	UDP-glucose pyrophosphorylase 2	Sum of Sum_pept_count	0	6	2	6
		Max of Max_Dif_pept	0	5	2	6
Txnrd1	thioredoxin reductase 1	Sum of Sum_pept_count	1	8	4	3
		Max of Max_Dif_pept	1	6	3	2
Txndc5	thioredoxin domain containing 5	Sum of Sum_pept_count	0	0	0	9
		Max of Max_Dif_pept	0	0	0	6
Tubal3	tubulin, alpha-like 3	Sum of Sum_pept_count	0	6	0	8
		Max of Max_Dif_pept	0	5	0	6
Ttc7	tetratricopeptide repeat domain 7	Sum of Sum_pept_count	3	9	3	1
		Max of Max_Dif_pept	3	6	2	1
0610010K06Rik	RIKEN cDNA 0610010K06 gene	Sum of Sum_pept_count	5	7	1	6
		Max of Max_Dif_pept	4	6	1	5
Chmp2a	chromatin modifying protein 2A	Sum of Sum_pept_count	13	12	12	13
		Max of Max_Dif_pept	3	3	4	6
Trf	transferrin	Sum of Sum_pept_count	0	6	0	1
		Max of Max_Dif_pept	0	6	0	1
1110038D17Rik	RIKEN cDNA 1110038D17 gene	Sum of Sum_pept_count	15	0	0	5
		Max of Max_Dif_pept	6	0	0	3
Tm9sf2	transmembrane 9 superfamily member 2	Sum of Sum_pept_count	6	2	16	9
		Max of Max_Dif_pept	4	1	6	4
Tars	threonyl-tRNA synthetase	Sum of Sum_pept_count	0	3	3	7
		Max of Max_Dif_pept	0	3	3	6

Gene Symbol	Gene Description	Data	proximal	central	distal	total mucosa
Casp7	caspase 7	Sum of Sum_pept_count	4	10	7	6
		Max of Max_Dif_pept	4	6	5	5
Arl2	ADP-ribosylation factor-like 2	Sum of Sum_pept_count	0	1	2	6
		Max of Max_Dif_pept	0	1	2	6
Sord	sorbitol dehydrogenase	Sum of Sum_pept_count	0	7	4	6
		Max of Max_Dif_pept	0	6	3	5
Sod1	superoxide dismutase 1, soluble	Sum of Sum_pept_count	3	6	5	4
		Max of Max_Dif_pept	3	6	4	4
Slc6a20a	solute carrier family 6 (neurotransmitter transporter), member 20A	Sum of Sum_pept_count	21	11	11	15
		Max of Max_Dif_pept	6	5	6	5
Slc5a9	solute carrier family 5 (sodium/glucose cotransporter), member 9	Sum of Sum_pept_count	23	20	19	16
		Max of Max_Dif_pept	6	6	6	6
Slc5a8	solute carrier family 5 (iodide transporter), member 8	Sum of Sum_pept_count	3	34	56	26
		Max of Max_Dif_pept	1	5	6	6
Slc25a3	solute carrier family 25 (mitochondrial carrier, phosphate carrier), member 3	Sum of Sum_pept_count	0	0	1	10
		Max of Max_Dif_pept	0	0	1	6
Slc23a1	solute carrier family 23 (nucleobase transporters), member 1	Sum of Sum_pept_count	13	14	5	10
		Max of Max_Dif_pept	6	5	2	4
Slc10a2	solute carrier family 10, member 2	Sum of Sum_pept_count	0	3	35	15
		Max of Max_Dif_pept	0	2	6	5
Sh3bgrl3	SH3 domain binding glutamic acid-rich protein-like 3	Sum of Sum_pept_count	4	14	9	10
		Max of Max_Dif_pept	3	6	3	5
Cryl1	crystallin, lambda 1	Sum of Sum_pept_count	1	7	3	7
		Max of Max_Dif_pept	1	6	3	6
Sar1b	SAR1 gene homolog B (S. cerevisiae)	Sum of Sum_pept_count	2	6	3	6
		Max of Max_Dif_pept	2	6	2	5
Samhd1	SAM domain and HD domain, 1	Sum of Sum_pept_count	0	0	0	7
		Max of Max_Dif_pept	0	0	0	6
Rps3	ribosomal protein S3	Sum of Sum_pept_count	5	2	4	7
		Max of Max_Dif_pept	3	2	4	6
Rps19	ribosomal protein S19	Sum of Sum_pept_count	10	5	9	2
		Max of Max_Dif_pept	6	4	5	2
Rps18	ribosomal protein S18	Sum of Sum_pept_count	11	5	14	13
		Max of Max_Dif_pept	5	4	5	6
Rhof	ras homolog gene family, member f	Sum of Sum_pept_count	10	8	18	7
		Max of Max_Dif_pept	5	5	6	4
Capza1	capping protein (actin filament) muscle Z-line, alpha 1	Sum of Sum_pept_count	0	10	4	11
		Max of Max_Dif_pept	0	4	2	6
BC022224	cDNA sequence BC022224	Sum of Sum_pept_count	2	2	4	9
		Max of Max_Dif_pept	1	2	2	6
Rabggta	Rab geranylgeranyl transferase, a subunit	Sum of Sum_pept_count	0	0	0	6
		Max of Max_Dif_pept	0	0	0	6
Akr1c19	aldo-keto reductase family 1, member C19	Sum of Sum_pept_count	0	0	2	8
		Max of Max_Dif_pept	0	0	2	6
Rab4b	RAB4B, member RAS oncogene family	Sum of Sum_pept_count	7	5	0	14
		Max of Max_Dif_pept	6	5	0	6
Ckmt1	creatine kinase, mitochondrial 1, ubiquitous	Sum of Sum_pept_count	0	5	6	14
		Max of Max_Dif_pept	0	3	3	6
Ctnna1	catenin (cadherin associated protein), alpha 1	Sum of Sum_pept_count	0	0	8	1
		Max of Max_Dif_pept	0	0	6	1
Rab33b	RAB33B, member of RAS oncogene family	Sum of Sum_pept_count	10	8	7	6
		Max of Max_Dif_pept	6	3	4	4
Rab30	RAB30, member RAS oncogene family	Sum of Sum_pept_count	10	0	0	0
		Max of Max_Dif_pept	6	0	0	0
Rab13	RAB13, member RAS oncogene family	Sum of Sum_pept_count	15	0	12	1
		Max of Max_Dif_pept	3	0	6	1
Psme2	proteasome (prosome, macropain) 28 subunit, beta	Sum of Sum_pept_count	2	5	6	14
		Max of Max_Dif_pept	2	4	3	6

Gene Symbol	Gene Description	Data	proximal	central	distal	total mucosa
Psmd6	proteasome (prosome, macropain) 26S subunit, non-ATPase, 6	Sum of Sum_pept_count	0	0	0	6
		Max of Max_Dif_pept	0	0	0	6
Psmd3	proteasome (prosome, macropain) 26S subunit, non-ATPase, 3	Sum of Sum_pept_count	0	0	0	8
		Max of Max_Dif_pept	0	0	0	6
Cct6a	chaperonin containing Tcp1, subunit 6a (zeta)	Sum of Sum_pept_count	1	3	5	10
		Max of Max_Dif_pept	1	3	4	6
Arf2	ADP-ribosylation factor 2	Sum of Sum_pept_count	6	12	9	11
		Max of Max_Dif_pept	5	6	5	4
Psma5	proteasome (prosome, macropain) subunit, alpha type 5	Sum of Sum_pept_count	0	0	6	7
		Max of Max_Dif_pept	0	0	6	6
Prdx2	peroxiredoxin 2	Sum of Sum_pept_count	6	6	10	12
		Max of Max_Dif_pept	2	4	4	6
Plscr1	phospholipid scramblase 1	Sum of Sum_pept_count	35	17	32	25
		Max of Max_Dif_pept	6	4	5	4
Plb1	phospholipase B1	Sum of Sum_pept_count	0	18	29	14
		Max of Max_Dif_pept	0	6	6	4
2610018G03Rik	RIKEN cDNA 2610018G03 gene	Sum of Sum_pept_count	2	13	11	7
		Max of Max_Dif_pept	1	6	4	1
Cyp2b10	cytochrome P450, family 2, subfamily b, polypeptide 10	Sum of Sum_pept_count	2	0	0	6
		Max of Max_Dif_pept	2	0	0	6
Pitpna	phosphatidylinositol transfer protein, alpha	Sum of Sum_pept_count	2	7	6	7
		Max of Max_Dif_pept	2	6	6	5
Phb2	prohibitin 2	Sum of Sum_pept_count	0	0	3	6
		Max of Max_Dif_pept	0	0	2	6
Pgls	6-phosphogluconolactonase	Sum of Sum_pept_count	1	1	1	9
		Max of Max_Dif_pept	1	1	1	6
Pdlim1	PDZ and LIM domain 1 (elfin)	Sum of Sum_pept_count	3	15	6	20
		Max of Max_Dif_pept	1	6	2	5
Pde9a	phosphodiesterase 9A	Sum of Sum_pept_count	0	8	8	3
		Max of Max_Dif_pept	0	6	5	3
Pcyt2	phosphate cytidylyltransferase 2, ethanolamine	Sum of Sum_pept_count	1	3	1	6
		Max of Max_Dif_pept	1	3	1	6
Pcbp1	poly(rC) binding protein 1	Sum of Sum_pept_count	15	11	15	22
		Max of Max_Dif_pept	5	4	5	6
Cyp2d26	cytochrome P450, family 2, subfamily d, polypeptide 26	Sum of Sum_pept_count	2	0	1	6
		Max of Max_Dif_pept	2	0	1	6
Pacsin2	protein kinase C and casein kinase substrate in neurons 2	Sum of Sum_pept_count	4	17	8	10
		Max of Max_Dif_pept	2	6	6	6
Ostf1	osteoclast stimulating factor 1	Sum of Sum_pept_count	4	7	9	11
		Max of Max_Dif_pept	3	4	6	6
Glod5	glyoxalase domain containing 5	Sum of Sum_pept_count	3	11	5	9
		Max of Max_Dif_pept	2	5	4	6
Glrx	glutaredoxin	Sum of Sum_pept_count	6	19	19	11
		Max of Max_Dif_pept	4	6	5	6
Ndufb10	NADH dehydrogenase (ubiquinone) 1 beta subcomplex, 10	Sum of Sum_pept_count	0	0	3	7
		Max of Max_Dif_pept	0	0	3	6
Acot11	acyl-CoA thioesterase 11	Sum of Sum_pept_count	1	1	0	7
		Max of Max_Dif_pept	1	1	0	6
Gmds	GDP-mannose 4, 6-dehydratase	Sum of Sum_pept_count	2	7	10	10
		Max of Max_Dif_pept	2	4	6	5
Ddost	dolichyl-di-phosphooligosaccharide-protein glycotransferase	Sum of Sum_pept_count	2	0	6	3
		Max of Max_Dif_pept	2	0	6	3
Mapbpip	mitogen-activated protein binding protein interacting protein	Sum of Sum_pept_count	4	10	15	5
		Max of Max_Dif_pept	2	4	6	3
Lipe	lipase, hormone sensitive	Sum of Sum_pept_count	15	12	4	6
		Max of Max_Dif_pept	6	5	3	2
Lgals8	lectin, galactose binding, soluble 8	Sum of Sum_pept_count	8	2	5	0
		Max of Max_Dif_pept	6	1	3	0
BC026682	cDNA sequence BC026682	Sum of Sum_pept_count	0	18	25	8
		Max of Max_Dif_pept	0	6	5	3
Krt77	keratin 77	Sum of Sum_pept_count	15	0	16	8
		Max of Max_Dif_pept	6	0	6	6
Aco2	aconitase 2, mitochondrial	Sum of Sum_pept_count	0	0	0	8
		Max of Max_Dif_pept	0	0	0	6

Gene Symbol	Gene Description	Data	proximal	central	distal	total mucosa
Krt76	keratin 76	Sum of Sum_pept_count	36	11	36	32
		Max of Max_Dif_pept	6	4	6	4
Krt75	keratin 75	Sum of Sum_pept_count	12	8	19	12
		Max of Max_Dif_pept	6	4	6	6
Cap1	CAP, adenylate cyclase-associated protein 1 (yeast)	Sum of Sum_pept_count	0	6	6	8
		Max of Max_Dif_pept	0	5	6	6
Krt72	keratin 72	Sum of Sum_pept_count	7	0	2	15
		Max of Max_Dif_pept	4	0	2	6
Krt34	keratin 34	Sum of Sum_pept_count	13	7	6	0
		Max of Max_Dif_pept	6	5	4	0
Krt17	keratin 17	Sum of Sum_pept_count	16	4	34	17
		Max of Max_Dif_pept	4	2	6	4
Krt13	keratin 13	Sum of Sum_pept_count	13	14	17	7
		Max of Max_Dif_pept	4	5	6	3
Apoa4	apolipoprotein A-IV	Sum of Sum_pept_count	9	1	4	12
		Max of Max_Dif_pept	5	1	4	6
Itfg1	integrin alpha FG-GAP repeat containing 1	Sum of Sum_pept_count	7	6	12	6
		Max of Max_Dif_pept	5	4	6	5
Atp9a	ATPase, class II, type 9A	Sum of Sum_pept_count	13	2	15	1
		Max of Max_Dif_pept	6	2	5	1
Ddx5	DEAD (Asp-Glu-Ala-Asp) box polypeptide 5	Sum of Sum_pept_count	0	4	3	15
		Max of Max_Dif_pept	0	4	2	6
Acy1	aminoacylase 1	Sum of Sum_pept_count	0	8	0	10
		Max of Max_Dif_pept	0	4	0	6
Camk2d	calcium/calmodulin-dependent protein kinase II, delta	Sum of Sum_pept_count	3	6	7	6
		Max of Max_Dif_pept	3	5	6	3
Hspa4	heat shock protein 4	Sum of Sum_pept_count	0	8	7	0
		Max of Max_Dif_pept	0	6	6	0
Hsp90b1	heat shock protein 90, beta (Grp94), member 1	Sum of Sum_pept_count	0	0	2	15
		Max of Max_Dif_pept	0	0	1	6
Hba-a1	hemoglobin alpha, adult chain 1	Sum of Sum_pept_count	6	4	3	9
		Max of Max_Dif_pept	5	3	3	6
2210412D01Rik	RIKEN cDNA 2210412D01 gene	Sum of Sum_pept_count	29	8	7	8
		Max of Max_Dif_pept	6	5	3	6
Dgka	diacylglycerol kinase, alpha	Sum of Sum_pept_count	0	1	7	1
		Max of Max_Dif_pept	0	1	6	1
Canx	calnexin	Sum of Sum_pept_count	1	0	7	3
		Max of Max_Dif_pept	1	0	6	1
Gsn	gelsolin	Sum of Sum_pept_count	0	6	3	2
		Max of Max_Dif_pept	0	6	3	2
Ceacam18	carcinoembryonic antigen-related cell adhesion molecule 1	Sum of Sum_pept_count	13	9	8	7
		Max of Max_Dif_pept	6	4	4	5
Gpx4	glutathione peroxidase 4	Sum of Sum_pept_count	7	5	4	8
		Max of Max_Dif_pept	6	5	4	6
Cnp	2',3'-cyclic nucleotide 3' phosphodiesterase	Sum of Sum_pept_count	9	6	7	4
		Max of Max_Dif_pept	6	3	4	3
Gpd1l	glycerol-3-phosphate dehydrogenase 1-like	Sum of Sum_pept_count	3	7	6	6
		Max of Max_Dif_pept	3	5	6	6
Sept11	septin 11	Sum of Sum_pept_count	0	1	1	10
		Max of Max_Dif_pept	0	1	1	5
1810034K20Rik	RIKEN cDNA 1810034K20 gene	Sum of Sum_pept_count	0	8	4	2
		Max of Max_Dif_pept	0	5	3	2
Dnajc7	DnaJ (Hsp40) homolog, subfamily C, member 7	Sum of Sum_pept_count	0	1	0	5
		Max of Max_Dif_pept	0	1	0	5
Vta1	Vps20-associated 1 homolog (S. cerevisiae)	Sum of Sum_pept_count	2	6	2	9
		Max of Max_Dif_pept	1	5	2	4
Bco2	beta-carotene oxygenase 2	Sum of Sum_pept_count	3	5	0	4
		Max of Max_Dif_pept	3	5	0	4
Aldoart1	aldolase 1, A isoform, retrogene 1	Sum of Sum_pept_count	5	0	0	5
		Max of Max_Dif_pept	5	0	0	5
Dnm2	dynamin 2	Sum of Sum_pept_count	0	1	0	12
		Max of Max_Dif_pept	0	1	0	5
Uqcrc2	ubiquinol cytochrome c reductase core protein 2	Sum of Sum_pept_count	0	0	6	0
		Max of Max_Dif_pept	0	0	5	0
Grtp1	GH regulated TBC protein 1	Sum of Sum_pept_count	8	7	10	4
		Max of Max_Dif_pept	5	4	5	4
Entpd8	ectonucleoside triphosphate diphosphohydrolase 8	Sum of Sum_pept_count	21	13	14	12
		Max of Max_Dif_pept	4	5	3	4

Gene Symbol	Gene Description	Data	proximal	central	distal	total mucosa
Got2	glutamate oxaloacetate transaminase 2, mitochondrial	Sum of Sum_pept_count	0	0	0	5
		Max of Max_Dif_pept	0	0	0	5
Uqcrc1	ubiquinol-cytochrome c reductase core protein 1	Sum of Sum_pept_count	2	0	8	4
		Max of Max_Dif_pept	2	0	5	3
Cox4i1	cytochrome c oxidase subunit IV isoform 1	Sum of Sum_pept_count	2	0	5	24
		Max of Max_Dif_pept	2	0	4	5
Gsta1	glutathione S-transferase, alpha 1 (Ya)	Sum of Sum_pept_count	2	0	0	6
		Max of Max_Dif_pept	2	0	0	5
Gsta2	glutathione S-transferase, alpha 2 (Yc2)	Sum of Sum_pept_count	2	0	0	6
		Max of Max_Dif_pept	2	0	0	5
Cftr	cystic fibrosis transmembrane conductance regulator homolog	Sum of Sum_pept_count	21	1	3	1
		Max of Max_Dif_pept	5	1	1	1
Ubl3	ubiquitin-like 3	Sum of Sum_pept_count	11	8	9	8
		Max of Max_Dif_pept	5	5	5	5
Ube2n	ubiquitin-conjugating enzyme E2N	Sum of Sum_pept_count	1	9	5	6
		Max of Max_Dif_pept	1	5	2	4
Ube2m	ubiquitin-conjugating enzyme E2M (UBC12 homolog, yeast)	Sum of Sum_pept_count	3	4	5	3
		Max of Max_Dif_pept	3	4	5	3
Ubc	ubiquitin C	Sum of Sum_pept_count	112	104	122	94
		Max of Max_Dif_pept	4	5	4	4
Ubb	ubiquitin B	Sum of Sum_pept_count	224	208	244	188
		Max of Max_Dif_pept	4	5	4	4
Gstt3	glutathione S-transferase, theta 3	Sum of Sum_pept_count	1	1	0	5
		Max of Max_Dif_pept	1	1	0	5
Uba52	ubiquitin A-52 residue ribosomal protein fusion product 1	Sum of Sum_pept_count	56	52	61	47
		Max of Max_Dif_pept	4	5	4	4
Gyk	glycerol kinase	Sum of Sum_pept_count	0	5	1	4
		Max of Max_Dif_pept	0	5	1	4
Tubb2a-ps1	tubulin, beta 2a, pseudogene 1	Sum of Sum_pept_count	0	7	9	0
		Max of Max_Dif_pept	0	5	5	0
Got1	glutamate oxaloacetate transaminase 1, soluble	Sum of Sum_pept_count	0	4	2	5
		Max of Max_Dif_pept	0	4	2	5
2400001E08Rik	RIKEN cDNA 2400001E08 gene	Sum of Sum_pept_count	12	10	16	4
		Max of Max_Dif_pept	5	5	4	2
Esd	esterase D/formylglutathione hydrolase	Sum of Sum_pept_count	3	8	6	4
		Max of Max_Dif_pept	3	5	5	4
Casp3	caspase 3	Sum of Sum_pept_count	7	8	1	15
		Max of Max_Dif_pept	4	5	1	5
Arhgap1	Rho GTPase activating protein 1	Sum of Sum_pept_count	0	1	4	9
		Max of Max_Dif_pept	0	1	3	5
Golph3l	golgi phosphoprotein 3-like	Sum of Sum_pept_count	4	2	6	5
		Max of Max_Dif_pept	4	2	3	5
Tes	testis derived transcript	Sum of Sum_pept_count	2	4	1	6
		Max of Max_Dif_pept	1	4	1	5
Taok3	TAO kinase 3	Sum of Sum_pept_count	2	3	8	2
		Max of Max_Dif_pept	1	2	5	2
F11r	F11 receptor	Sum of Sum_pept_count	19	10	18	12
		Max of Max_Dif_pept	5	4	5	4
H2-T3	histocompatibility 2, T region locus 3	Sum of Sum_pept_count	2	0	5	5
		Max of Max_Dif_pept	2	0	5	5
Anxa1	annexin A1	Sum of Sum_pept_count	3	5	17	2
		Max of Max_Dif_pept	2	4	5	1
Stx8	syntaxin 8	Sum of Sum_pept_count	5	0	6	7
		Max of Max_Dif_pept	4	0	5	5
Stx7	syntaxin 7	Sum of Sum_pept_count	10	6	8	7
		Max of Max_Dif_pept	3	4	4	5
St3gal4	ST3 beta-galactoside alpha-2,3-sialyltransferase 4	Sum of Sum_pept_count	5	1	6	4
		Max of Max_Dif_pept	5	1	5	4
Chmp4c	chromatin modifying protein 4C	Sum of Sum_pept_count	13	11	6	8
		Max of Max_Dif_pept	4	4	3	5
B930007M17Rik	RIKEN cDNA B930007M17 gene	Sum of Sum_pept_count	2	11	3	10
		Max of Max_Dif_pept	2	4	2	5
Atp1a3	ATPase, Na+/K+ transporting, alpha 3 polypeptide	Sum of Sum_pept_count	12	8	4	9
		Max of Max_Dif_pept	5	3	2	2
5730469M10Rik	RIKEN cDNA 5730469M10 gene	Sum of Sum_pept_count	1	0	0	6
		Max of Max_Dif_pept	1	0	0	5
Slc39a4	solute carrier family 39 (zinc transporter), member 4	Sum of Sum_pept_count	35	18	42	16
		Max of Max_Dif_pept	5	5	4	3

Gene Symbol	Gene Description	Data	proximal	central	distal	total mucosa
Slc36a1	solute carrier family 36 (proton/amino acid symporter), member 1	Sum of Sum_pept_count	12	13	12	9
		Max of Max_Dif_pept	2	5	2	2
Slc35f2	solute carrier family 35, member F2	Sum of Sum_pept_count	27	18	21	13
		Max of Max_Dif_pept	4	5	3	3
Slc2a7	solute carrier family 2 (facilitated glucose transporter), member 7	Sum of Sum_pept_count	13	7	4	6
		Max of Max_Dif_pept	5	4	3	3
Slc22a5	solute carrier family 22 (organic cation transporter), member 5	Sum of Sum_pept_count	8	8	7	6
		Max of Max_Dif_pept	4	4	5	3
Sh3bgrl2	SH3 domain binding glutamic acid-rich protein like 2	Sum of Sum_pept_count	0	8	3	4
		Max of Max_Dif_pept	0	5	2	4
Sec24c	SEC24 related gene family, member C (S. cerevisiae)	Sum of Sum_pept_count	0	0	1	5
		Max of Max_Dif_pept	0	0	1	5
Scp2	sterol carrier protein 2, liver	Sum of Sum_pept_count	2	6	0	2
		Max of Max_Dif_pept	2	5	0	2
Fbp2	fructose bisphosphatase 2	Sum of Sum_pept_count	4	9	5	10
		Max of Max_Dif_pept	4	5	3	5
Scarb2	scavenger receptor class B, member 2	Sum of Sum_pept_count	2	3	5	2
		Max of Max_Dif_pept	2	3	5	2
Rps27a	ribosomal protein S27a	Sum of Sum_pept_count	56	52	61	50
		Max of Max_Dif_pept	4	5	4	4
Atp1a2	ATPase, Na+/K+ transporting, alpha 2 polypeptide	Sum of Sum_pept_count	12	8	4	9
		Max of Max_Dif_pept	5	3	2	2
Rpl12	ribosomal protein L12	Sum of Sum_pept_count	6	0	6	8
		Max of Max_Dif_pept	2	0	3	5
2210016F16Rik	RIKEN cDNA 2210016F16 gene	Sum of Sum_pept_count	2	2	1	5
		Max of Max_Dif_pept	2	2	1	5
Anxa3	annexin A3	Sum of Sum_pept_count	0	6	4	1
		Max of Max_Dif_pept	0	5	4	1
Chn2	chimerin (chimaerin) 2	Sum of Sum_pept_count	7	1	1	4
		Max of Max_Dif_pept	5	1	1	3
Rhob	ras homolog gene family, member B	Sum of Sum_pept_count	0	9	1	0
		Max of Max_Dif_pept	0	5	1	0
EG625929	predicted gene, EG625929	Sum of Sum_pept_count	3	5	9	2
		Max of Max_Dif_pept	3	4	5	2
Rab9	RAB9, member RAS oncogene family	Sum of Sum_pept_count	8	6	7	8
		Max of Max_Dif_pept	5	5	4	5
Cth	cystathionase (cystathionine gamma-lyase)	Sum of Sum_pept_count	2	3	2	8
		Max of Max_Dif_pept	1	2	1	5
Rab43	RAB43, member RAS oncogene family	Sum of Sum_pept_count	4	3	4	5
		Max of Max_Dif_pept	4	3	4	5
Rab3a	RAB3A, member RAS oncogene family	Sum of Sum_pept_count	1	0	5	6
		Max of Max_Dif_pept	1	0	5	4
Rab27a	RAB27A, member RAS oncogene family	Sum of Sum_pept_count	2	2	3	5
		Max of Max_Dif_pept	2	2	3	5
Ctss	cathepsin S	Sum of Sum_pept_count	0	5	1	4
		Max of Max_Dif_pept	0	5	1	4
Rab19	RAB19, member RAS oncogene family	Sum of Sum_pept_count	4	6	4	2
		Max of Max_Dif_pept	3	5	3	1
Arpc1b	actin related protein 2/3 complex, subunit 1B	Sum of Sum_pept_count	2	5	8	7
		Max of Max_Dif_pept	2	4	4	5
Ptpmt1	protein tyrosine phosphatase, mitochondrial 1	Sum of Sum_pept_count	6	0	3	0
		Max of Max_Dif_pept	5	0	2	0
Ptk6	PTK6 protein tyrosine kinase 6	Sum of Sum_pept_count	0	4	2	5
		Max of Max_Dif_pept	0	4	2	5
Cct7	chaperonin containing Tcp1, subunit 7 (eta)	Sum of Sum_pept_count	0	5	3	8
		Max of Max_Dif_pept	0	4	2	5
Cxadr	coxsackievirus and adenovirus receptor	Sum of Sum_pept_count	12	7	15	13
		Max of Max_Dif_pept	5	4	5	5
Bub3	budding uninhibited by benzimidazoles 3 homolog (S. cerevisiae)	Sum of Sum_pept_count	0	0	0	6
		Max of Max_Dif_pept	0	0	0	5

207

Gene Symbol	Gene Description	Data	proximal	central	distal	total mucosa
Psma2	proteasome (prosome, macropain) subunit, alpha type 2	Sum of Sum_pept_count	1	0	4	5
		Max of Max_Dif_pept	1	0	4	5
Cyb5	cytochrome b-5	Sum of Sum_pept_count	5	1	3	8
		Max of Max_Dif_pept	3	1	3	5
Prkcz	protein kinase C, zeta	Sum of Sum_pept_count	9	11	10	2
		Max of Max_Dif_pept	4	5	3	1
Ppp2cb	protein phosphatase 2 (formerly 2A), catalytic subunit, beta isoform	Sum of Sum_pept_count	0	0	0	5
		Max of Max_Dif_pept	0	0	0	5
Ppp2ca	protein phosphatase 2 (formerly 2A), catalytic subunit, alpha isoform	Sum of Sum_pept_count	0	5	1	5
		Max of Max_Dif_pept	0	3	1	5
Ppap2c	phosphatidic acid phosphatase type 2c	Sum of Sum_pept_count	8	7	8	10
		Max of Max_Dif_pept	3	3	3	5
Pmm2	phosphomannomutase 2	Sum of Sum_pept_count	1	5	3	9
		Max of Max_Dif_pept	1	4	2	5
Pi4k2b	phosphatidylinositol 4-kinase type 2 beta	Sum of Sum_pept_count	5	1	3	3
		Max of Max_Dif_pept	5	1	2	2
Pef1	penta-EF hand domain containing 1	Sum of Sum_pept_count	5	10	6	6
		Max of Max_Dif_pept	2	5	4	3
Hist1h4a	histone cluster 1, H4a	Sum of Sum_pept_count	18	4	11	16
		Max of Max_Dif_pept	5	1	3	4
Hist1h4b	histone cluster 1, H4b	Sum of Sum_pept_count	18	4	11	16
		Max of Max_Dif_pept	5	1	3	4
Hist1h4c	histone cluster 1, H4c	Sum of Sum_pept_count	18	4	11	16
		Max of Max_Dif_pept	5	1	3	4
Hist1h4d	histone cluster 1, H4d	Sum of Sum_pept_count	18	4	11	16
		Max of Max_Dif_pept	5	1	3	4
Hist1h4f	histone cluster 1, H4f	Sum of Sum_pept_count	18	4	11	16
		Max of Max_Dif_pept	5	1	3	4
Hist1h4h	histone cluster 1, H4h	Sum of Sum_pept_count	18	4	11	16
		Max of Max_Dif_pept	5	1	3	4
Hist1h4i	histone cluster 1, H4i	Sum of Sum_pept_count	18	4	11	16
		Max of Max_Dif_pept	5	1	3	4
Hist1h4j	histone cluster 1, H4j	Sum of Sum_pept_count	18	4	11	16
		Max of Max_Dif_pept	5	1	3	4
Hist1h4k	histone cluster 1, H4k	Sum of Sum_pept_count	18	4	11	16
		Max of Max_Dif_pept	5	1	3	4
Hist1h4m	histone cluster 1, H4m	Sum of Sum_pept_count	18	4	11	16
		Max of Max_Dif_pept	5	1	3	4
Cyp2c65	cytochrome P450, family 2, subfamily c, polypeptide 65	Sum of Sum_pept_count	2	0	1	5
		Max of Max_Dif_pept	2	0	1	5
Cldn3	claudin 3	Sum of Sum_pept_count	22	14	27	23
		Max of Max_Dif_pept	5	4	5	3
Otub1	OTU domain, ubiquitin aldehyde binding 1	Sum of Sum_pept_count	0	1	1	6
		Max of Max_Dif_pept	0	1	1	5
Ap1g2	adaptor protein complex AP-1, gamma 2 subunit	Sum of Sum_pept_count	0	0	0	5
		Max of Max_Dif_pept	0	0	0	5
Hist2h4	histone cluster 2, H4	Sum of Sum_pept_count	18	4	11	16
		Max of Max_Dif_pept	5	1	3	4
Arpc4	actin related protein 2/3 complex, subunit 4	Sum of Sum_pept_count	9	11	16	23
		Max of Max_Dif_pept	4	5	4	5
Nek6	NIMA (never in mitosis gene a)-related expressed kinase 6	Sum of Sum_pept_count	5	6	3	4
		Max of Max_Dif_pept	5	5	3	4
Gm1821	gene model 1821, (NCBI)	Sum of Sum_pept_count	112	104	122	94
		Max of Max_Dif_pept	4	5	4	4
Hist4h4	histone cluster 4, H4	Sum of Sum_pept_count	18	4	11	16
		Max of Max_Dif_pept	5	1	3	4
Acp2	acid phosphatase 2, lysosomal	Sum of Sum_pept_count	6	3	7	4
		Max of Max_Dif_pept	5	3	4	4
Hmgb1	high mobility group box 1	Sum of Sum_pept_count	0	5	1	4
		Max of Max_Dif_pept	0	5	1	4
Myo5b	myosin Vb	Sum of Sum_pept_count	10	9	9	14
		Max of Max_Dif_pept	3	3	4	5
Mylc2b	myosin light chain, regulatory B	Sum of Sum_pept_count	0	1	1	7
		Max of Max_Dif_pept	0	1	1	5
Ms4a10	membrane-spanning 4-domains, subfamily A, member 10	Sum of Sum_pept_count	30	4	2	12
		Max of Max_Dif_pept	5	2	1	3
Mogat2	monoacylglycerol O-acyltransferase 2	Sum of Sum_pept_count	0	0	1	6
		Max of Max_Dif_pept	0	0	1	5

Gene Symbol	Gene Description	Data	proximal	central	distal	total mucosa
Clta	clathrin, light polypeptide (Lca)	Sum of Sum_pept_count	0	0	0	10
		Max of Max_Dif_pept	0	0	0	5
Map2k1ip1	mitogen-activated protein kinase kinase 1 interacting protein 1	Sum of Sum_pept_count	1	4	11	1
		Max of Max_Dif_pept	1	3	5	1
Lyz1	lysozyme 1	Sum of Sum_pept_count	8	13	11	17
		Max of Max_Dif_pept	1	5	4	4
Lypla1	lysophospholipase 1	Sum of Sum_pept_count	7	3	4	8
		Max of Max_Dif_pept	4	3	3	5
Hpgd	hydroxyprostaglandin dehydrogenase 15 (NAD)	Sum of Sum_pept_count	2	3	1	6
		Max of Max_Dif_pept	2	3	1	5
Hprt1	hypoxanthine guanine phosphoribosyl transferase 1	Sum of Sum_pept_count	2	1	4	5
		Max of Max_Dif_pept	2	1	4	5
2900073G15Rik	RIKEN cDNA 2900073G15 gene	Sum of Sum_pept_count	0	1	1	7
		Max of Max_Dif_pept	0	1	1	5
Ddt	D-dopachrome tautomerase	Sum of Sum_pept_count	2	6	5	4
		Max of Max_Dif_pept	2	4	5	3
Acaa2	acetyl-Coenzyme A acyltransferase 2 (mitochondrial 3-oxoacyl-Coenzyme A thiolase)	Sum of Sum_pept_count	0	0	0	6
		Max of Max_Dif_pept	0	0	0	5
Ddx39	DEAD (Asp-Glu-Ala-Asp) box polypeptide 39	Sum of Sum_pept_count	2	6	5	10
		Max of Max_Dif_pept	2	5	5	5
Hsd17b6	hydroxysteroid (17-beta) dehydrogenase 6	Sum of Sum_pept_count	0	0	0	6
		Max of Max_Dif_pept	0	0	0	5
B020010K11Rik	RIKEN cDNA B020010K11 gene	Sum of Sum_pept_count	36	15	18	29
		Max of Max_Dif_pept	4	3	4	5
Gnao1	guanine nucleotide binding protein, alpha O	Sum of Sum_pept_count	26	12	23	25
		Max of Max_Dif_pept	5	5	4	5
Krt6a	keratin 6A	Sum of Sum_pept_count	0	0	7	0
		Max of Max_Dif_pept	0	0	5	0
Krt35	keratin 35	Sum of Sum_pept_count	12	0	0	0
		Max of Max_Dif_pept	5	0	0	0
Eif5a	eukaryotic translation initiation factor 5A	Sum of Sum_pept_count	0	6	2	3
		Max of Max_Dif_pept	0	5	1	2
Krt19	keratin 19	Sum of Sum_pept_count	2	6	1	19
		Max of Max_Dif_pept	2	2	1	5
Cmas	cytidine monophospho-N-acetylneuraminic acid synthetase	Sum of Sum_pept_count	0	0	0	6
		Max of Max_Dif_pept	0	0	0	5
Keap1	kelch-like ECH-associated protein 1	Sum of Sum_pept_count	1	1	1	5
		Max of Max_Dif_pept	1	1	1	5
Gnpda1	glucosamine-6-phosphate deaminase 1	Sum of Sum_pept_count	0	2	1	6
		Max of Max_Dif_pept	0	2	1	5
Gng12	guanine nucleotide binding protein (G protein), gamma 12	Sum of Sum_pept_count	12	10	10	4
		Max of Max_Dif_pept	5	5	4	2
Itln1	intelectin 1 (galactofuranose binding)	Sum of Sum_pept_count	0	2	4	5
		Max of Max_Dif_pept	0	2	4	5
Agr2	anterior gradient 2 (Xenopus laevis)	Sum of Sum_pept_count	0	2	1	22
		Max of Max_Dif_pept	0	1	1	5
Ilf3	interleukin enhancer binding factor 3	Sum of Sum_pept_count	0	1	0	8
		Max of Max_Dif_pept	0	1	0	5
Arhgap17	Rho GTPase activating protein 17	Sum of Sum_pept_count	8	12	3	8
		Max of Max_Dif_pept	4	3	2	3
Apol10a	apolipoprotein L 10a	Sum of Sum_pept_count	3	6	6	10
		Max of Max_Dif_pept	3	4	4	4
Idh3g	isocitrate dehydrogenase 3 (NAD+), gamma	Sum of Sum_pept_count	0	0	0	5
		Max of Max_Dif_pept	0	0	0	4
Anp32b	acidic nuclear phosphoprotein 32 family, member B	Sum of Sum_pept_count	4	7	6	4
		Max of Max_Dif_pept	4	3	3	3
Ykt6	YKT6 homolog (S. Cerevisiae)	Sum of Sum_pept_count	5	4	3	4
		Max of Max_Dif_pept	4	4	2	2
Gne	glucosamine	Sum of Sum_pept_count	0	0	0	4
		Max of Max_Dif_pept	0	0	0	4
Ddx3y	DEAD (Asp-Glu-Ala-Asp) box polypeptide 3, Y-linked	Sum of Sum_pept_count	8	11	23	10
		Max of Max_Dif_pept	3	2	4	4

Gene Symbol	Gene Description	Data	proximal	central	distal	total mucosa
Dnajc5	DnaJ (Hsp40) homolog, subfamily C, member 5	Sum of Sum_pept_count	12	9	8	10
		Max of Max_Dif_pept	4	4	3	4
Agrp	agouti related protein	Sum of Sum_pept_count	2	0	5	1
		Max of Max_Dif_pept	2	0	4	1
Impa1	inositol (myo)-1(or 4)-monophosphatase 1	Sum of Sum_pept_count	0	2	2	6
		Max of Max_Dif_pept	0	2	2	4
Wasf2	WAS protein family, member 2	Sum of Sum_pept_count	0	8	0	3
		Max of Max_Dif_pept	0	4	0	1
Vps33b	vacuolar protein sorting 33B (yeast)	Sum of Sum_pept_count	0	1	0	4
		Max of Max_Dif_pept	0	1	0	4
Vps26a	vacuolar protein sorting 26 homolog A (yeast)	Sum of Sum_pept_count	4	1	2	4
		Max of Max_Dif_pept	3	1	1	4
Cfl2	cofilin 2, muscle	Sum of Sum_pept_count	0	0	5	6
		Max of Max_Dif_pept	0	0	3	4
Vamp8	vesicle-associated membrane protein 8	Sum of Sum_pept_count	10	10	10	9
		Max of Max_Dif_pept	4	4	4	4
Vamp3	vesicle-associated membrane protein 3	Sum of Sum_pept_count	5	2	5	12
		Max of Max_Dif_pept	2	1	2	4
Uqcrq	ubiquinol-cytochrome c reductase, complex III subunit VII	Sum of Sum_pept_count	0	0	1	4
		Max of Max_Dif_pept	0	0	1	4
Itch	itchy, E3 ubiquitin protein ligase	Sum of Sum_pept_count	1	5	4	1
		Max of Max_Dif_pept	1	4	3	1
Ugt1a1	UDP glucuronosyltransferase 1 family, polypeptide A1	Sum of Sum_pept_count	2	0	4	4
		Max of Max_Dif_pept	2	0	4	4
Gstm2	glutathione S-transferase, mu 2	Sum of Sum_pept_count	0	0	0	4
		Max of Max_Dif_pept	0	0	0	4
Ube2d3	ubiquitin-conjugating enzyme E2D 3 (UBC4/5 homolog, yeast)	Sum of Sum_pept_count	3	7	3	4
		Max of Max_Dif_pept	1	4	1	2
H2-Aa	histocompatibility 2, class II antigen A, alpha	Sum of Sum_pept_count	8	8	21	8
		Max of Max_Dif_pept	3	3	4	3
Itgb1	integrin beta 1 (fibronectin receptor beta)	Sum of Sum_pept_count	0	0	6	6
		Max of Max_Dif_pept	0	0	3	4
Uap1l1	UDP-N-acteylglucosamine pyrophosphorylase 1-like 1	Sum of Sum_pept_count	1	2	1	4
		Max of Max_Dif_pept	1	2	1	4
Arhgef16	Rho guanine nucleotide exchange factor (GEF) 16	Sum of Sum_pept_count	6	8	5	5
		Max of Max_Dif_pept	3	4	3	2
Cox5a	cytochrome c oxidase, subunit Va	Sum of Sum_pept_count	1	2	7	12
		Max of Max_Dif_pept	1	1	4	4
Chmp1a	chromatin modifying protein 1A	Sum of Sum_pept_count	5	8	4	8
		Max of Max_Dif_pept	3	3	3	4
Amy2-1	amylase 2-1, pancreatic	Sum of Sum_pept_count	0	1	2	4
		Max of Max_Dif_pept	0	1	2	4
Copa	coatomer protein complex subunit alpha	Sum of Sum_pept_count	0	0	1	6
		Max of Max_Dif_pept	0	0	1	4
Akr1b3	aldo-keto reductase family 1, member B3 (aldose reductase)	Sum of Sum_pept_count	0	2	3	5
		Max of Max_Dif_pept	0	2	3	4
Cnnm4	cyclin M4	Sum of Sum_pept_count	0	0	5	0
		Max of Max_Dif_pept	0	0	4	0
H2-DMb2	histocompatibility 2, class II, locus Mb2	Sum of Sum_pept_count	5	5	17	6
		Max of Max_Dif_pept	3	4	4	4
Duoxa2	dual oxidase maturation factor 2	Sum of Sum_pept_count	0	1	36	6
		Max of Max_Dif_pept	0	1	4	3
Kif5b	kinesin family member 5B	Sum of Sum_pept_count	0	0	0	4
		Max of Max_Dif_pept	0	0	0	4
Espn	espin	Sum of Sum_pept_count	0	6	0	8
		Max of Max_Dif_pept	0	3	0	4
Tollip	toll interacting protein	Sum of Sum_pept_count	0	3	4	5
		Max of Max_Dif_pept	0	2	3	4
Dync1li1	dynein cytoplasmic 1 light intermediate chain 1	Sum of Sum_pept_count	0	0	0	6
		Max of Max_Dif_pept	0	0	0	4
Krt1	keratin 1	Sum of Sum_pept_count	27	21	28	33
		Max of Max_Dif_pept	4	3	4	3

Gene Symbol	Gene Description	Data	proximal	central	distal	total mucosa
Tmem55b	transmembrane protein 55b	Sum of Sum_pept_count	7	2	9	3
		Max of Max_Dif_pept	3	2	4	3
Tmem55a	transmembrane protein 55A	Sum of Sum_pept_count	5	1	8	2
		Max of Max_Dif_pept	3	1	4	1
Epb4.1l5	erythrocyte protein band 4.1-like 5	Sum of Sum_pept_count	9	4	2	2
		Max of Max_Dif_pept	4	1	1	1
Tmem33	transmembrane protein 33	Sum of Sum_pept_count	1	0	3	4
		Max of Max_Dif_pept	1	0	2	4
Krt2	keratin 2	Sum of Sum_pept_count	13	7	4	6
		Max of Max_Dif_pept	3	4	3	2
Tmed9	transmembrane emp24 protein transport domain containing 9	Sum of Sum_pept_count	2	0	3	8
		Max of Max_Dif_pept	1	0	3	4
Tmed7	transmembrane emp24 protein transport domain containing 7	Sum of Sum_pept_count	1	0	3	4
		Max of Max_Dif_pept	1	0	2	4
Tmed4	transmembrane emp24 protein transport domain containing 4	Sum of Sum_pept_count	2	0	4	7
		Max of Max_Dif_pept	1	0	4	4
Tmed10	transmembrane emp24-like trafficking protein 10 (yeast)	Sum of Sum_pept_count	0	0	1	6
		Max of Max_Dif_pept	0	0	1	4
Thop1	thimet oligopeptidase 1	Sum of Sum_pept_count	0	0	0	4
		Max of Max_Dif_pept	0	0	0	4
Tgm2	transglutaminase 2, C polypeptide	Sum of Sum_pept_count	0	1	0	4
		Max of Max_Dif_pept	0	1	0	4
Atp8a2	ATPase, aminophospholipid transporter-like, class I, type 8A, member 2	Sum of Sum_pept_count	0	0	4	1
		Max of Max_Dif_pept	0	0	4	1
Tfg	Trk-fused gene	Sum of Sum_pept_count	3	4	6	4
		Max of Max_Dif_pept	3	3	4	2
Tceb2	transcription elongation factor B (SIII), polypeptide 2	Sum of Sum_pept_count	1	6	5	7
		Max of Max_Dif_pept	1	4	4	4
Casp8	caspase 8	Sum of Sum_pept_count	0	6	1	7
		Max of Max_Dif_pept	0	3	1	4
Krt73	keratin 73	Sum of Sum_pept_count	4	1	0	0
		Max of Max_Dif_pept	4	1	0	0
Sult2b1	sulfotransferase family, cytosolic, 2B, member 1	Sum of Sum_pept_count	1	7	2	5
		Max of Max_Dif_pept	1	4	2	2
Stxbp3a	syntaxin binding protein 3A	Sum of Sum_pept_count	0	2	1	4
		Max of Max_Dif_pept	0	2	1	4
Faah	fatty acid amide hydrolase	Sum of Sum_pept_count	0	0	0	4
		Max of Max_Dif_pept	0	0	0	4
Spr	sepiapterin reductase	Sum of Sum_pept_count	1	2	4	3
		Max of Max_Dif_pept	1	2	4	3
Snx6	sorting nexin 6	Sum of Sum_pept_count	1	0	0	7
		Max of Max_Dif_pept	1	0	0	4
Krt81	keratin 81	Sum of Sum_pept_count	0	0	4	0
		Max of Max_Dif_pept	0	0	4	0
Snx5	sorting nexin 5	Sum of Sum_pept_count	0	2	4	5
		Max of Max_Dif_pept	0	2	3	4
Krt83	keratin 83	Sum of Sum_pept_count	5	0	4	0
		Max of Max_Dif_pept	4	0	4	0
Snx3	sorting nexin 3	Sum of Sum_pept_count	7	5	3	8
		Max of Max_Dif_pept	4	3	2	4
Efr3b	EFR3 homolog B (S. cerevisiae)	Sum of Sum_pept_count	17	0	2	1
		Max of Max_Dif_pept	4	0	1	1
Slc7a9	solute carrier family 7 (cationic amino acid transporter, y+ system), member 9	Sum of Sum_pept_count	12	6	19	9
		Max of Max_Dif_pept	4	3	3	4
Slc6a8	solute carrier family 6 (neurotransmitter transporter, creatine), member 8	Sum of Sum_pept_count	3	9	9	5
		Max of Max_Dif_pept	2	4	3	3
Lamp2	lysosomal-associated membrane protein 2	Sum of Sum_pept_count	4	4	11	5
		Max of Max_Dif_pept	2	2	4	2
Aqp1	aquaporin 1	Sum of Sum_pept_count	30	22	28	26
		Max of Max_Dif_pept	4	3	3	3
Ap1s1	adaptor protein complex AP-1, sigma 1	Sum of Sum_pept_count	0	1	1	7
		Max of Max_Dif_pept	0	1	1	4
Blvra	biliverdin reductase A	Sum of Sum_pept_count	2	2	3	4
		Max of Max_Dif_pept	2	2	3	4
Slc46a1	solute carrier family 46, member 1	Sum of Sum_pept_count	47	13	5	27
		Max of Max_Dif_pept	4	3	3	3

211

Gene Symbol	Gene Description	Data	proximal	central	distal	total mucosa
Slc44a1	solute carrier family 44, member 1	Sum of Sum_pept_count	2	0	6	2
		Max of Max_Dif_pept	1	0	4	2
Adss	adenylosuccinate synthetase, non muscle	Sum of Sum_pept_count	1	6	4	4
		Max of Max_Dif_pept	1	4	4	4
2310046K01Rik	RIKEN cDNA 2310046K01 gene	Sum of Sum_pept_count	8	28	40	17
		Max of Max_Dif_pept	2	4	3	3
Hint1	histidine triad nucleotide binding protein 1	Sum of Sum_pept_count	3	4	4	10
		Max of Max_Dif_pept	2	3	2	4
Slc2a9	solute carrier family 2 (facilitated glucose transporter), member 9	Sum of Sum_pept_count	20	17	22	14
		Max of Max_Dif_pept	4	4	4	4
Chmp5	chromatin modifying protein 5	Sum of Sum_pept_count	4	7	4	7
		Max of Max_Dif_pept	2	3	2	4
Bpgm	2,3-bisphosphoglycerate mutase	Sum of Sum_pept_count	0	0	0	5
		Max of Max_Dif_pept	0	0	0	4
EG432987	predicted gene, EG432987	Sum of Sum_pept_count	3	0	10	0
		Max of Max_Dif_pept	3	0	4	0
Cr1l	complement component (3b/4b) receptor 1-like	Sum of Sum_pept_count	0	0	6	3
		Max of Max_Dif_pept	0	0	4	3
Slc13a1	solute carrier family 13 (sodium/sulfate symporters), member 1	Sum of Sum_pept_count	8	0	9	4
		Max of Max_Dif_pept	4	0	4	3
Shmt2	serine hydroxymethyltransferase 2 (mitochondrial)	Sum of Sum_pept_count	0	0	0	4
		Max of Max_Dif_pept	0	0	0	4
Sec24d	SEC24 related gene family, member D (S. cerevisiae)	Sum of Sum_pept_count	1	1	0	10
		Max of Max_Dif_pept	1	1	0	4
Chmp6	chromatin modifying protein 6	Sum of Sum_pept_count	17	13	6	18
		Max of Max_Dif_pept	4	4	3	4
Crot	carnitine O-octanoyltransferase	Sum of Sum_pept_count	0	0	0	4
		Max of Max_Dif_pept	0	0	0	4
Fbxl20	F-box and leucine-rich repeat protein 20	Sum of Sum_pept_count	3	1	4	1
		Max of Max_Dif_pept	3	1	4	1
Cs	citrate synthase	Sum of Sum_pept_count	0	2	1	4
		Max of Max_Dif_pept	0	2	1	4
Rras	Harvey rat sarcoma oncogene, subgroup R	Sum of Sum_pept_count	12	9	13	10
		Max of Max_Dif_pept	4	3	4	4
Rps9	ribosomal protein S9	Sum of Sum_pept_count	2	2	5	4
		Max of Max_Dif_pept	1	2	3	4
Rps5	ribosomal protein S5	Sum of Sum_pept_count	5	2	6	4
		Max of Max_Dif_pept	1	1	4	2
Rps2	ribosomal protein S2	Sum of Sum_pept_count	2	0	5	1
		Max of Max_Dif_pept	2	0	4	1
Hsd17b11	hydroxysteroid (17-beta) dehydrogenase 11	Sum of Sum_pept_count	1	1	4	6
		Max of Max_Dif_pept	1	1	2	4
Rps16	ribosomal protein S16	Sum of Sum_pept_count	3	3	6	4
		Max of Max_Dif_pept	2	3	3	4
Rps15a	ribosomal protein S15a	Sum of Sum_pept_count	5	2	3	5
		Max of Max_Dif_pept	4	1	2	3
Rps13	ribosomal protein S13	Sum of Sum_pept_count	4	2	5	5
		Max of Max_Dif_pept	4	2	4	4
Rpl35a	ribosomal protein L35a	Sum of Sum_pept_count	0	0	2	6
		Max of Max_Dif_pept	0	0	2	4
Rpl23	ribosomal protein L23	Sum of Sum_pept_count	3	2	5	9
		Max of Max_Dif_pept	1	1	2	4
Rhoq	ras homolog gene family, member Q	Sum of Sum_pept_count	5	2	0	3
		Max of Max_Dif_pept	4	2	0	3
Rasgef1a	RasGEF domain family, member 1A	Sum of Sum_pept_count	2	4	1	4
		Max of Max_Dif_pept	1	4	1	2
EG625055	predicted gene, EG625055	Sum of Sum_pept_count	0	0	4	1
		Max of Max_Dif_pept	0	0	4	1
Arbp	acidic ribosomal phosphoprotein P0	Sum of Sum_pept_count	1	0	1	8
		Max of Max_Dif_pept	1	0	1	4
1700009N14Rik	RIKEN cDNA 1700009N14 gene	Sum of Sum_pept_count	1	6	10	5
		Max of Max_Dif_pept	1	4	4	2
Csnk2a1	casein kinase 2, alpha 1 polypeptide	Sum of Sum_pept_count	0	4	4	0
		Max of Max_Dif_pept	0	3	4	0
Hpcal1	hippocalcin-like 1	Sum of Sum_pept_count	3	4	4	2
		Max of Max_Dif_pept	3	4	4	2

Gene Symbol	Gene Description	Data	proximal	central	distal	total mucosa
Csrp1	cysteine and glycine-rich protein 1	Sum of Sum_pept_count	0	5	0	4
		Max of Max_Dif_pept	0	2	0	4
Akr7a5	aldo-keto reductase family 7, member A5 (aflatoxin aldehyde reductase)	Sum of Sum_pept_count	1	4	1	6
		Max of Max_Dif_pept	1	4	1	4
Atp5h	ATP synthase, H+ transporting, mitochondrial F0 complex, subunit d	Sum of Sum_pept_count	0	0	0	5
		Max of Max_Dif_pept	0	0	0	4
Btbd7	BTB (POZ) domain containing 7	Sum of Sum_pept_count	2	0	5	1
		Max of Max_Dif_pept	2	0	4	1
Ptp4a1	protein tyrosine phosphatase 4a1	Sum of Sum_pept_count	4	4	3	4
		Max of Max_Dif_pept	3	3	3	4
Ptgr1	prostaglandin reductase 1	Sum of Sum_pept_count	0	6	1	2
		Max of Max_Dif_pept	0	4	1	2
Psme1	proteasome (prosome, macropain) 28 subunit, alpha	Sum of Sum_pept_count	1	4	3	4
		Max of Max_Dif_pept	1	4	2	4
Psmd7	proteasome (prosome, macropain) 26S subunit, non-ATPase, 7	Sum of Sum_pept_count	0	0	0	7
		Max of Max_Dif_pept	0	0	0	4
Psmd5	proteasome (prosome, macropain) 26S subunit, non-ATPase, 5	Sum of Sum_pept_count	0	0	0	5
		Max of Max_Dif_pept	0	0	0	4
Psmb10	proteasome (prosome, macropain) subunit, beta type 10	Sum of Sum_pept_count	6	3	3	9
		Max of Max_Dif_pept	3	1	2	4
Psmb1	proteasome (prosome, macropain) subunit, beta type 1	Sum of Sum_pept_count	1	3	5	2
		Max of Max_Dif_pept	1	3	4	2
Psma7	proteasome (prosome, macropain) subunit, alpha type 7	Sum of Sum_pept_count	1	4	5	7
		Max of Max_Dif_pept	1	3	3	4
Mapk14	mitogen-activated protein kinase 14	Sum of Sum_pept_count	0	0	0	6
		Max of Max_Dif_pept	0	0	0	4
Arg2	arginase type II	Sum of Sum_pept_count	0	0	0	5
		Max of Max_Dif_pept	0	0	0	4
Psma6	proteasome (prosome, macropain) subunit, alpha type 6	Sum of Sum_pept_count	0	2	2	5
		Max of Max_Dif_pept	0	1	1	4
Psma1	proteasome (prosome, macropain) subunit, alpha type 1	Sum of Sum_pept_count	1	2	4	3
		Max of Max_Dif_pept	1	1	4	2
Mat2b	methionine adenosyltransferase II, beta	Sum of Sum_pept_count	0	0	0	7
		Max of Max_Dif_pept	0	0	0	4
Bzw1	basic leucine zipper and W2 domains 1	Sum of Sum_pept_count	3	2	3	4
		Max of Max_Dif_pept	3	2	3	4
3110049J23Rik	RIKEN cDNA 3110049J23 gene	Sum of Sum_pept_count	0	4	2	4
		Max of Max_Dif_pept	0	3	2	4
Ppp2r5d	protein phosphatase 2, regulatory subunit B (B56), delta isoform	Sum of Sum_pept_count	0	3	1	8
		Max of Max_Dif_pept	0	2	1	4
Ppp2r4	protein phosphatase 2A, regulatory subunit B (PR 53)	Sum of Sum_pept_count	1	3	1	6
		Max of Max_Dif_pept	1	2	1	4
4930539N22Rik	RIKEN cDNA 4930539N22 gene	Sum of Sum_pept_count	0	0	1	4
		Max of Max_Dif_pept	0	0	1	4
Cd38	CD38 antigen	Sum of Sum_pept_count	2	0	9	6
		Max of Max_Dif_pept	2	0	4	3
Arpc3	actin related protein 2/3 complex, subunit 3	Sum of Sum_pept_count	2	5	7	6
		Max of Max_Dif_pept	2	3	4	4
Ppcs	phosphopantothenoylcysteine synthetase	Sum of Sum_pept_count	0	0	0	4
		Max of Max_Dif_pept	0	0	0	4
Cycs	cytochrome c, somatic	Sum of Sum_pept_count	2	5	1	3
		Max of Max_Dif_pept	2	4	1	3
Ppa1	pyrophosphatase (inorganic) 1	Sum of Sum_pept_count	6	7	1	7
		Max of Max_Dif_pept	3	4	1	4
Plcd1	phospholipase C, delta 1	Sum of Sum_pept_count	1	2	6	1
		Max of Max_Dif_pept	1	2	4	1
Gclm	glutamate-cysteine ligase , modifier subunit	Sum of Sum_pept_count	4	2	1	7
		Max of Max_Dif_pept	2	1	1	4

Gene Symbol	Gene Description	Data	proximal	central	distal	total mucosa
Pkp3	plakophilin 3	Sum of Sum_pept_count	4	8	4	6
		Max of Max_Dif_pept	1	4	2	4
1810065E05Rik	RIKEN cDNA 1810065E05 gene	Sum of Sum_pept_count	1	1	15	2
		Max of Max_Dif_pept	1	1	4	1
Abhd14b	abhydrolase domain containing 14b	Sum of Sum_pept_count	1	1	0	6
		Max of Max_Dif_pept	1	1	0	4
Pitpnb	phosphatidylinositol transfer protein, beta	Sum of Sum_pept_count	3	4	2	2
		Max of Max_Dif_pept	3	4	2	2
Gdpd1	glycerophosphodiester phosphodiesterase domain containing 1	Sum of Sum_pept_count	0	0	4	6
		Max of Max_Dif_pept	0	0	2	4
Mosc2	MOCO sulphurase C-terminal domain containing 2	Sum of Sum_pept_count	1	0	0	9
		Max of Max_Dif_pept	1	0	0	4
Arhgdia	Rho GDP dissociation inhibitor (GDI) alpha	Sum of Sum_pept_count	1	5	3	7
		Max of Max_Dif_pept	1	4	2	3
Pfkl	phosphofructokinase, liver, B-type	Sum of Sum_pept_count	0	2	0	5
		Max of Max_Dif_pept	0	2	0	4
Pebp1	phosphatidylethanolamine binding protein 1	Sum of Sum_pept_count	1	5	1	3
		Max of Max_Dif_pept	1	4	1	3
Hnrnpf	heterogeneous nuclear ribonucleoprotein F	Sum of Sum_pept_count	5	8	2	6
		Max of Max_Dif_pept	2	3	2	4
Pdxdc1	pyridoxal-dependent decarboxylase domain containing 1	Sum of Sum_pept_count	0	0	2	7
		Max of Max_Dif_pept	0	0	1	4
Pdlim5	PDZ and LIM domain 5	Sum of Sum_pept_count	1	2	1	6
		Max of Max_Dif_pept	1	1	1	4
Pcyt1a	phosphate cytidylyltransferase 1, choline, alpha isoform	Sum of Sum_pept_count	4	4	2	5
		Max of Max_Dif_pept	4	2	2	3
Pctk3	PCTAIRE-motif protein kinase 3	Sum of Sum_pept_count	9	5	4	4
		Max of Max_Dif_pept	4	2	2	3
2310057J18Rik	RIKEN cDNA 2310057J18 gene	Sum of Sum_pept_count	1	5	1	3
		Max of Max_Dif_pept	1	4	1	3
Gk5	glycerol kinase 5 (putative)	Sum of Sum_pept_count	4	0	0	0
		Max of Max_Dif_pept	4	0	0	0
Pbld	phenazine biosynthesis-like protein domain containing	Sum of Sum_pept_count	7	1	3	3
		Max of Max_Dif_pept	4	1	1	2
Pard6b	par-6 (partitioning defective 6) homolog beta (C. elegans)	Sum of Sum_pept_count	2	2	1	4
		Max of Max_Dif_pept	1	1	1	4
Glipr2	GLI pathogenesis-related 2	Sum of Sum_pept_count	0	5	6	5
		Max of Max_Dif_pept	0	4	4	2
Oxsr1	oxidative-stress responsive 1	Sum of Sum_pept_count	0	0	0	6
		Max of Max_Dif_pept	0	0	0	4
Muc2	mucin 2	Sum of Sum_pept_count	0	0	0	6
		Max of Max_Dif_pept	0	0	0	4
Nptn	neuroplastin	Sum of Sum_pept_count	8	6	8	7
		Max of Max_Dif_pept	3	3	3	4
Clrn3	clarin 3	Sum of Sum_pept_count	60	45	53	38
		Max of Max_Dif_pept	4	4	4	4
1810046J19Rik	RIKEN cDNA 1810046J19 gene	Sum of Sum_pept_count	0	4	0	1
		Max of Max_Dif_pept	0	4	0	1
Myh14	myosin, heavy polypeptide 14	Sum of Sum_pept_count	0	0	0	8
		Max of Max_Dif_pept	0	0	0	4
D6Wsu176e	DNA segment, Chr 6, Wayne State University 176, expressed	Sum of Sum_pept_count	1	0	4	3
		Max of Max_Dif_pept	1	0	4	3
Nlrp4e	NLR family, pyrin domain containing 4E	Sum of Sum_pept_count	3	2	7	14
		Max of Max_Dif_pept	2	2	3	4
Nlrc4	NLR family, CARD domain containing 4	Sum of Sum_pept_count	0	1	1	8
		Max of Max_Dif_pept	0	1	1	4
Mylk	myosin, light polypeptide kinase	Sum of Sum_pept_count	2	8	2	11
		Max of Max_Dif_pept	1	4	1	4
Dad1	defender against cell death 1	Sum of Sum_pept_count	1	0	4	3
		Max of Max_Dif_pept	1	0	4	3
Clint1	clathrin interactor 1	Sum of Sum_pept_count	0	0	0	6
		Max of Max_Dif_pept	0	0	0	4
Copg	coatomer protein complex, subunit gamma	Sum of Sum_pept_count	0	2	4	16
		Max of Max_Dif_pept	0	1	2	4

214

Gene Symbol	Gene Description	Data	proximal	central	distal	total mucosa
Cd81	CD81 antigen	Sum of Sum_pept_count	6	3	7	4
		Max of Max_Dif_pept	3	3	4	2
2700060E02Rik	RIKEN cDNA 2700060E02 gene	Sum of Sum_pept_count	0	0	2	5
		Max of Max_Dif_pept	0	0	2	4
01.09.2005	septin 5	Sum of Sum_pept_count	1	1	3	7
		Max of Max_Dif_pept	1	1	2	3
Zdhhc9	zinc finger, DHHC domain containing 9	Sum of Sum_pept_count	7	0	5	1
		Max of Max_Dif_pept	3	0	2	1
Bcap31	B-cell receptor-associated protein 31	Sum of Sum_pept_count	0	0	2	7
		Max of Max_Dif_pept	0	0	1	3
Idi1	isopentenyl-diphosphate delta isomerase	Sum of Sum_pept_count	0	0	2	3
		Max of Max_Dif_pept	0	0	2	3
Ap3m1	adaptor-related protein complex 3, mu 1 subunit	Sum of Sum_pept_count	0	1	0	3
		Max of Max_Dif_pept	0	1	0	3
Hnrnpa2b1	heterogeneous nuclear ribonucleoprotein A2/B1	Sum of Sum_pept_count	0	4	0	10
		Max of Max_Dif_pept	0	2	0	3
Bcl2l14	Bcl2-like 14 (apoptosis facilitator)	Sum of Sum_pept_count	6	1	1	1
		Max of Max_Dif_pept	3	1	1	1
5730446C15Rik	RIKEN cDNA 5730446C15 gene	Sum of Sum_pept_count	2	2	4	3
		Max of Max_Dif_pept	2	2	2	3
Grb2	growth factor receptor bound protein 2	Sum of Sum_pept_count	0	5	3	3
		Max of Max_Dif_pept	0	3	3	3
Vps28	vacuolar protein sorting 28 (yeast)	Sum of Sum_pept_count	2	3	0	3
		Max of Max_Dif_pept	2	3	0	3
Vps24	vacuolar protein sorting 24 (yeast)	Sum of Sum_pept_count	8	7	5	8
		Max of Max_Dif_pept	3	2	2	3
Vdac3	voltage-dependent anion channel 3	Sum of Sum_pept_count	0	0	3	3
		Max of Max_Dif_pept	0	0	2	3
Cope	coatomer protein complex, subunit epsilon	Sum of Sum_pept_count	0	0	0	4
		Max of Max_Dif_pept	0	0	0	3
1810020D17Rik	RIKEN cDNA 1810020D17 gene	Sum of Sum_pept_count	0	4	0	3
		Max of Max_Dif_pept	0	3	0	3
Myo1c	myosin IC	Sum of Sum_pept_count	0	0	4	3
		Max of Max_Dif_pept	0	0	2	3
Dab1	disabled homolog 1 (Drosophila)	Sum of Sum_pept_count	4	7	7	3
		Max of Max_Dif_pept	2	3	3	2
Ndufa13	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 13	Sum of Sum_pept_count	1	0	1	5
		Max of Max_Dif_pept	1	0	1	3
Isoc1	isochorismatase domain containing 1	Sum of Sum_pept_count	1	2	1	3
		Max of Max_Dif_pept	1	2	1	3
9130404H23Rik	RIKEN cDNA 9130404H23 gene	Sum of Sum_pept_count	0	4	2	3
		Max of Max_Dif_pept	0	3	1	3
Epb4.1l4b	erythrocyte protein band 4.1-like 4b	Sum of Sum_pept_count	4	6	7	4
		Max of Max_Dif_pept	1	2	3	3
Ndufb4	NADH dehydrogenase (ubiquinone) 1 beta subcomplex 4	Sum of Sum_pept_count	1	0	1	4
		Max of Max_Dif_pept	1	0	1	3
Ndufb5	NADH dehydrogenase (ubiquinone) 1 beta subcomplex, 5	Sum of Sum_pept_count	0	0	1	5
		Max of Max_Dif_pept	0	0	1	3
Uqcrfs1	ubiquinol-cytochrome c reductase, Rieske iron-sulfur polypeptide 1	Sum of Sum_pept_count	0	0	0	4
		Max of Max_Dif_pept	0	0	0	3
Ndufb9	NADH dehydrogenase (ubiquinone) 1 beta subcomplex, 9	Sum of Sum_pept_count	0	0	0	3
		Max of Max_Dif_pept	0	0	0	3
Dgat1	diacylglycerol O-acyltransferase 1	Sum of Sum_pept_count	0	0	0	3
		Max of Max_Dif_pept	0	0	0	3
Gss	glutathione synthetase	Sum of Sum_pept_count	0	3	0	2
		Max of Max_Dif_pept	0	3	0	2
Uqcrb	ubiquinol-cytochrome c reductase binding protein	Sum of Sum_pept_count	0	0	3	0
		Max of Max_Dif_pept	0	0	3	0

215

Gene Symbol	Gene Description	Data	proximal	central	distal	total mucosa
Ngef	neuronal guanine nucleotide exchange factor	Sum of Sum_pept_count	2	9	6	2
		Max of Max_Dif_pept	1	3	3	2
Nit2	nitrilase family, member 2	Sum of Sum_pept_count	0	0	0	3
		Max of Max_Dif_pept	0	0	0	3
Cd82	CD82 antigen	Sum of Sum_pept_count	13	10	16	8
		Max of Max_Dif_pept	3	3	3	2
Ugt2b34	UDP glucuronosyltransferase 2 family, polypeptide B34	Sum of Sum_pept_count	0	0	2	5
		Max of Max_Dif_pept	0	0	1	3
Gstm1	glutathione S-transferase, mu 1	Sum of Sum_pept_count	0	0	0	3
		Max of Max_Dif_pept	0	0	0	3
Abcb6	ATP-binding cassette, sub-family B (MDR/TAP), member 6	Sum of Sum_pept_count	3	0	6	3
		Max of Max_Dif_pept	3	0	3	2
Itfg3	integrin alpha FG-GAP repeat containing 3	Sum of Sum_pept_count	0	1	3	1
		Max of Max_Dif_pept	0	1	3	1
Uevld	UEV and lactate/malate dehyrogenase domains	Sum of Sum_pept_count	0	3	1	0
		Max of Max_Dif_pept	0	3	1	0
Ube2v2	ubiquitin-conjugating enzyme E2 variant 2	Sum of Sum_pept_count	1	5	1	3
		Max of Max_Dif_pept	1	3	1	3
Ube2v1	ubiquitin-conjugating enzyme E2 variant 1	Sum of Sum_pept_count	1	5	1	3
		Max of Max_Dif_pept	1	3	1	3
Itga6	integrin alpha 6	Sum of Sum_pept_count	2	0	5	3
		Max of Max_Dif_pept	1	0	3	2
Gstt1	glutathione S-transferase, theta 1	Sum of Sum_pept_count	1	0	0	3
		Max of Max_Dif_pept	1	0	0	3
Aadac	arylacetamide deacetylase (esterase)	Sum of Sum_pept_count	0	0	0	3
		Max of Max_Dif_pept	0	0	0	3
Aspa	aspartoacylase	Sum of Sum_pept_count	0	2	3	1
		Max of Max_Dif_pept	0	2	3	1
EG668668	predicted gene, EG668668	Sum of Sum_pept_count	6	5	5	5
		Max of Max_Dif_pept	3	1	2	3
Iah1	isoamyl acetate-hydrolyzing esterase 1 homolog (S. cerevisiae)	Sum of Sum_pept_count	0	1	3	7
		Max of Max_Dif_pept	0	1	2	3
Uba1	ubiquitin-like modifier activating enzyme 1	Sum of Sum_pept_count	0	2	0	5
		Max of Max_Dif_pept	0	2	0	3
Itm2b	integral membrane protein 2B	Sum of Sum_pept_count	4	0	2	1
		Max of Max_Dif_pept	3	0	2	1
Akap7	A kinase (PRKA) anchor protein 7	Sum of Sum_pept_count	10	6	5	6
		Max of Max_Dif_pept	3	2	2	3
Arhgap18	Rho GTPase activating protein 18	Sum of Sum_pept_count	0	3	0	1
		Max of Max_Dif_pept	0	3	0	1
Clec2h	C-type lectin domain family 2, member h	Sum of Sum_pept_count	0	1	7	3
		Max of Max_Dif_pept	0	1	3	1
Dsg4	desmoglein 4	Sum of Sum_pept_count	4	0	1	0
		Max of Max_Dif_pept	3	0	1	0
Ogdh	oxoglutarate dehydrogenase (lipoamide)	Sum of Sum_pept_count	0	0	0	4
		Max of Max_Dif_pept	0	0	0	3
Tubb4	tubulin, beta 4	Sum of Sum_pept_count	2	2	0	8
		Max of Max_Dif_pept	1	1	0	3
ORF9	open reading frame 9	Sum of Sum_pept_count	1	0	3	1
		Max of Max_Dif_pept	1	0	3	1
Tuba3b	tubulin, alpha 3B	Sum of Sum_pept_count	3	3	3	3
		Max of Max_Dif_pept	2	1	2	3
Glrx3	glutaredoxin 3	Sum of Sum_pept_count	0	3	2	3
		Max of Max_Dif_pept	0	2	2	3
Otop3	otopetrin 3	Sum of Sum_pept_count	15	0	0	2
		Max of Max_Dif_pept	3	0	0	2
OTTMUSG0000000	predicted gene, OTTMUSG00000001634	Sum of Sum_pept_count	20	23	20	13
		Max of Max_Dif_pept	2	3	2	2
Tuba3a	tubulin, alpha 3A	Sum of Sum_pept_count	3	3	3	3
		Max of Max_Dif_pept	2	1	2	3
Adsl	adenylosuccinate lyase	Sum of Sum_pept_count	0	3	2	3
		Max of Max_Dif_pept	0	3	2	3
Chmp1b	chromatin modifying protein 1B	Sum of Sum_pept_count	9	9	4	10
		Max of Max_Dif_pept	3	2	1	3

Gene Symbol	Gene Description	Data	proximal	central	distal	total mucosa
Ttc38	tetratricopeptide repeat domain 38	Sum of Sum_pept_count	0	3	1	0
		Max of Max_Dif_pept	0	3	1	0
Duox1	dual oxidase 1	Sum of Sum_pept_count	1	0	9	0
		Max of Max_Dif_pept	1	0	3	0
Tspan15	tetraspanin 15	Sum of Sum_pept_count	2	0	6	2
		Max of Max_Dif_pept	1	0	3	1
Tsg101	tumor susceptibility gene 101	Sum of Sum_pept_count	1	3	0	3
		Max of Max_Dif_pept	1	2	0	3
Try4	trypsin 4	Sum of Sum_pept_count	0	3	1	6
		Max of Max_Dif_pept	0	2	1	3
Arl15	ADP-ribosylation factor-like 15	Sum of Sum_pept_count	3	3	1	2
		Max of Max_Dif_pept	3	3	1	2
Pabpc1	poly A binding protein, cytoplasmic 1	Sum of Sum_pept_count	0	0	0	4
		Max of Max_Dif_pept	0	0	0	3
Glo1	glyoxalase 1	Sum of Sum_pept_count	0	1	1	3
		Max of Max_Dif_pept	0	1	1	3
BC085271	cDNA sequence BC085271	Sum of Sum_pept_count	0	7	2	4
		Max of Max_Dif_pept	0	3	2	2
Ela1	elastase 1, pancreatic	Sum of Sum_pept_count	3	0	2	4
		Max of Max_Dif_pept	3	0	2	3
Kng2	kininogen 2	Sum of Sum_pept_count	0	0	1	5
		Max of Max_Dif_pept	0	0	1	3
Cpne1	copine I	Sum of Sum_pept_count	0	3	3	0
		Max of Max_Dif_pept	0	3	3	0
Kpnb1	karyopherin (importin) beta 1	Sum of Sum_pept_count	0	1	0	5
		Max of Max_Dif_pept	0	1	0	3
Mtus1	mitochondrial tumor suppressor 1	Sum of Sum_pept_count	0	0	3	0
		Max of Max_Dif_pept	0	0	3	0
Tpd52	tumor protein D52	Sum of Sum_pept_count	0	0	0	3
		Max of Max_Dif_pept	0	0	0	3
Aldh3a2	aldehyde dehydrogenase family 3, subfamily A2	Sum of Sum_pept_count	0	0	2	3
		Max of Max_Dif_pept	0	0	2	3
Tnpo1	transportin 1	Sum of Sum_pept_count	0	0	1	5
		Max of Max_Dif_pept	0	0	1	3
Tmprss8	transmembrane protease, serine 8 (intestinal)	Sum of Sum_pept_count	9	13	30	8
		Max of Max_Dif_pept	2	3	3	1
Pcbp2	poly(rC) binding protein 2	Sum of Sum_pept_count	0	4	1	3
		Max of Max_Dif_pept	0	3	1	3
Mtmr2	myotubularin related protein 2	Sum of Sum_pept_count	0	2	3	2
		Max of Max_Dif_pept	0	2	3	2
Tmprss4	transmembrane protease, serine 4	Sum of Sum_pept_count	6	2	6	4
		Max of Max_Dif_pept	3	1	2	2
Pck2	phosphoenolpyruvate carboxykinase 2 (mitochondrial)	Sum of Sum_pept_count	0	0	0	3
		Max of Max_Dif_pept	0	0	0	3
Arhgap27	Rho GTPase activating protein 27	Sum of Sum_pept_count	0	1	3	2
		Max of Max_Dif_pept	0	1	3	2
Eif6	eukaryotic translation initiation factor 6	Sum of Sum_pept_count	2	1	3	3
		Max of Max_Dif_pept	2	1	1	3
Arhgef3	Rho guanine nucleotide exchange factor (GEF) 3	Sum of Sum_pept_count	0	0	3	3
		Max of Max_Dif_pept	0	0	2	3
Eef1a2	eukaryotic translation elongation factor 1 alpha 2	Sum of Sum_pept_count	7	7	5	9
		Max of Max_Dif_pept	2	2	2	3
Tmem189	transmembrane protein 189	Sum of Sum_pept_count	1	5	1	3
		Max of Max_Dif_pept	1	3	1	3
Cldn23	claudin 23	Sum of Sum_pept_count	13	3	7	6
		Max of Max_Dif_pept	3	1	3	2
Tmem106b	transmembrane protein 106B	Sum of Sum_pept_count	0	0	3	0
		Max of Max_Dif_pept	0	0	3	0
4930471M23Rik	RIKEN cDNA 4930471M23 gene	Sum of Sum_pept_count	11	4	10	7
		Max of Max_Dif_pept	3	2	2	2
Cdk5	cyclin-dependent kinase 5	Sum of Sum_pept_count	0	0	1	3
		Max of Max_Dif_pept	0	0	1	3
Tmed2	transmembrane emp24 domain trafficking protein 2	Sum of Sum_pept_count	0	0	0	3
		Max of Max_Dif_pept	0	0	0	3
Anp32a	acidic (leucine-rich) nuclear phosphoprotein 32 family, member A	Sum of Sum_pept_count	0	4	0	4
		Max of Max_Dif_pept	0	2	0	3
Tm4sf5	transmembrane 4 superfamily member 5	Sum of Sum_pept_count	20	9	6	4
		Max of Max_Dif_pept	3	2	1	2

Gene Symbol	Gene Description	Data	proximal	central	distal	total mucosa
Eef1d	eukaryotic translation elongation factor 1 delta (guanine nucleotide exchange protein)	Sum of Sum_pept_count	0	0	0	8
		Max of Max_Dif_pept	0	0	0	3
Dctn2	dynactin 2	Sum of Sum_pept_count	0	0	0	3
		Max of Max_Dif_pept	0	0	0	3
Pdxk	pyridoxal (pyridoxine, vitamin B6) kinase	Sum of Sum_pept_count	0	3	0	4
		Max of Max_Dif_pept	0	3	0	3
Acadl	acyl-Coenzyme A dehydrogenase, long-chain	Sum of Sum_pept_count	0	0	0	5
		Max of Max_Dif_pept	0	0	0	3
Thy1	thymus cell antigen 1, theta	Sum of Sum_pept_count	0	0	2	4
		Max of Max_Dif_pept	0	0	1	3
Ada	adenosine deaminase	Sum of Sum_pept_count	1	3	0	2
		Max of Max_Dif_pept	1	3	0	1
Thnsl2	threonine synthase-like 2 (bacterial)	Sum of Sum_pept_count	0	1	0	3
		Max of Max_Dif_pept	0	1	0	3
H2-Q6	histocompatibility 2, Q region locus 6	Sum of Sum_pept_count	2	2	3	3
		Max of Max_Dif_pept	2	2	3	2
Cda	cytidine deaminase	Sum of Sum_pept_count	2	6	0	5
		Max of Max_Dif_pept	1	3	0	2
Tcirg1	T-cell, immune regulator 1, ATPase, H+ transporting, lysosomal V0 protein A3	Sum of Sum_pept_count	2	0	3	0
		Max of Max_Dif_pept	1	0	3	0
H2-Q7	histocompatibility 2, Q region locus 7	Sum of Sum_pept_count	2	2	3	3
		Max of Max_Dif_pept	2	2	3	2
Dhfr	dihydrofolate reductase	Sum of Sum_pept_count	3	1	0	3
		Max of Max_Dif_pept	2	1	0	3
Hadh	hydroxyacyl-Coenzyme A dehydrogenase	Sum of Sum_pept_count	1	4	1	5
		Max of Max_Dif_pept	1	3	1	3
Stx4a	syntaxin 4A (placental)	Sum of Sum_pept_count	2	0	4	3
		Max of Max_Dif_pept	1	0	2	3
Stk16	serine/threonine kinase 16	Sum of Sum_pept_count	0	2	3	2
		Max of Max_Dif_pept	0	2	3	2
Acyp1	acylphosphatase 1, erythrocyte (common) type	Sum of Sum_pept_count	0	3	0	0
		Max of Max_Dif_pept	0	3	0	0
Pgm2l1	phosphoglucomutase 2-like 1	Sum of Sum_pept_count	0	3	0	0
		Max of Max_Dif_pept	0	3	0	0
Pgm3	phosphoglucomutase 3	Sum of Sum_pept_count	0	3	1	1
		Max of Max_Dif_pept	0	3	1	1
Pgrmc1	progesterone receptor membrane component 1	Sum of Sum_pept_count	0	0	0	4
		Max of Max_Dif_pept	0	0	0	3
Phb	prohibitin	Sum of Sum_pept_count	0	0	2	3
		Max of Max_Dif_pept	0	0	1	3
BC046404	cDNA sequence BC046404	Sum of Sum_pept_count	5	5	3	0
		Max of Max_Dif_pept	3	3	2	0
Pi4k2a	phosphatidylinositol 4-kinase type 2 alpha	Sum of Sum_pept_count	2	0	4	0
		Max of Max_Dif_pept	2	0	3	0
Capn5	calpain 5	Sum of Sum_pept_count	0	2	4	0
		Max of Max_Dif_pept	0	2	3	0
C77080	expressed sequence C77080	Sum of Sum_pept_count	2	5	2	0
		Max of Max_Dif_pept	1	3	1	0
Hagh	hydroxyacyl glutathione hydrolase	Sum of Sum_pept_count	0	2	0	4
		Max of Max_Dif_pept	0	1	0	3
St14	suppression of tumorigenicity 14 (colon carcinoma)	Sum of Sum_pept_count	0	0	5	3
		Max of Max_Dif_pept	0	0	3	3
St13	suppression of tumorigenicity 13	Sum of Sum_pept_count	1	1	1	4
		Max of Max_Dif_pept	1	1	1	3
Ssr4	signal sequence receptor, delta	Sum of Sum_pept_count	1	0	3	3
		Max of Max_Dif_pept	1	0	3	2
Capg	capping protein (actin filament), gelsolin-like	Sum of Sum_pept_count	0	7	1	1
		Max of Max_Dif_pept	0	3	1	1
Pkn2	protein kinase N2	Sum of Sum_pept_count	5	2	7	0
		Max of Max_Dif_pept	3	2	3	0
Pkn3	protein kinase N3	Sum of Sum_pept_count	0	7	2	4
		Max of Max_Dif_pept	0	3	2	2
Pkp1	plakophilin 1	Sum of Sum_pept_count	4	0	3	0
		Max of Max_Dif_pept	3	0	2	0
Chmp2b	chromatin modifying protein 2B	Sum of Sum_pept_count	2	7	2	3
		Max of Max_Dif_pept	1	3	1	1

Gene Symbol	Gene Description	Data	proximal	central	distal	total mucosa
Plac8	placenta-specific 8	Sum of Sum_pept_count	19	12	16	8
		Max of Max_Dif_pept	3	2	2	1
Mif	macrophage migration inhibitory factor	Sum of Sum_pept_count	1	4	4	3
		Max of Max_Dif_pept	1	3	2	2
Aprt	adenine phosphoribosyl transferase	Sum of Sum_pept_count	1	3	0	2
		Max of Max_Dif_pept	1	3	0	2
Ddc	dopa decarboxylase	Sum of Sum_pept_count	0	0	0	3
		Max of Max_Dif_pept	0	0	0	3
Acly	ATP citrate lyase	Sum of Sum_pept_count	0	0	3	8
		Max of Max_Dif_pept	0	0	1	3
Cnbp	cellular nucleic acid binding protein	Sum of Sum_pept_count	0	0	0	3
		Max of Max_Dif_pept	0	0	0	3
Plekhf1	pleckstrin homology domain containing, family F (with FYVE domain) member 1	Sum of Sum_pept_count	4	1	0	2
		Max of Max_Dif_pept	3	1	0	2
Krt82	keratin 82	Sum of Sum_pept_count	3	1	0	0
		Max of Max_Dif_pept	3	1	0	0
Eif5	eukaryotic translation initiation factor 5	Sum of Sum_pept_count	0	0	0	4
		Max of Max_Dif_pept	0	0	0	3
Krtcap3	keratinocyte associated protein 3	Sum of Sum_pept_count	1	0	4	3
		Max of Max_Dif_pept	1	0	3	2
Snx2	sorting nexin 2	Sum of Sum_pept_count	0	0	0	10
		Max of Max_Dif_pept	0	0	0	3
Snx1	sorting nexin 1	Sum of Sum_pept_count	0	0	0	10
		Max of Max_Dif_pept	0	0	0	3
Blmh	bleomycin hydrolase	Sum of Sum_pept_count	0	2	3	2
		Max of Max_Dif_pept	0	2	3	1
Cav1	caveolin, caveolae protein 1	Sum of Sum_pept_count	1	1	2	9
		Max of Max_Dif_pept	1	1	1	3
Smad2	MAD homolog 2 (Drosophila)	Sum of Sum_pept_count	4	1	2	1
		Max of Max_Dif_pept	3	1	1	1
Slc7a8	solute carrier family 7 (cationic amino acid transporter, y+ system), member 8	Sum of Sum_pept_count	9	2	6	6
		Max of Max_Dif_pept	3	2	3	3
0910001A06Rik	RIKEN cDNA 0910001A06 gene	Sum of Sum_pept_count	2	3	2	4
		Max of Max_Dif_pept	2	3	1	3
Lancl2	LanC (bacterial lantibiotic synthetase component C)-like 2	Sum of Sum_pept_count	0	1	0	3
		Max of Max_Dif_pept	0	1	0	3
Ppme1	protein phosphatase methylesterase 1	Sum of Sum_pept_count	0	0	0	5
		Max of Max_Dif_pept	0	0	0	3
Slc6a20b	solute carrier family 6 (neurotransmitter transporter), member 20B	Sum of Sum_pept_count	1	0	3	0
		Max of Max_Dif_pept	1	0	3	0
Comt	catechol-O-methyltransferase	Sum of Sum_pept_count	0	0	1	4
		Max of Max_Dif_pept	0	0	1	3
Mfge8	milk fat globule-EGF factor 8 protein	Sum of Sum_pept_count	0	0	3	0
		Max of Max_Dif_pept	0	0	3	0
Slc5a6	solute carrier family 5 (sodium-dependent vitamin transporter), member 6	Sum of Sum_pept_count	7	0	9	3
		Max of Max_Dif_pept	3	0	3	2
Lck	lymphocyte protein tyrosine kinase	Sum of Sum_pept_count	2	0	4	2
		Max of Max_Dif_pept	2	0	3	2
EG243302	predicted gene, EG243302	Sum of Sum_pept_count	4	1	4	4
		Max of Max_Dif_pept	2	1	2	3
Slc4a1	solute carrier family 4 (anion exchanger), member 1	Sum of Sum_pept_count	0	0	3	3
		Max of Max_Dif_pept	0	0	3	2
Atp5o	ATP synthase, H+ transporting, mitochondrial F1 complex, O subunit	Sum of Sum_pept_count	1	0	1	3
		Max of Max_Dif_pept	1	0	1	3
Blvrb	biliverdin reductase B (flavin reductase (NADPH))	Sum of Sum_pept_count	0	1	0	4
		Max of Max_Dif_pept	0	1	0	3
Gmfb	glia maturation factor, beta	Sum of Sum_pept_count	1	3	1	1
		Max of Max_Dif_pept	1	3	1	1
Slc43a2	solute carrier family 43, member 2	Sum of Sum_pept_count	1	0	2	8
		Max of Max_Dif_pept	1	0	1	3

Gene Symbol	Gene Description	Data	proximal	central	distal	total mucosa
Cyc1	cytochrome c-1	Sum of Sum_pept_count	0	0	2	7
		Max of Max_Dif_pept	0	0	2	3
1110012J17Rik	RIKEN cDNA 1110012J17 gene	Sum of Sum_pept_count	4	1	5	2
		Max of Max_Dif_pept	2	1	3	2
Slc37a2	solute carrier family 37 (glycerol-3-phosphate transporter), member 2	Sum of Sum_pept_count	12	0	0	3
		Max of Max_Dif_pept	3	0	0	1
Hint3	histidine triad nucleotide binding protein 3	Sum of Sum_pept_count	0	3	0	3
		Max of Max_Dif_pept	0	3	0	2
Bzw2	basic leucine zipper and W2 domains 2	Sum of Sum_pept_count	2	3	1	3
		Max of Max_Dif_pept	2	2	1	3
Prg2	proteoglycan 2, bone marrow	Sum of Sum_pept_count	5	0	0	3
		Max of Max_Dif_pept	3	0	0	2
EG667525	predicted gene, EG667525	Sum of Sum_pept_count	3	1	3	0
		Max of Max_Dif_pept	3	1	3	0
Prkaa2	protein kinase, AMP-activated, alpha 2 catalytic subunit	Sum of Sum_pept_count	0	1	0	3
		Max of Max_Dif_pept	0	1	0	3
Prkab1	protein kinase, AMP-activated, beta 1 non-catalytic subunit	Sum of Sum_pept_count	0	0	1	5
		Max of Max_Dif_pept	0	0	1	3
Hip1r	huntingtin interacting protein 1 related	Sum of Sum_pept_count	3	0	0	4
		Max of Max_Dif_pept	2	0	0	3
Slc2a2	solute carrier family 2 (facilitated glucose transporter), member 2	Sum of Sum_pept_count	5	3	1	4
		Max of Max_Dif_pept	3	2	1	1
Cps1	carbamoyl-phosphate synthetase 1	Sum of Sum_pept_count	0	0	0	5
		Max of Max_Dif_pept	0	0	0	3
Cpt2	carnitine palmitoyltransferase 2	Sum of Sum_pept_count	0	0	0	3
		Max of Max_Dif_pept	0	0	0	3
Slc26a2	solute carrier family 26 (sulfate transporter), member 2	Sum of Sum_pept_count	24	23	30	13
		Max of Max_Dif_pept	3	3	3	3
Slc25a1	solute carrier family 25 (mitochondrial carrier, citrate transporter), member 1	Sum of Sum_pept_count	0	0	1	3
		Max of Max_Dif_pept	0	0	1	3
Lin7c	lin-7 homolog C (C. elegans)	Sum of Sum_pept_count	3	5	3	3
		Max of Max_Dif_pept	3	3	3	3
Slc12a2	solute carrier family 12, member 2	Sum of Sum_pept_count	3	0	10	2
		Max of Max_Dif_pept	1	0	3	1
Prkci	protein kinase C, iota	Sum of Sum_pept_count	5	7	6	2
		Max of Max_Dif_pept	2	3	1	1
Slc11a2	solute carrier family 11 (proton-coupled divalent metal ion transporters), member 2	Sum of Sum_pept_count	7	1	3	2
		Max of Max_Dif_pept	3	1	1	1
2610002M06Rik	RIKEN cDNA 2610002M06 gene	Sum of Sum_pept_count	6	2	4	5
		Max of Max_Dif_pept	3	2	3	3
Skp1a	S-phase kinase-associated protein 1A	Sum of Sum_pept_count	2	4	0	2
		Max of Max_Dif_pept	2	3	0	2
Shq1	SHQ1 homolog (S. cerevisiae)	Sum of Sum_pept_count	1	5	1	3
		Max of Max_Dif_pept	1	3	1	3
Fbp1	fructose bisphosphatase 1	Sum of Sum_pept_count	0	4	0	4
		Max of Max_Dif_pept	0	3	0	3
Arl5a	ADP-ribosylation factor-like 5A	Sum of Sum_pept_count	3	2	1	1
		Max of Max_Dif_pept	3	2	1	1
Set	SET translocation	Sum of Sum_pept_count	0	7	2	4
		Max of Max_Dif_pept	0	3	2	2
Crip1	cysteine-rich protein 1 (intestinal)	Sum of Sum_pept_count	2	3	2	7
		Max of Max_Dif_pept	1	1	1	3
Psen1	presenilin 1	Sum of Sum_pept_count	16	7	21	15
		Max of Max_Dif_pept	3	2	3	2
Serpina1d	serine (or cysteine) peptidase inhibitor, clade A, member 1d	Sum of Sum_pept_count	0	5	1	3
		Max of Max_Dif_pept	0	3	1	3
Serpina1c	serine (or cysteine) peptidase inhibitor, clade A, member 1c	Sum of Sum_pept_count	0	5	0	0
		Max of Max_Dif_pept	0	3	0	0
Serpina1b	serine (or cysteine) preptidase inhibitor, clade A, member 1b	Sum of Sum_pept_count	0	5	0	0
		Max of Max_Dif_pept	0	3	0	0
Psma3	proteasome (prosome, macropain) subunit, alpha type 3	Sum of Sum_pept_count	1	3	3	6
		Max of Max_Dif_pept	1	2	2	3

220

Gene Symbol	Gene Description	Data	proximal	central	distal	total mucosa
Psma4	proteasome (prosome, macropain) subunit, alpha type 4	Sum of Sum_pept_count	1	2	3	3
		Max of Max_Dif_pept	1	2	2	3
Serhl	serine hydrolase-like	Sum of Sum_pept_count	0	0	0	3
		Max of Max_Dif_pept	0	0	0	3
Selenbp1	selenium binding protein 1	Sum of Sum_pept_count	1	5	2	3
		Max of Max_Dif_pept	1	3	1	3
Mark2	MAP/microtubule affinity-regulating kinase 2	Sum of Sum_pept_count	6	3	4	1
		Max of Max_Dif_pept	3	1	1	1
Cct2	chaperonin containing Tcp1, subunit 2 (beta)	Sum of Sum_pept_count	0	1	1	6
		Max of Max_Dif_pept	0	1	1	3
Sec24b	SEC24 related gene family, member B (S. cerevisiae)	Sum of Sum_pept_count	0	1	0	3
		Max of Max_Dif_pept	0	1	0	3
Lman1	lectin, mannose-binding, 1	Sum of Sum_pept_count	0	0	2	4
		Max of Max_Dif_pept	0	0	1	3
Atp5b	ATP synthase, H+ transporting mitochondrial F1 complex, beta subunit	Sum of Sum_pept_count	1	2	1	6
		Max of Max_Dif_pept	1	1	1	3
Psmb8	proteasome (prosome, macropain) subunit, beta type 8 (large multifunctional peptidase 7)	Sum of Sum_pept_count	0	1	1	3
		Max of Max_Dif_pept	0	1	1	3
Sec14l2	SEC14-like 2 (S. cerevisiae)	Sum of Sum_pept_count	6	4	1	4
		Max of Max_Dif_pept	3	3	1	3
Galm	galactose mutarotase	Sum of Sum_pept_count	0	3	2	5
		Max of Max_Dif_pept	0	3	2	3
Sdcbp2	syndecan binding protein (syntenin) 2	Sum of Sum_pept_count	1	5	4	5
		Max of Max_Dif_pept	1	2	2	3
Scarb1	scavenger receptor class B, member 1	Sum of Sum_pept_count	3	0	0	0
		Max of Max_Dif_pept	3	0	0	0
EG638487	predicted gene, EG638487	Sum of Sum_pept_count	0	7	2	4
		Max of Max_Dif_pept	0	3	2	2
S100a10	S100 calcium binding protein A10 (calpactin)	Sum of Sum_pept_count	4	9	9	6
		Max of Max_Dif_pept	3	3	3	3
Cct8	chaperonin containing Tcp1, subunit 8 (theta)	Sum of Sum_pept_count	0	2	3	3
		Max of Max_Dif_pept	0	1	1	3
Lmtk2	lemur tyrosine kinase 2	Sum of Sum_pept_count	4	0	1	2
		Max of Max_Dif_pept	3	0	1	2
Rragd	Ras-related GTP binding D	Sum of Sum_pept_count	2	1	0	3
		Max of Max_Dif_pept	2	1	0	3
Rragc	Ras-related GTP binding C	Sum of Sum_pept_count	2	1	2	3
		Max of Max_Dif_pept	2	1	2	3
Rraga	Ras-related GTP binding A	Sum of Sum_pept_count	2	0	2	3
		Max of Max_Dif_pept	2	0	2	3
LOC620497	solute carrier family 5 (sodium-dependent vitamin transporter), member 6 pseudogene	Sum of Sum_pept_count	7	0	9	3
		Max of Max_Dif_pept	3	0	3	2
Rps7	ribosomal protein S7	Sum of Sum_pept_count	3	0	3	3
		Max of Max_Dif_pept	3	0	3	3
Rps6	ribosomal protein S6	Sum of Sum_pept_count	5	0	3	3
		Max of Max_Dif_pept	3	0	2	1
Psmd8	proteasome (prosome, macropain) 26S subunit, non-ATPase, 8	Sum of Sum_pept_count	0	0	0	6
		Max of Max_Dif_pept	0	0	0	3
Rps4x	ribosomal protein S4, X-linked	Sum of Sum_pept_count	6	5	8	5
		Max of Max_Dif_pept	3	1	3	3
Rps27l	ribosomal protein S27-like	Sum of Sum_pept_count	0	2	4	9
		Max of Max_Dif_pept	0	2	3	3
Rps25	ribosomal protein S25	Sum of Sum_pept_count	4	1	4	4
		Max of Max_Dif_pept	2	1	2	3
Rps20	ribosomal protein S20	Sum of Sum_pept_count	3	1	1	4
		Max of Max_Dif_pept	3	1	1	3
Fgd4	FYVE, RhoGEF and PH domain containing 4	Sum of Sum_pept_count	2	1	3	2
		Max of Max_Dif_pept	2	1	3	1
Ptges3	prostaglandin E synthase 3 (cytosolic)	Sum of Sum_pept_count	4	7	2	6
		Max of Max_Dif_pept	2	2	1	3
Rps17	ribosomal protein S17	Sum of Sum_pept_count	2	2	5	4
		Max of Max_Dif_pept	1	1	2	3

Gene Symbol	Gene Description	Data	proximal	central	distal	total mucosa
Rps11	ribosomal protein S11	Sum of Sum_pept_count	3	0	2	3
		Max of Max_Dif_pept	2	0	2	3
Csk	c-src tyrosine kinase	Sum of Sum_pept_count	0	3	1	3
		Max of Max_Dif_pept	0	3	1	3
Ptp4a2	protein tyrosine phosphatase 4a2	Sum of Sum_pept_count	2	4	3	3
		Max of Max_Dif_pept	2	3	3	3
Rpl34	ribosomal protein L34	Sum of Sum_pept_count	4	0	5	2
		Max of Max_Dif_pept	2	0	3	2
Rpl27a	ribosomal protein L27a	Sum of Sum_pept_count	3	2	6	6
		Max of Max_Dif_pept	1	2	2	3
Ptprj	protein tyrosine phosphatase, receptor type, J	Sum of Sum_pept_count	1	6	5	3
		Max of Max_Dif_pept	1	3	2	2
Pttg1ip	pituitary tumor-transforming 1 interacting protein	Sum of Sum_pept_count	9	1	6	5
		Max of Max_Dif_pept	3	1	2	2
Rpl24	ribosomal protein L24	Sum of Sum_pept_count	7	3	7	7
		Max of Max_Dif_pept	2	1	2	3
Rpl18a	ribosomal protein L18A	Sum of Sum_pept_count	0	0	3	3
		Max of Max_Dif_pept	0	0	3	3
Rpl14	ribosomal protein L14	Sum of Sum_pept_count	1	0	1	4
		Max of Max_Dif_pept	1	0	1	3
Cltb	clathrin, light polypeptide (Lcb)	Sum of Sum_pept_count	0	0	0	5
		Max of Max_Dif_pept	0	0	0	3
Rpl10	ribosomal protein 10	Sum of Sum_pept_count	0	0	3	3
		Max of Max_Dif_pept	0	0	3	2
Maob	monoamine oxidase B	Sum of Sum_pept_count	0	0	0	3
		Max of Max_Dif_pept	0	0	0	3
RP23-357I14.1	novel protein similar to solute carrier family 28 (sodium-coupled nucleoside transporter) member 2 (Slc28a2)	Sum of Sum_pept_count	1	3	3	0
		Max of Max_Dif_pept	1	3	3	0
Rnf128	ring finger protein 128	Sum of Sum_pept_count	9	1	5	3
		Max of Max_Dif_pept	3	1	2	1
Rit1	Ras-like without CAAX 1	Sum of Sum_pept_count	4	1	0	3
		Max of Max_Dif_pept	3	1	0	3
Alad	aminolevulinate, delta-, dehydratase	Sum of Sum_pept_count	0	2	2	3
		Max of Max_Dif_pept	0	2	2	3
Fgr	Gardner-Rasheed feline sarcoma viral (Fgr) oncogene homolog	Sum of Sum_pept_count	0	0	3	1
		Max of Max_Dif_pept	0	0	3	1
EG545332	predicted gene, EG545332	Sum of Sum_pept_count	1	5	1	0
		Max of Max_Dif_pept	1	3	1	0
Rfk	riboflavin kinase	Sum of Sum_pept_count	2	3	0	2
		Max of Max_Dif_pept	2	3	0	1
Reep6	receptor accessory protein 6	Sum of Sum_pept_count	2	1	2	10
		Max of Max_Dif_pept	2	1	2	3
Csnk1g1	casein kinase 1, gamma 1	Sum of Sum_pept_count	6	3	2	4
		Max of Max_Dif_pept	3	2	1	2
Dlat	dihydrolipoamide S-acetyltransferase (E2 component of pyruvate dehydrogenase complex)	Sum of Sum_pept_count	0	0	0	5
		Max of Max_Dif_pept	0	0	0	3
Arpc5	actin related protein 2/3 complex, subunit 5	Sum of Sum_pept_count	3	9	4	7
		Max of Max_Dif_pept	2	3	2	2
Ranbp1	RAN binding protein 1	Sum of Sum_pept_count	1	0	3	6
		Max of Max_Dif_pept	1	0	2	3
M6pr	mannose-6-phosphate receptor, cation dependent	Sum of Sum_pept_count	2	0	5	9
		Max of Max_Dif_pept	2	0	3	3
Cisd2	CDGSH iron sulfur domain 2	Sum of Sum_pept_count	0	0	0	4
		Max of Max_Dif_pept	0	0	0	3
Rabif	RAB interacting factor	Sum of Sum_pept_count	3	3	2	4
		Max of Max_Dif_pept	2	2	2	3
EG629575	predicted gene, EG629575	Sum of Sum_pept_count	5	0	3	3
		Max of Max_Dif_pept	3	0	2	1
Ctdspl	CTD (carboxy-terminal domain, RNA polymerase II, polypeptide A) small phosphatase-like	Sum of Sum_pept_count	6	5	4	1
		Max of Max_Dif_pept	3	3	2	1
Rab31	RAB31, member RAS oncogene family	Sum of Sum_pept_count	4	4	1	0
		Max of Max_Dif_pept	3	3	1	0
Aacs	acetoacetyl-CoA synthetase	Sum of Sum_pept_count	0	0	0	3
		Max of Max_Dif_pept	0	0	0	3

Gene Symbol	Gene Description	Data	proximal	central	distal	total mucosa
Frs2	fibroblast growth factor receptor substrate 2	Sum of Sum_pept_count	4	1	1	0
		Max of Max_Dif_pept	3	1	1	0
Ctps2	cytidine 5'-triphosphate synthase 2	Sum of Sum_pept_count	0	0	0	3
		Max of Max_Dif_pept	0	0	0	3
Frmd8	FERM domain containing 8	Sum of Sum_pept_count	2	1	1	3
		Max of Max_Dif_pept	2	1	1	3
Rab3c	RAB3C, member RAS oncogene family	Sum of Sum_pept_count	0	2	3	0
		Max of Max_Dif_pept	0	2	3	0
2210010C04Rik	RIKEN cDNA 2210010C04 gene	Sum of Sum_pept_count	13	16	18	18
		Max of Max_Dif_pept	1	2	1	2
Hist1h2bf	histone cluster 1, H2bf	Sum of Sum_pept_count	5	1	4	7
		Max of Max_Dif_pept	2	1	2	2
Hist1h2be	histone cluster 1, H2be	Sum of Sum_pept_count	5	1	4	7
		Max of Max_Dif_pept	2	1	2	2
Frmd3	FERM domain containing 3	Sum of Sum_pept_count	0	0	0	3
		Max of Max_Dif_pept	0	0	0	2
Rab39	RAB39, member RAS oncogene family	Sum of Sum_pept_count	4	0	1	4
		Max of Max_Dif_pept	1	0	1	2
Hist1h2bc	histone cluster 1, H2bc	Sum of Sum_pept_count	5	1	4	7
		Max of Max_Dif_pept	2	1	2	2
Rab32	RAB32, member RAS oncogene family	Sum of Sum_pept_count	2	1	0	2
		Max of Max_Dif_pept	2	1	0	2
EG629557	predicted gene, EG629557	Sum of Sum_pept_count	4	0	3	0
		Max of Max_Dif_pept	2	0	2	0
EG628596	predicted gene, EG628596	Sum of Sum_pept_count	0	3	2	2
		Max of Max_Dif_pept	0	1	1	2
Actn1	actinin, alpha 1	Sum of Sum_pept_count	0	5	11	7
		Max of Max_Dif_pept	0	2	2	2
Fmo5	flavin containing monooxygenase 5	Sum of Sum_pept_count	0	0	0	2
		Max of Max_Dif_pept	0	0	0	2
Arcn1	archain 1	Sum of Sum_pept_count	0	2	1	6
		Max of Max_Dif_pept	0	2	1	2
Hist1h2bb	histone cluster 1, H2bb	Sum of Sum_pept_count	5	1	4	7
		Max of Max_Dif_pept	2	1	2	2
Dirc2	disrupted in renal carcinoma 2 (human)	Sum of Sum_pept_count	0	1	6	1
		Max of Max_Dif_pept	0	1	2	1
Ftl2	ferritin light chain 2	Sum of Sum_pept_count	0	0	0	2
		Max of Max_Dif_pept	0	0	0	2
M6prbp1	mannose-6-phosphate receptor binding protein 1	Sum of Sum_pept_count	6	1	1	2
		Max of Max_Dif_pept	2	1	1	2
Capns1	calpain, small subunit 1	Sum of Sum_pept_count	2	1	3	2
		Max of Max_Dif_pept	1	1	2	1
Rac3	RAS-related C3 botulinum substrate 3	Sum of Sum_pept_count	4	4	3	4
		Max of Max_Dif_pept	2	1	1	2
EG632352	predicted gene, EG632352	Sum of Sum_pept_count	0	2	2	0
		Max of Max_Dif_pept	0	2	2	0
Hist1h2bg	histone cluster 1, H2bg	Sum of Sum_pept_count	5	1	4	7
		Max of Max_Dif_pept	2	1	2	2
EG627371	predicted gene, EG627371	Sum of Sum_pept_count	3	0	2	4
		Max of Max_Dif_pept	2	0	1	1
Ctrb1	chymotrypsinogen B1	Sum of Sum_pept_count	0	0	0	2
		Max of Max_Dif_pept	0	0	0	2
Lypla2	lysophospholipase 2	Sum of Sum_pept_count	1	0	2	2
		Max of Max_Dif_pept	1	0	2	2
Hist1h2ba	histone cluster 1, H2ba	Sum of Sum_pept_count	0	1	2	1
		Max of Max_Dif_pept	0	1	2	1
Rap1gds1	RAP1, GTP-GDP dissociation stimulator 1	Sum of Sum_pept_count	0	0	0	2
		Max of Max_Dif_pept	0	0	0	2
Rap2a	RAS related protein 2a	Sum of Sum_pept_count	0	1	4	0
		Max of Max_Dif_pept	0	1	2	0
Rab24	RAB24, member RAS oncogene family	Sum of Sum_pept_count	2	0	1	1
		Max of Max_Dif_pept	2	0	1	1
Hist1h2bh	histone cluster 1, H2bh	Sum of Sum_pept_count	5	1	4	7
		Max of Max_Dif_pept	2	1	2	2
Cisd1	CDGSH iron sulfur domain 1	Sum of Sum_pept_count	0	0	1	2
		Max of Max_Dif_pept	0	0	1	2
Ly6d	lymphocyte antigen 6 complex, locus D	Sum of Sum_pept_count	0	0	2	2
		Max of Max_Dif_pept	0	0	2	2

Gene Symbol	Gene Description	Data	proximal	central	distal	total mucosa
Qdpr	quinoid dihydropteridine reductase	Sum of Sum_pept_count	0	4	0	2
		Max of Max_Dif_pept	0	2	0	2
Rpl21	ribosomal protein L21	Sum of Sum_pept_count	1	0	0	2
		Max of Max_Dif_pept	1	0	0	2
Rpl22	ribosomal protein L22	Sum of Sum_pept_count	3	2	1	2
		Max of Max_Dif_pept	2	2	1	2
Eif3e	eukaryotic translation initiation factor 3, subunit E	Sum of Sum_pept_count	0	0	1	2
		Max of Max_Dif_pept	0	0	1	2
Pycrl	pyrroline-5-carboxylate reductase-like	Sum of Sum_pept_count	2	2	2	3
		Max of Max_Dif_pept	2	2	2	2
EG637273	predicted gene, EG637273	Sum of Sum_pept_count	0	0	4	2
		Max of Max_Dif_pept	0	0	2	1
Rpl28	ribosomal protein L28	Sum of Sum_pept_count	0	0	0	2
		Max of Max_Dif_pept	0	0	0	2
Rpl30	ribosomal protein L30	Sum of Sum_pept_count	4	0	2	4
		Max of Max_Dif_pept	2	0	2	2
Rpl31	ribosomal protein L31	Sum of Sum_pept_count	7	0	3	4
		Max of Max_Dif_pept	2	0	1	2
Ptprc	protein tyrosine phosphatase, receptor type, C	Sum of Sum_pept_count	0	0	1	2
		Max of Max_Dif_pept	0	0	1	2
Lrrc40	leucine rich repeat containing 40	Sum of Sum_pept_count	0	0	0	2
		Max of Max_Dif_pept	0	0	0	2
Rpl36	ribosomal protein L36	Sum of Sum_pept_count	2	0	2	3
		Max of Max_Dif_pept	2	0	2	2
Rpl4	ribosomal protein L4	Sum of Sum_pept_count	2	0	0	1
		Max of Max_Dif_pept	2	0	0	1
Rpl7a	ribosomal protein L7a	Sum of Sum_pept_count	0	0	0	3
		Max of Max_Dif_pept	0	0	0	2
Hist1h2ag	histone cluster 1, H2ag	Sum of Sum_pept_count	2	3	2	3
		Max of Max_Dif_pept	1	2	1	2
Hist1h2bl	histone cluster 1, H2bl	Sum of Sum_pept_count	5	1	4	7
		Max of Max_Dif_pept	2	1	2	2
Hnrnpu	heterogeneous nuclear ribonucleoprotein U	Sum of Sum_pept_count	4	0	0	0
		Max of Max_Dif_pept	2	0	0	0
Lrrc1	leucine rich repeat containing 1	Sum of Sum_pept_count	0	0	3	0
		Max of Max_Dif_pept	0	0	2	0
Rps14	ribosomal protein S14	Sum of Sum_pept_count	6	1	3	3
		Max of Max_Dif_pept	2	1	2	2
Lrp1	low density lipoprotein receptor-related protein 1	Sum of Sum_pept_count	0	0	0	3
		Max of Max_Dif_pept	0	0	0	2
LOC665622	H2b histone family member	Sum of Sum_pept_count	5	1	4	7
		Max of Max_Dif_pept	2	1	2	2
Hist1h2bm	histone cluster 1, H2bm	Sum of Sum_pept_count	5	1	4	7
		Max of Max_Dif_pept	2	1	2	2
Eif1b	eukaryotic translation initiation factor 1B	Sum of Sum_pept_count	0	3	2	2
		Max of Max_Dif_pept	0	1	1	2
Epha6	Eph receptor A6	Sum of Sum_pept_count	0	0	2	0
		Max of Max_Dif_pept	0	0	2	0
LOC632454	similar to tumor protein, translationally-controlled 1	Sum of Sum_pept_count	0	6	0	3
		Max of Max_Dif_pept	0	2	0	2
Ptbp1	polypyrimidine tract binding protein 1	Sum of Sum_pept_count	0	0	0	3
		Max of Max_Dif_pept	0	0	0	2
Rps24	ribosomal protein S24	Sum of Sum_pept_count	3	1	2	2
		Max of Max_Dif_pept	2	1	1	1
Psmg3	proteasome (prosome, macropain) assembly chaperone 3	Sum of Sum_pept_count	0	0	0	2
		Max of Max_Dif_pept	0	0	0	2
Rps26	ribosomal protein S26	Sum of Sum_pept_count	4	2	6	5
		Max of Max_Dif_pept	2	2	2	2
Rps27	ribosomal protein S27	Sum of Sum_pept_count	0	2	0	1
		Max of Max_Dif_pept	0	2	0	1
Hist1h2af	histone cluster 1, H2af	Sum of Sum_pept_count	2	3	2	3
		Max of Max_Dif_pept	1	2	1	2
G6pd2	glucose-6-phosphate dehydrogenase 2	Sum of Sum_pept_count	0	1	2	0
		Max of Max_Dif_pept	0	1	2	0
Fcgrt	Fc receptor, IgG, alpha chain transporter	Sum of Sum_pept_count	0	0	0	2
		Max of Max_Dif_pept	0	0	0	2
Rps3a	ribosomal protein S3a	Sum of Sum_pept_count	2	0	0	1
		Max of Max_Dif_pept	2	0	0	1
Map2k3	mitogen-activated protein kinase kinase 3	Sum of Sum_pept_count	0	0	2	1
		Max of Max_Dif_pept	0	0	2	1

Gene Symbol	Gene Description	Data	proximal	central	distal	total mucosa
		Max of Max_Dif_pept	0	0	2	1
EG545267	predicted gene, EG545267	Sum of Sum_pept_count	0	0	2	3
		Max of Max_Dif_pept	0	0	2	2
G6pdx	glucose-6-phosphate dehydrogenase X-linked	Sum of Sum_pept_count	0	1	2	2
		Max of Max_Dif_pept	0	1	2	2
Rqcd1	rcd1 (required for cell differentiation) homolog 1 (S. pombe)	Sum of Sum_pept_count	0	0	0	2
		Max of Max_Dif_pept	0	0	0	2
Gmppa	GDP-mannose pyrophosphorylase A	Sum of Sum_pept_count	0	2	0	2
		Max of Max_Dif_pept	0	2	0	2
Rragb	Ras-related GTP binding B	Sum of Sum_pept_count	0	0	2	0
		Max of Max_Dif_pept	0	0	2	0
Diap1	diaphanous homolog 1 (Drosophila)	Sum of Sum_pept_count	0	0	0	2
		Max of Max_Dif_pept	0	0	0	2
Psmd2	proteasome (prosome, macropain) 26S subunit, non-ATPase, 2	Sum of Sum_pept_count	0	1	0	3
		Max of Max_Dif_pept	0	1	0	2
LOC554292	UbiE-YGHL1 fusion protein	Sum of Sum_pept_count	1	0	0	2
		Max of Max_Dif_pept	1	0	0	2
Galk1	galactokinase 1	Sum of Sum_pept_count	2	0	0	4
		Max of Max_Dif_pept	1	0	0	2
Rtn3	reticulon 3	Sum of Sum_pept_count	0	0	3	6
		Max of Max_Dif_pept	0	0	2	2
Hist1h2bn	histone cluster 1, H2bn	Sum of Sum_pept_count	5	1	4	7
		Max of Max_Dif_pept	2	1	2	2
S100a11	S100 calcium binding protein A11 (calgizzarin)	Sum of Sum_pept_count	1	3	4	1
		Max of Max_Dif_pept	1	2	2	1
Fbxo6	F-box protein 6	Sum of Sum_pept_count	2	1	1	1
		Max of Max_Dif_pept	2	1	1	1
Sar1a	SAR1 gene homolog A (S. cerevisiae)	Sum of Sum_pept_count	2	0	1	2
		Max of Max_Dif_pept	2	0	1	2
Eif4a2	eukaryotic translation initiation factor 4A2	Sum of Sum_pept_count	0	1	1	3
		Max of Max_Dif_pept	0	1	1	2
Sars	seryl-aminoacyl-tRNA synthetase	Sum of Sum_pept_count	0	2	3	2
		Max of Max_Dif_pept	0	2	2	2
Sbds	Shwachman-Bodian-Diamond syndrome homolog (human)	Sum of Sum_pept_count	1	0	2	2
		Max of Max_Dif_pept	1	0	2	2
Cul4b	cullin 4B	Sum of Sum_pept_count	0	0	0	2
		Max of Max_Dif_pept	0	0	0	2
Hist1h2ae	histone cluster 1, H2ae	Sum of Sum_pept_count	2	3	2	3
		Max of Max_Dif_pept	1	2	1	2
Hist1h2ac	histone cluster 1, H2ac	Sum of Sum_pept_count	2	3	2	3
		Max of Max_Dif_pept	1	2	1	2
Hist1h2ab	histone cluster 1, H2ab	Sum of Sum_pept_count	2	3	2	3
		Max of Max_Dif_pept	1	2	1	2
Scye1	small inducible cytokine subfamily E, member 1	Sum of Sum_pept_count	0	0	0	3
		Max of Max_Dif_pept	0	0	0	2
Sdcbp	syndecan binding protein	Sum of Sum_pept_count	0	1	2	1
		Max of Max_Dif_pept	0	1	2	1
Hist1h2bp	histone cluster 1, H2bp	Sum of Sum_pept_count	5	1	4	7
		Max of Max_Dif_pept	2	1	2	2
Sdha	succinate dehydrogenase complex, subunit A, flavoprotein (Fp)	Sum of Sum_pept_count	0	0	0	2
		Max of Max_Dif_pept	0	0	0	2
Sdhc	succinate dehydrogenase complex, subunit C, integral membrane protein	Sum of Sum_pept_count	0	0	0	2
		Max of Max_Dif_pept	0	0	0	2
Sec11a	SEC11 homolog A (S. cerevisiae)	Sum of Sum_pept_count	0	0	1	2
		Max of Max_Dif_pept	0	0	1	2
EG544954	predicted gene, EG544954	Sum of Sum_pept_count	0	0	0	2
		Max of Max_Dif_pept	0	0	0	2
Atp5j2	ATP synthase, H+ transporting, mitochondrial F0 complex, subunit f, isoform 2	Sum of Sum_pept_count	0	1	3	7
		Max of Max_Dif_pept	0	1	1	2
Sec22b	SEC22 vesicle trafficking protein homolog B (S. cerevisiae)	Sum of Sum_pept_count	0	0	1	2
		Max of Max_Dif_pept	0	0	1	2

Gene Symbol	Gene Description	Data	proximal	central	distal	total mucosa
Hist1h3a	histone cluster 1, H3a	Sum of Sum_pept_count	3	0	2	4
		Max of Max_Dif_pept	2	0	1	1
Psmb3	proteasome (prosome, macropain) subunit, beta type 3	Sum of Sum_pept_count	0	2	3	2
		Max of Max_Dif_pept	0	2	1	1
Litaf	LPS-induced TN factor	Sum of Sum_pept_count	3	0	2	0
		Max of Max_Dif_pept	2	0	1	0
Psmb2	proteasome (prosome, macropain) subunit, beta type 2	Sum of Sum_pept_count	1	1	2	2
		Max of Max_Dif_pept	1	1	2	2
Hist1h2aa	histone cluster 1, H2aa	Sum of Sum_pept_count	2	3	2	3
		Max of Max_Dif_pept	1	2	1	2
Gmppb	GDP-mannose pyrophosphorylase B	Sum of Sum_pept_count	0	2	0	1
		Max of Max_Dif_pept	0	2	0	1
Hnrnpk	heterogeneous nuclear ribonucleoprotein K	Sum of Sum_pept_count	0	2	1	5
		Max of Max_Dif_pept	0	1	1	2
Gmfg	glia maturation factor, gamma	Sum of Sum_pept_count	0	2	1	0
		Max of Max_Dif_pept	0	2	1	0
Mat2a	methionine adenosyltransferase II, alpha	Sum of Sum_pept_count	0	0	0	2
		Max of Max_Dif_pept	0	0	0	2
Serinc3	serine incorporator 3	Sum of Sum_pept_count	0	0	2	1
		Max of Max_Dif_pept	0	0	2	1
Bat1a	HLA-B-associated transcript 1A	Sum of Sum_pept_count	2	1	0	0
		Max of Max_Dif_pept	2	1	0	0
Hist1h3b	histone cluster 1, H3b	Sum of Sum_pept_count	3	0	2	4
		Max of Max_Dif_pept	2	0	1	1
Ddi2	DNA-damage inducible protein 2	Sum of Sum_pept_count	0	2	0	0
		Max of Max_Dif_pept	0	2	0	0
Serpina3k	serine (or cysteine) peptidase inhibitor, clade A, member 3K	Sum of Sum_pept_count	0	2	1	2
		Max of Max_Dif_pept	0	2	1	2
Acp1	acid phosphatase 1, soluble	Sum of Sum_pept_count	0	0	0	3
		Max of Max_Dif_pept	0	0	0	2
Psenen	presenilin enhancer 2 homolog (C. elegans)	Sum of Sum_pept_count	4	2	5	2
		Max of Max_Dif_pept	2	1	2	2
Pscd2	pleckstrin homology, Sec7 and coiled-coil domains 2	Sum of Sum_pept_count	0	2	0	2
		Max of Max_Dif_pept	0	1	0	2
Hist1h3c	histone cluster 1, H3c	Sum of Sum_pept_count	3	0	2	4
		Max of Max_Dif_pept	2	0	1	1
Sft2d2	SFT2 domain containing 2	Sum of Sum_pept_count	3	1	3	4
		Max of Max_Dif_pept	1	1	2	2
Crb3	crumbs homolog 3 (Drosophila)	Sum of Sum_pept_count	4	0	3	0
		Max of Max_Dif_pept	2	0	1	0
Hist1h1d	histone cluster 1, H1d	Sum of Sum_pept_count	3	0	1	0
		Max of Max_Dif_pept	2	0	1	0
Prss32	protease, serine, 32	Sum of Sum_pept_count	4	2	12	2
		Max of Max_Dif_pept	1	2	2	1
EG434460	predicted gene, EG434460	Sum of Sum_pept_count	0	2	3	1
		Max of Max_Dif_pept	0	1	2	1
Lincr	lung-inducible neuralized-related C3HC4 RING domain protein	Sum of Sum_pept_count	4	1	1	3
		Max of Max_Dif_pept	2	1	1	1
Prss2	protease, serine, 2	Sum of Sum_pept_count	0	0	3	0
		Max of Max_Dif_pept	0	0	2	0
Sirt2	sirtuin 2 (silent mating type information regulation 2, homolog) 2 (S. cerevisiae)	Sum of Sum_pept_count	0	0	1	3
		Max of Max_Dif_pept	0	0	1	2
Prpsap1	phosphoribosyl pyrophosphate synthetase-associated protein 1	Sum of Sum_pept_count	1	2	1	0
		Max of Max_Dif_pept	1	2	1	0
Epha7	Eph receptor A7	Sum of Sum_pept_count	0	0	2	0
		Max of Max_Dif_pept	0	0	2	0
Prosc	proline synthetase co-transcribed	Sum of Sum_pept_count	0	1	0	2
		Max of Max_Dif_pept	0	1	0	2
Hist1h3d	histone cluster 1, H3d	Sum of Sum_pept_count	3	0	2	4
		Max of Max_Dif_pept	2	0	1	1
Prkce	protein kinase C, epsilon	Sum of Sum_pept_count	4	2	5	0
		Max of Max_Dif_pept	1	1	2	0
Aqp7	aquaporin 7	Sum of Sum_pept_count	12	1	2	7
		Max of Max_Dif_pept	2	1	1	2
EG434426	predicted gene, EG434426	Sum of Sum_pept_count	0	0	0	3
		Max of Max_Dif_pept	0	0	0	2

Gene Symbol	Gene Description	Data	proximal	central	distal	total mucosa
Slc16a1	solute carrier family 16 (monocarboxylic acid transporters), member 1	Sum of Sum_pept_count	7	1	9	7
		Max of Max_Dif_pept	2	1	2	2
Slc1a1	solute carrier family 1 (neuronal/epithelial high affinity glutamate transporter, system Xag), member 1	Sum of Sum_pept_count	6	6	33	8
		Max of Max_Dif_pept	1	2	2	2
Slc1a5	solute carrier family 1 (neutral amino acid transporter), member 5	Sum of Sum_pept_count	1	0	8	2
		Max of Max_Dif_pept	1	0	2	1
Slc22a4	solute carrier family 22 (organic cation transporter), member 4	Sum of Sum_pept_count	3	5	6	2
		Max of Max_Dif_pept	2	2	2	2
Hist1h1c	histone cluster 1, H1c	Sum of Sum_pept_count	3	0	1	0
		Max of Max_Dif_pept	2	0	1	0
Farsa	phenylalanyl-tRNA synthetase, alpha subunit	Sum of Sum_pept_count	0	0	0	2
		Max of Max_Dif_pept	0	0	0	2
4921506J03Rik	RIKEN cDNA 4921506J03 gene	Sum of Sum_pept_count	0	0	0	2
		Max of Max_Dif_pept	0	0	0	2
Slc25a10	solute carrier family 25 (mitochondrial carrier, dicarboxylate transporter), member 10	Sum of Sum_pept_count	0	0	0	2
		Max of Max_Dif_pept	0	0	0	2
Fahd1	fumarylacetoacetate hydrolase domain containing 1	Sum of Sum_pept_count	1	0	0	2
		Max of Max_Dif_pept	1	0	0	2
EG433328	predicted gene, EG433328	Sum of Sum_pept_count	0	0	3	0
		Max of Max_Dif_pept	0	0	2	0
Cyb5b	cytochrome b5 type B	Sum of Sum_pept_count	0	0	0	3
		Max of Max_Dif_pept	0	0	0	2
Hsd17b4	hydroxysteroid (17-beta) dehydrogenase 4	Sum of Sum_pept_count	2	2	3	2
		Max of Max_Dif_pept	1	1	2	1
Lhfpl2	lipoma HMGIC fusion partner-like 2	Sum of Sum_pept_count	0	0	4	1
		Max of Max_Dif_pept	0	0	2	1
Ddx1	DEAD (Asp-Glu-Ala-Asp) box polypeptide 1	Sum of Sum_pept_count	0	1	0	3
		Max of Max_Dif_pept	0	1	0	2
Epha2	Eph receptor A2	Sum of Sum_pept_count	0	0	2	0
		Max of Max_Dif_pept	0	0	2	0
Aarsd1	alanyl-tRNA synthetase domain containing 1	Sum of Sum_pept_count	0	0	0	2
		Max of Max_Dif_pept	0	0	0	2
Amy1	amylase 1, salivary	Sum of Sum_pept_count	0	1	2	0
		Max of Max_Dif_pept	0	1	2	0
Cdc2a	cell division cycle 2 homolog A (S. pombe)	Sum of Sum_pept_count	0	0	1	2
		Max of Max_Dif_pept	0	0	1	2
EG386551	predicted gene, EG386551	Sum of Sum_pept_count	13	15	10	11
		Max of Max_Dif_pept	2	2	2	2
EG664891	predicted gene, EG664891	Sum of Sum_pept_count	4	2	6	5
		Max of Max_Dif_pept	2	2	2	2
Atp6v1e1	ATPase, H+ transporting, lysosomal V1 subunit E1	Sum of Sum_pept_count	1	0	4	3
		Max of Max_Dif_pept	1	0	2	2
Arl4a	ADP-ribosylation factor-like 4A	Sum of Sum_pept_count	2	0	0	0
		Max of Max_Dif_pept	2	0	0	0
Slc31a1	solute carrier family 31, member 1	Sum of Sum_pept_count	12	3	3	4
		Max of Max_Dif_pept	2	1	1	1
Agbl3	ATP/GTP binding protein-like 3	Sum of Sum_pept_count	0	0	0	3
		Max of Max_Dif_pept	0	0	0	2
Atp5l	ATP synthase, H+ transporting, mitochondrial F0 complex, subunit g	Sum of Sum_pept_count	0	0	3	4
		Max of Max_Dif_pept	0	0	1	2
Cdh1	cadherin 1	Sum of Sum_pept_count	1	1	5	6
		Max of Max_Dif_pept	1	1	1	2
Med21	mediator complex subunit 21	Sum of Sum_pept_count	4	2	6	5
		Max of Max_Dif_pept	2	2	2	2
Hdhd2	haloacid dehalogenase-like hydrolase domain containing 2	Sum of Sum_pept_count	0	0	0	4
		Max of Max_Dif_pept	0	0	0	2
Arpc5l	actin related protein 2/3 complex, subunit 5-like	Sum of Sum_pept_count	0	1	0	2
		Max of Max_Dif_pept	0	1	0	2

Gene Symbol	Gene Description	Data	proximal	central	distal	total mucosa
Prdx4	peroxiredoxin 4	Sum of Sum_pept_count	2	3	4	7
		Max of Max_Dif_pept	1	2	2	2
Gars	glycyl-tRNA synthetase	Sum of Sum_pept_count	0	2	1	1
		Max of Max_Dif_pept	0	2	1	1
1810011O10Rik	RIKEN cDNA 1810011O10 gene	Sum of Sum_pept_count	1	3	2	0
		Max of Max_Dif_pept	1	2	2	0
Ppp3r1	protein phosphatase 3, regulatory subunit B, alpha isoform (calcineurin B, type I)	Sum of Sum_pept_count	0	2	1	1
		Max of Max_Dif_pept	0	2	1	1
As3mt	arsenic (+3 oxidation state) methyltransferase	Sum of Sum_pept_count	0	2	0	0
		Max of Max_Dif_pept	0	2	0	0
Mettl7a1	methyltransferase like 7A1	Sum of Sum_pept_count	1	0	2	2
		Max of Max_Dif_pept	1	0	2	2
Hck	hemopoietic cell kinase	Sum of Sum_pept_count	5	4	1	7
		Max of Max_Dif_pept	2	2	1	2
2210010C17Rik	RIKEN cDNA 2210010C17 gene	Sum of Sum_pept_count	5	2	1	2
		Max of Max_Dif_pept	2	1	1	1
Hist1h3e	histone cluster 1, H3e	Sum of Sum_pept_count	3	0	2	4
		Max of Max_Dif_pept	2	0	1	1
Hist1h3f	histone cluster 1, H3f	Sum of Sum_pept_count	3	0	2	4
		Max of Max_Dif_pept	2	0	1	1
Gna12	guanine nucleotide binding protein, alpha 12	Sum of Sum_pept_count	5	4	5	6
		Max of Max_Dif_pept	2	2	2	2
Ppp1r7	protein phosphatase 1, regulatory (inhibitor) subunit 7	Sum of Sum_pept_count	0	2	2	0
		Max of Max_Dif_pept	0	2	2	0
Fadd	Fas (TNFRSF6)-associated via death domain	Sum of Sum_pept_count	2	0	2	4
		Max of Max_Dif_pept	1	0	2	2
Ppp1r16a	protein phosphatase 1, regulatory (inhibitor) subunit 16A	Sum of Sum_pept_count	3	2	0	0
		Max of Max_Dif_pept	2	2	0	0
EG241053	predicted gene, EG241053	Sum of Sum_pept_count	2	0	0	2
		Max of Max_Dif_pept	2	0	0	1
Cbara1	calcium binding atopy-related autoantigen 1	Sum of Sum_pept_count	0	0	0	3
		Max of Max_Dif_pept	0	0	0	2
Fabp4	fatty acid binding protein 4, adipocyte	Sum of Sum_pept_count	0	2	0	2
		Max of Max_Dif_pept	0	1	0	2
Cd47	CD47 antigen (Rh-related antigen, integrin-associated signal transducer)	Sum of Sum_pept_count	2	2	5	5
		Max of Max_Dif_pept	1	1	2	1
C530008M17Rik	RIKEN cDNA C530008M17 gene	Sum of Sum_pept_count	0	5	2	0
		Max of Max_Dif_pept	0	2	2	0
Slc7a15	solute carrier family 7 (cationic amino acid transporter, y+ system), member 15	Sum of Sum_pept_count	6	2	3	3
		Max of Max_Dif_pept	2	1	2	2
Ppl	periplakin	Sum of Sum_pept_count	0	0	3	0
		Max of Max_Dif_pept	0	0	2	0
Lamp1	lysosomal-associated membrane protein 1	Sum of Sum_pept_count	7	5	9	6
		Max of Max_Dif_pept	2	2	2	2
BC022651	cDNA sequence BC022651	Sum of Sum_pept_count	4	3	5	4
		Max of Max_Dif_pept	1	1	2	1
A130092J06Rik	RIKEN cDNA A130092J06 gene	Sum of Sum_pept_count	4	1	1	1
		Max of Max_Dif_pept	2	1	1	1
Slco6d1	solute carrier organic anion transporter family, member 6d1	Sum of Sum_pept_count	0	0	0	2
		Max of Max_Dif_pept	0	0	0	2
Smad1	MAD homolog 1 (Drosophila)	Sum of Sum_pept_count	3	1	2	0
		Max of Max_Dif_pept	2	1	1	0
Mgrn1	mahogunin, ring finger 1	Sum of Sum_pept_count	2	0	0	0
		Max of Max_Dif_pept	2	0	0	0
Depdc7	DEP domain containing 7	Sum of Sum_pept_count	1	0	3	1
		Max of Max_Dif_pept	1	0	2	1
Hist1h3g	histone cluster 1, H3g	Sum of Sum_pept_count	3	0	2	4
		Max of Max_Dif_pept	2	0	1	1
Mgst1	microsomal glutathione S-transferase 1	Sum of Sum_pept_count	1	0	1	3
		Max of Max_Dif_pept	1	0	1	2
Snap29	synaptosomal-associated protein 29	Sum of Sum_pept_count	3	2	3	4
		Max of Max_Dif_pept	2	1	2	2

Gene Symbol	Gene Description	Data	proximal	central	distal	total mucosa
Mgst3	microsomal glutathione S-transferase 3	Sum of Sum_pept_count	0	0	0	4
		Max of Max_Dif_pept	0	0	0	2
Pnpo	pyridoxine 5'-phosphate oxidase	Sum of Sum_pept_count	0	0	0	2
		Max of Max_Dif_pept	0	0	0	2
EG667806	predicted gene, EG667806	Sum of Sum_pept_count	4	0	3	0
		Max of Max_Dif_pept	2	0	2	0
Snx4	sorting nexin 4	Sum of Sum_pept_count	0	0	0	2
		Max of Max_Dif_pept	0	0	0	2
Hist1h3h	histone cluster 1, H3h	Sum of Sum_pept_count	3	0	2	4
		Max of Max_Dif_pept	2	0	1	1
Plxnb2	plexin B2	Sum of Sum_pept_count	1	0	6	0
		Max of Max_Dif_pept	1	0	2	0
Gas2	growth arrest specific 2	Sum of Sum_pept_count	0	2	1	1
		Max of Max_Dif_pept	0	2	1	1
Dmn	desmuslin	Sum of Sum_pept_count	8	6	14	6
		Max of Max_Dif_pept	1	1	2	1
Spata6	spermatogenesis associated 6	Sum of Sum_pept_count	0	0	0	3
		Max of Max_Dif_pept	0	0	0	2
Spcs2	signal peptidase complex subunit 2 homolog (S. cerevisiae)	Sum of Sum_pept_count	0	0	3	1
		Max of Max_Dif_pept	0	0	2	1
Spcs3	signal peptidase complex subunit 3 homolog (S. cerevisiae)	Sum of Sum_pept_count	0	0	2	2
		Max of Max_Dif_pept	0	0	2	2
Golm1	golgi membrane protein 1	Sum of Sum_pept_count	0	0	5	0
		Max of Max_Dif_pept	0	0	2	0
Spint2	serine protease inhibitor, Kunitz type 2	Sum of Sum_pept_count	2	0	4	3
		Max of Max_Dif_pept	1	0	2	1
6330409N04Rik	RIKEN cDNA 6330409N04 gene	Sum of Sum_pept_count	0	0	0	2
		Max of Max_Dif_pept	0	0	0	2
Spsb2	splA/ryanodine receptor domain and SOCS box containing 2	Sum of Sum_pept_count	0	0	2	0
		Max of Max_Dif_pept	0	0	2	0
Plekhb2	pleckstrin homology domain containing, family B (evectins) member 2	Sum of Sum_pept_count	2	3	2	3
		Max of Max_Dif_pept	2	1	1	1
Abcb4	ATP-binding cassette, sub-family B (MDR/TAP), member 4	Sum of Sum_pept_count	0	3	0	1
		Max of Max_Dif_pept	0	2	0	1
Ssr1	signal sequence receptor, alpha	Sum of Sum_pept_count	0	0	3	1
		Max of Max_Dif_pept	0	0	2	1
Mobkl2b	MOB1, Mps One Binder kinase activator-like 2B (yeast)	Sum of Sum_pept_count	0	0	0	2
		Max of Max_Dif_pept	0	0	0	2
Ssu72	Ssu72 RNA polymerase II CTD phosphatase homolog (yeast)	Sum of Sum_pept_count	0	0	0	2
		Max of Max_Dif_pept	0	0	0	2
Mobkl3	MOB1, Mps One Binder kinase activator-like 3 (yeast)	Sum of Sum_pept_count	0	0	2	3
		Max of Max_Dif_pept	0	0	1	2
Hnrnph1	heterogeneous nuclear ribonucleoprotein H1	Sum of Sum_pept_count	0	1	2	1
		Max of Max_Dif_pept	0	1	2	1
Mpi	mannose phosphate isomerase	Sum of Sum_pept_count	0	1	0	2
		Max of Max_Dif_pept	0	1	0	2
Stard3	START domain containing 3	Sum of Sum_pept_count	1	0	2	1
		Max of Max_Dif_pept	1	0	2	1
Stard4	StAR-related lipid transfer (START) domain containing 4	Sum of Sum_pept_count	1	0	0	2
		Max of Max_Dif_pept	1	0	0	2
Pip5k1b	phosphatidylinositol-4-phosphate 5-kinase, type 1 beta	Sum of Sum_pept_count	1	3	1	2
		Max of Max_Dif_pept	1	2	1	2
Stip1	stress-induced phosphoprotein 1	Sum of Sum_pept_count	0	1	0	2
		Max of Max_Dif_pept	0	1	0	2
Hist1h3i	histone cluster 1, H3i	Sum of Sum_pept_count	3	0	2	4
		Max of Max_Dif_pept	2	0	1	1
Cast	calpastatin	Sum of Sum_pept_count	0	0	0	2
		Max of Max_Dif_pept	0	0	0	2
Krt78	keratin 78	Sum of Sum_pept_count	0	2	2	2
		Max of Max_Dif_pept	0	2	2	2
Stk38	serine/threonine kinase 38	Sum of Sum_pept_count	1	0	0	3
		Max of Max_Dif_pept	1	0	0	2
Stk38l	serine/threonine kinase 38 like	Sum of Sum_pept_count	0	0	0	2
		Max of Max_Dif_pept	0	0	0	2

Gene Symbol	Gene Description	Data	proximal	central	distal	total mucosa
Atp2b4	ATPase, Ca++ transporting, plasma membrane 4	Sum of Sum_pept_count	2	0	4	1
		Max of Max_Dif_pept	1	0	2	1
Stub1	STIP1 homology and U-Box containing protein 1	Sum of Sum_pept_count	0	2	2	0
		Max of Max_Dif_pept	0	2	2	0
Cops2	COP9 (constitutive photomorphogenic) homolog, subunit 2 (Arabidopsis thaliana)	Sum of Sum_pept_count	0	1	1	2
		Max of Max_Dif_pept	0	1	1	2
Atp2b2	ATPase, Ca++ transporting, plasma membrane 2	Sum of Sum_pept_count	0	1	2	0
		Max of Max_Dif_pept	0	1	2	0
Pgk2	phosphoglycerate kinase 2	Sum of Sum_pept_count	0	2	0	0
		Max of Max_Dif_pept	0	2	0	0
Hadha	hydroxyacyl-Coenzyme A dehydrogenase/3-ketoacyl-Coenzyme A thiolase/enoyl-Coenzyme A hydratase (trifunctional protein), alpha subunit	Sum of Sum_pept_count	0	0	0	2
		Max of Max_Dif_pept	0	0	0	2
Actn2	actinin alpha 2	Sum of Sum_pept_count	0	1	9	4
		Max of Max_Dif_pept	0	1	2	2
Abcb11	ATP-binding cassette, sub-family B (MDR/TAP), member 11	Sum of Sum_pept_count	0	4	1	7
		Max of Max_Dif_pept	0	1	1	2
Actn3	actinin alpha 3	Sum of Sum_pept_count	0	1	5	4
		Max of Max_Dif_pept	0	1	2	2
Stxbp5	syntaxin binding protein 5 (tomosyn)	Sum of Sum_pept_count	2	0	0	0
		Max of Max_Dif_pept	2	0	0	0
H3f3b	H3 histone, family 3B	Sum of Sum_pept_count	3	0	2	4
		Max of Max_Dif_pept	2	0	1	1
Bdh1	3-hydroxybutyrate dehydrogenase, type 1	Sum of Sum_pept_count	0	0	0	2
		Max of Max_Dif_pept	0	0	0	2
Sult1d1	sulfotransferase family 1D, member 1	Sum of Sum_pept_count	0	1	1	4
		Max of Max_Dif_pept	0	1	1	2
Gnai1	guanine nucleotide binding protein (G protein), alpha inhibiting 1	Sum of Sum_pept_count	5	0	2	8
		Max of Max_Dif_pept	2	0	1	2
1700022A21Rik	RIKEN cDNA 1700022A21 gene	Sum of Sum_pept_count	0	2	0	0
		Max of Max_Dif_pept	0	2	0	0
Sypl	synaptophysin-like protein	Sum of Sum_pept_count	5	0	7	6
		Max of Max_Dif_pept	1	0	2	2
2900024O10Rik	RIKEN cDNA 2900024O10 gene	Sum of Sum_pept_count	1	0	2	0
		Max of Max_Dif_pept	1	0	2	0
Tagln	transgelin	Sum of Sum_pept_count	0	4	2	1
		Max of Max_Dif_pept	0	2	1	1
H3f3a	H3 histone, family 3A	Sum of Sum_pept_count	3	0	2	4
		Max of Max_Dif_pept	2	0	1	1
Abcb7	ATP-binding cassette, sub-family B (MDR/TAP), member 7	Sum of Sum_pept_count	0	0	4	0
		Max of Max_Dif_pept	0	0	2	0
EG668182	predicted gene, EG668182	Sum of Sum_pept_count	3	0	1	1
		Max of Max_Dif_pept	2	0	1	1
2010109I03Rik	RIKEN cDNA 2010109I03 gene	Sum of Sum_pept_count	0	1	2	3
		Max of Max_Dif_pept	0	1	1	2
Tbc1d2b	TBC1 domain family, member 2B	Sum of Sum_pept_count	0	0	2	0
		Max of Max_Dif_pept	0	0	2	0
Tceb1	transcription elongation factor B (SIII), polypeptide 1	Sum of Sum_pept_count	0	3	2	4
		Max of Max_Dif_pept	0	2	2	2
Epb4.1l3	erythrocyte protein band 4.1-like 3	Sum of Sum_pept_count	1	2	3	4
		Max of Max_Dif_pept	1	2	2	2
Cldn15	claudin 15	Sum of Sum_pept_count	4	2	4	4
		Max of Max_Dif_pept	2	2	2	2
Eef1g	eukaryotic translation elongation factor 1 gamma	Sum of Sum_pept_count	1	0	0	4
		Max of Max_Dif_pept	1	0	0	2
Aldh1l1	aldehyde dehydrogenase 1 family, member L1	Sum of Sum_pept_count	0	1	4	1
		Max of Max_Dif_pept	0	1	2	1
Krt6b	keratin 6B	Sum of Sum_pept_count	0	0	2	0
		Max of Max_Dif_pept	0	0	2	0
Asns	asparagine synthetase	Sum of Sum_pept_count	0	0	1	2
		Max of Max_Dif_pept	0	0	1	2

Gene Symbol	Gene Description	Data	proximal	central	distal	total mucosa
Them2	thioesterase superfamily member 2	Sum of Sum_pept_count	0	0	0	3
		Max of Max_Dif_pept	0	0	0	2
Cd63	Cd63 antigen	Sum of Sum_pept_count	2	2	3	2
		Max of Max_Dif_pept	1	1	2	1
Hist2h2aa2	histone cluster 2, H2aa2	Sum of Sum_pept_count	2	3	2	3
		Max of Max_Dif_pept	1	2	1	2
Msn	moesin	Sum of Sum_pept_count	3	0	1	1
		Max of Max_Dif_pept	2	0	1	1
Tjp2	tight junction protein 2	Sum of Sum_pept_count	0	6	0	0
		Max of Max_Dif_pept	0	2	0	0
1810049H19Rik	RIKEN cDNA 1810049H19 gene	Sum of Sum_pept_count	0	3	1	3
		Max of Max_Dif_pept	0	2	1	2
Agpat1	1-acylglycerol-3-phosphate O-acyltransferase 1 (lysophosphatidic acid acyltransferase, alpha)	Sum of Sum_pept_count	0	0	0	2
		Max of Max_Dif_pept	0	0	0	2
Tm4sf20	transmembrane 4 L six family member 20	Sum of Sum_pept_count	6	2	4	0
		Max of Max_Dif_pept	2	1	2	0
Mtch2	mitochondrial carrier homolog 2 (C. elegans)	Sum of Sum_pept_count	0	0	0	7
		Max of Max_Dif_pept	0	0	0	2
Etf1	eukaryotic translation termination factor 1	Sum of Sum_pept_count	2	1	0	0
		Max of Max_Dif_pept	2	1	0	0
Apol9b	apolipoprotein L 9b	Sum of Sum_pept_count	4	0	0	6
		Max of Max_Dif_pept	2	0	0	2
H2-Q1	histocompatibility 2, Q region locus 1	Sum of Sum_pept_count	2	2	0	3
		Max of Max_Dif_pept	2	2	0	2
Gimap3	GTPase, IMAP family member 3	Sum of Sum_pept_count	0	0	2	0
		Max of Max_Dif_pept	0	0	2	0
Hist2h2bb	histone cluster 2, H2bb	Sum of Sum_pept_count	5	1	4	7
		Max of Max_Dif_pept	2	1	2	2
4930544G11Rik	RIKEN cDNA 4930544G11 gene	Sum of Sum_pept_count	3	6	1	1
		Max of Max_Dif_pept	2	2	1	1
Hspa1l	heat shock protein 1-like	Sum of Sum_pept_count	3	1	1	2
		Max of Max_Dif_pept	1	1	1	2
Pdhb	pyruvate dehydrogenase (lipoamide) beta	Sum of Sum_pept_count	0	0	0	3
		Max of Max_Dif_pept	0	0	0	2
Tmem106a	transmembrane protein 106A	Sum of Sum_pept_count	4	0	2	1
		Max of Max_Dif_pept	2	0	1	1
Gimap5	GTPase, IMAP family member 5	Sum of Sum_pept_count	0	0	2	0
		Max of Max_Dif_pept	0	0	2	0
1700003O08Rik	RIKEN cDNA 1700003O08 gene	Sum of Sum_pept_count	1	2	0	0
		Max of Max_Dif_pept	1	2	0	0
EG668192	predicted gene, EG668192	Sum of Sum_pept_count	4	0	3	0
		Max of Max_Dif_pept	2	0	2	0
B4galnt1	beta-1,4-N-acetyl-galactosaminyl transferase 1	Sum of Sum_pept_count	0	0	3	1
		Max of Max_Dif_pept	0	0	2	1
Pcyt1b	phosphate cytidylyltransferase 1, choline, beta isoform	Sum of Sum_pept_count	0	2	0	0
		Max of Max_Dif_pept	0	2	0	0
Krt20	keratin 20	Sum of Sum_pept_count	3	3	3	3
		Max of Max_Dif_pept	2	2	1	2
Hspa2	heat shock protein 2	Sum of Sum_pept_count	3	1	1	2
		Max of Max_Dif_pept	1	1	1	2
Mthfd1	methylenetetrahydrofolate dehydrogenase (NADP+ dependent), methenyltetrahydrofolate cyclohydrolase, formyltetrahydrofolate synthase	Sum of Sum_pept_count	0	0	0	3
		Max of Max_Dif_pept	0	0	0	2
Tmem59	transmembrane protein 59	Sum of Sum_pept_count	0	0	2	3
		Max of Max_Dif_pept	0	0	1	2
Atp6v1d	ATPase, H+ transporting, lysosomal V1 subunit D	Sum of Sum_pept_count	0	0	0	3
		Max of Max_Dif_pept	0	0	0	2
Tmod3	tropomodulin 3	Sum of Sum_pept_count	0	3	0	0
		Max of Max_Dif_pept	0	2	0	0
Pcna	proliferating cell nuclear antigen	Sum of Sum_pept_count	0	0	0	2
		Max of Max_Dif_pept	0	0	0	2
Cyp2c66	cytochrome P450, family 2, subfamily c, polypeptide 66	Sum of Sum_pept_count	2	0	1	0
		Max of Max_Dif_pept	2	0	1	0
Tnni1	troponin I, skeletal, slow 1	Sum of Sum_pept_count	1	0	0	3
		Max of Max_Dif_pept	1	0	0	2

Gene Symbol	Gene Description	Data	proximal	central	distal	total mucosa
Mtpn	myotrophin	Sum of Sum_pept_count	1	2	1	2
		Max of Max_Dif_pept	1	2	1	2
Gnpnat1	glucosamine-phosphate N-acetyltransferase 1	Sum of Sum_pept_count	0	2	2	1
		Max of Max_Dif_pept	0	2	2	1
Tom1l2	target of myb1-like 2 (chicken)	Sum of Sum_pept_count	2	6	4	2
		Max of Max_Dif_pept	1	2	2	2
Pcbd1	pterin 4 alpha carbinolamine dehydratase/dimerization cofactor of hepatocyte nuclear factor 1 alpha (TCF1) 1	Sum of Sum_pept_count	0	2	1	2
		Max of Max_Dif_pept	0	2	1	2
Anp32c	acidic leucine-rich nuclear phosphoprotein 32 family member C	Sum of Sum_pept_count	0	2	0	0
		Max of Max_Dif_pept	0	2	0	0
Park7	Parkinson disease (autosomal recessive, early onset) 7	Sum of Sum_pept_count	3	4	3	3
		Max of Max_Dif_pept	2	2	2	2
Tpt1	tumor protein, translationally-controlled 1	Sum of Sum_pept_count	0	6	0	3
		Max of Max_Dif_pept	0	2	0	2
Hist2h2be	histone cluster 2, H2be	Sum of Sum_pept_count	3	1	4	7
		Max of Max_Dif_pept	1	1	2	2
Dbnl	drebrin-like	Sum of Sum_pept_count	2	3	2	11
		Max of Max_Dif_pept	1	2	1	2
Pak4	p21 (CDKN1A)-activated kinase 4	Sum of Sum_pept_count	1	4	3	2
		Max of Max_Dif_pept	1	2	2	1
Cd74	CD74 antigen (invariant polypeptide of major histocompatibility complex, class II antigen-associated)	Sum of Sum_pept_count	0	1	3	1
		Max of Max_Dif_pept	0	1	2	1
Erlin2	ER lipid raft associated 2	Sum of Sum_pept_count	0	0	1	3
		Max of Max_Dif_pept	0	0	1	2
Try10	trypsin 10	Sum of Sum_pept_count	13	15	10	11
		Max of Max_Dif_pept	2	2	2	2
Atp6v0c	ATPase, H+ transporting, lysosomal V0 subunit C	Sum of Sum_pept_count	1	0	2	2
		Max of Max_Dif_pept	1	0	2	2
EG668319	predicted gene, EG668319	Sum of Sum_pept_count	1	0	0	2
		Max of Max_Dif_pept	1	0	0	2
Tspan13	tetraspanin 13	Sum of Sum_pept_count	1	0	4	4
		Max of Max_Dif_pept	1	0	2	2
Copz1	coatomer protein complex, subunit zeta 1	Sum of Sum_pept_count	0	1	0	2
		Max of Max_Dif_pept	0	1	0	2
Tspan3	tetraspanin 3	Sum of Sum_pept_count	1	0	0	2
		Max of Max_Dif_pept	1	0	0	2
Erlin1	ER lipid raft associated 1	Sum of Sum_pept_count	0	0	1	3
		Max of Max_Dif_pept	0	0	1	2
Hist2h3b	histone cluster 2, H3b	Sum of Sum_pept_count	3	0	2	4
		Max of Max_Dif_pept	2	0	1	1
OTTMUSG0000001	predicted gene, OTTMUSG00000012957	Sum of Sum_pept_count	2	0	2	3
		Max of Max_Dif_pept	2	0	2	2
Hsph1	heat shock 105kDa/110kDa protein 1	Sum of Sum_pept_count	0	0	1	2
		Max of Max_Dif_pept	0	0	1	2
OTTMUSG0000000	predicted gene, OTTMUSG00000008659	Sum of Sum_pept_count	5	5	3	2
		Max of Max_Dif_pept	2	2	1	1
2810459M11Rik	RIKEN cDNA 2810459M11 gene	Sum of Sum_pept_count	5	1	4	4
		Max of Max_Dif_pept	1	1	2	1
OTTMUSG0000000	predicted gene, OTTMUSG00000007855	Sum of Sum_pept_count	3	0	2	4
		Max of Max_Dif_pept	2	0	1	1
OTTMUSG0000000	predicted gene, OTTMUSG00000005885	Sum of Sum_pept_count	2	5	0	2
		Max of Max_Dif_pept	2	2	0	2
Hist3h2a	histone cluster 3, H2a	Sum of Sum_pept_count	2	3	2	3
		Max of Max_Dif_pept	1	2	1	2
Kdr	kinase insert domain protein receptor	Sum of Sum_pept_count	0	0	0	2
		Max of Max_Dif_pept	0	0	0	2
Kdelr1	KDEL (Lys-Asp-Glu-Leu) endoplasmic reticulum protein retention receptor 1	Sum of Sum_pept_count	0	0	0	2
		Max of Max_Dif_pept	0	0	0	2
Tubb1	tubulin, beta 1	Sum of Sum_pept_count	1	1	0	4
		Max of Max_Dif_pept	1	1	0	2

Gene Symbol	Gene Description	Data	proximal	central	distal	total mucosa
Cox5b	cytochrome c oxidase, subunit Vb	Sum of Sum_pept_count	1	0	1	4
		Max of Max_Dif_pept	1	0	1	2
H2afv	H2A histone family, member V	Sum of Sum_pept_count	3	3	2	6
		Max of Max_Dif_pept	1	2	1	2
Kank1	KN motif and ankyrin repeat domains 1	Sum of Sum_pept_count	0	0	0	3
		Max of Max_Dif_pept	0	0	0	2
9130404D14Rik	RIKEN cDNA 9130404D14 gene	Sum of Sum_pept_count	0	1	3	1
		Max of Max_Dif_pept	0	1	2	1
Aars	alanyl-tRNA synthetase	Sum of Sum_pept_count	0	0	0	4
		Max of Max_Dif_pept	0	0	0	2
Ostb	organic solute transporter beta	Sum of Sum_pept_count	1	1	7	2
		Max of Max_Dif_pept	1	1	2	1
Oit1	oncoprotein induced transcript 1	Sum of Sum_pept_count	0	0	4	2
		Max of Max_Dif_pept	0	0	2	1
Cldn7	claudin 7	Sum of Sum_pept_count	7	5	17	9
		Max of Max_Dif_pept	2	2	2	1
Twf1	twinfilin, actin-binding protein, homolog 1 (Drosophila)	Sum of Sum_pept_count	1	2	2	3
		Max of Max_Dif_pept	1	2	2	2
Dsg1b	desmoglein 1 beta	Sum of Sum_pept_count	0	0	3	1
		Max of Max_Dif_pept	0	0	2	1
Txndc17	thioredoxin domain containing 17	Sum of Sum_pept_count	0	5	3	2
		Max of Max_Dif_pept	0	2	2	1
Numbl	numb-like	Sum of Sum_pept_count	1	1	2	1
		Max of Max_Dif_pept	1	1	2	1
Txnl1	thioredoxin-like 1	Sum of Sum_pept_count	1	3	0	1
		Max of Max_Dif_pept	1	2	0	1
Itpka	inositol 1,4,5-trisphosphate 3-kinase A	Sum of Sum_pept_count	4	5	1	3
		Max of Max_Dif_pept	2	2	1	2
EG668559	predicted gene, EG668559	Sum of Sum_pept_count	0	0	0	3
		Max of Max_Dif_pept	0	0	0	2
Nudcd3	NudC domain containing 3	Sum of Sum_pept_count	0	2	3	1
		Max of Max_Dif_pept	0	1	2	1
Nt5e	5' nucleotidase, ecto	Sum of Sum_pept_count	3	1	2	0
		Max of Max_Dif_pept	1	1	2	0
Nt5c	5',3'-nucleotidase, cytosolic	Sum of Sum_pept_count	1	2	2	4
		Max of Max_Dif_pept	1	2	2	2
D10Wsu52e	DNA segment, Chr 10, Wayne State University 52, expressed	Sum of Sum_pept_count	0	2	0	1
		Max of Max_Dif_pept	0	2	0	1
Npepps	aminopeptidase puromycin sensitive	Sum of Sum_pept_count	0	1	1	3
		Max of Max_Dif_pept	0	1	1	2
Ube2l3	ubiquitin-conjugating enzyme E2L 3	Sum of Sum_pept_count	1	2	1	2
		Max of Max_Dif_pept	1	1	1	2
2310008M10Rik	RIKEN cDNA 2310008M10 gene	Sum of Sum_pept_count	0	0	1	2
		Max of Max_Dif_pept	0	0	1	2
Amn	amnionless	Sum of Sum_pept_count	0	1	4	1
		Max of Max_Dif_pept	0	1	2	1
2010106G01Rik	RIKEN cDNA 2010106G01 gene	Sum of Sum_pept_count	3	0	2	1
		Max of Max_Dif_pept	2	0	2	1
Npc1	Niemann Pick type C1	Sum of Sum_pept_count	9	7	12	5
		Max of Max_Dif_pept	2	1	2	1
Ubie	UbiE-YGHL1 fusion protein	Sum of Sum_pept_count	1	0	0	2
		Max of Max_Dif_pept	1	0	0	2
Dfna5h	deafness, autosomal dominant 5 homolog (human)	Sum of Sum_pept_count	0	0	1	2
		Max of Max_Dif_pept	0	0	1	2
Npas2	neuronal PAS domain protein 2	Sum of Sum_pept_count	2	0	3	2
		Max of Max_Dif_pept	2	0	2	2
Itga3	integrin alpha 3	Sum of Sum_pept_count	3	1	3	4
		Max of Max_Dif_pept	2	1	2	2
Ephb4	Eph receptor B4	Sum of Sum_pept_count	0	0	2	0
		Max of Max_Dif_pept	0	0	2	0
Clpb	ClpB caseinolytic peptidase B homolog (E. coli)	Sum of Sum_pept_count	7	0	3	4
		Max of Max_Dif_pept	2	0	1	2
Cacybp	calcyclin binding protein	Sum of Sum_pept_count	0	4	2	1
		Max of Max_Dif_pept	0	2	1	1
Hist3h2ba	histone cluster 3, H2ba	Sum of Sum_pept_count	3	1	4	7
		Max of Max_Dif_pept	1	1	2	2
Ephb3	Eph receptor B3	Sum of Sum_pept_count	0	0	2	0
		Max of Max_Dif_pept	0	0	2	0
Dsg1a	desmoglein 1 alpha	Sum of Sum_pept_count	0	0	3	1
		Max of Max_Dif_pept	0	0	2	1
Hnrnpa3	heterogeneous nuclear ribonucleoprotein A3	Sum of Sum_pept_count	0	0	0	3
		Max of Max_Dif_pept	0	0	0	2

233

Gene Symbol	Gene Description	Data	proximal	central	distal	total mucosa
Hist3h2bb	histone cluster 3, H2bb	Sum of Sum_pept_count	3	1	4	7
		Max of Max_Dif_pept	1	1	2	2
Nedd4l	neural precursor cell expressed, developmentally down-regulated gene 4-like	Sum of Sum_pept_count	8	5	4	2
		Max of Max_Dif_pept	2	1	2	1
Nedd4	neural precursor cell expressed, developmentally down-regulated gene 4	Sum of Sum_pept_count	2	5	4	2
		Max of Max_Dif_pept	1	1	2	1
Uqcrh	ubiquinol-cytochrome c reductase hinge protein	Sum of Sum_pept_count	0	0	1	2
		Max of Max_Dif_pept	0	0	1	2
Ndufb7	NADH dehydrogenase (ubiquinone) 1 beta subcomplex, 7	Sum of Sum_pept_count	2	0	1	6
		Max of Max_Dif_pept	2	0	1	2
Ndufb3	NADH dehydrogenase (ubiquinone) 1 beta subcomplex 3	Sum of Sum_pept_count	0	0	0	3
		Max of Max_Dif_pept	0	0	0	2
Usp12	ubiquitin specific peptidase 12	Sum of Sum_pept_count	1	3	0	0
		Max of Max_Dif_pept	1	2	0	0
Usp46	ubiquitin specific peptidase 46	Sum of Sum_pept_count	1	3	0	0
		Max of Max_Dif_pept	1	2	0	0
Vamp2	vesicle-associated membrane protein 2	Sum of Sum_pept_count	5	2	5	1
		Max of Max_Dif_pept	2	1	2	1
Hkdc1	hexokinase domain containing 1	Sum of Sum_pept_count	1	1	1	2
		Max of Max_Dif_pept	1	1	1	2
Vamp5	vesicle-associated membrane protein 5	Sum of Sum_pept_count	2	0	0	1
		Max of Max_Dif_pept	2	0	0	1
Ick	intestinal cell kinase	Sum of Sum_pept_count	2	0	3	1
		Max of Max_Dif_pept	2	0	2	1
Abr	active BCR-related gene	Sum of Sum_pept_count	1	2	0	1
		Max of Max_Dif_pept	1	2	0	1
Dopey2	dopey family member 2	Sum of Sum_pept_count	0	0	2	0
		Max of Max_Dif_pept	0	0	2	0
Ap2s1	adaptor-related protein complex 2, sigma 1 subunit	Sum of Sum_pept_count	1	3	2	4
		Max of Max_Dif_pept	1	2	2	2
Dok4	docking protein 4	Sum of Sum_pept_count	2	0	3	0
		Max of Max_Dif_pept	2	0	2	0
Copg2	coatomer protein complex, subunit gamma 2	Sum of Sum_pept_count	0	1	2	2
		Max of Max_Dif_pept	0	1	2	2
Ndufa8	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 8	Sum of Sum_pept_count	0	0	0	2
		Max of Max_Dif_pept	0	0	0	2
AA986860	expressed sequence AA986860	Sum of Sum_pept_count	0	5	2	0
		Max of Max_Dif_pept	0	2	2	0
Bcr	breakpoint cluster region homolog	Sum of Sum_pept_count	1	2	1	0
		Max of Max_Dif_pept	1	2	1	0
Napg	N-ethylmaleimide sensitive fusion protein attachment protein gamma	Sum of Sum_pept_count	0	2	3	1
		Max of Max_Dif_pept	0	2	2	1
Emd	emerin	Sum of Sum_pept_count	0	3	2	2
		Max of Max_Dif_pept	0	1	1	2
Myo5a	myosin Va	Sum of Sum_pept_count	0	0	0	2
		Max of Max_Dif_pept	0	0	0	2
Vps29	vacuolar protein sorting 29 (S. pombe)	Sum of Sum_pept_count	0	1	1	2
		Max of Max_Dif_pept	0	1	1	2
Ipo5	importin 5	Sum of Sum_pept_count	0	0	3	0
		Max of Max_Dif_pept	0	0	2	0
Nap1l4	nucleosome assembly protein 1-like 4	Sum of Sum_pept_count	0	2	0	0
		Max of Max_Dif_pept	0	2	0	0
Gpc4	glypican 4	Sum of Sum_pept_count	0	0	2	0
		Max of Max_Dif_pept	0	0	2	0
Ass1	argininosuccinate synthetase 1	Sum of Sum_pept_count	0	0	0	2
		Max of Max_Dif_pept	0	0	0	2
Entpd2	ectonucleoside triphosphate diphosphohydrolase 2	Sum of Sum_pept_count	6	1	2	3
		Max of Max_Dif_pept	2	1	2	1
Ephb2	Eph receptor B2	Sum of Sum_pept_count	0	0	2	0
		Max of Max_Dif_pept	0	0	2	0

Gene Symbol	Gene Description	Data	proximal	central	distal	total mucosa
Cmbl	carboxymethylenebutenolidase-like (Pseudomonas)	Sum of Sum_pept_count	0	1	1	2
		Max of Max_Dif_pept	0	1	1	2
Gnb4	guanine nucleotide binding protein (G protein), beta 4	Sum of Sum_pept_count	0	3	0	0
		Max of Max_Dif_pept	0	2	0	0
Wdr61	WD repeat domain 61	Sum of Sum_pept_count	0	0	0	2
		Max of Max_Dif_pept	0	0	0	2
Xdh	xanthine dehydrogenase	Sum of Sum_pept_count	0	0	0	5
		Max of Max_Dif_pept	0	0	0	2
Car3	carbonic anhydrase 3	Sum of Sum_pept_count	0	0	0	2
		Max of Max_Dif_pept	0	0	0	2
Eif1	eukaryotic translation initiation factor 1	Sum of Sum_pept_count	0	3	2	2
		Max of Max_Dif_pept	0	1	1	2
Xpo1	exportin 1, CRM1 homolog (yeast)	Sum of Sum_pept_count	0	0	0	2
		Max of Max_Dif_pept	0	0	0	2
Nagk	N-acetylglucosamine kinase	Sum of Sum_pept_count	0	0	0	2
		Max of Max_Dif_pept	0	0	0	2
Aco1	aconitase 1	Sum of Sum_pept_count	0	0	0	2
		Max of Max_Dif_pept	0	0	0	2
Abi1	abl-interactor 1	Sum of Sum_pept_count	0	4	1	1
		Max of Max_Dif_pept	0	2	1	1
Acot7	acyl-CoA thioesterase 7	Sum of Sum_pept_count	0	1	2	0
		Max of Max_Dif_pept	0	1	2	0
Nadsyn1	NAD synthetase 1	Sum of Sum_pept_count	0	0	0	2
		Max of Max_Dif_pept	0	0	0	2
Ephb1	Eph receptor B1	Sum of Sum_pept_count	0	0	2	0
		Max of Max_Dif_pept	0	0	2	0
3110043O21Rik	RIKEN cDNA 3110043O21 gene	Sum of Sum_pept_count	2	0	1	1
		Max of Max_Dif_pept	2	0	1	1
Calr	calreticulin	Sum of Sum_pept_count	0	0	1	3
		Max of Max_Dif_pept	0	0	1	2
Hnrnpa1	heterogeneous nuclear ribonucleoprotein A1	Sum of Sum_pept_count	0	0	0	2
		Max of Max_Dif_pept	0	0	0	2
Zfp119	zinc finger protein 119	Sum of Sum_pept_count	1	0	0	2
		Max of Max_Dif_pept	1	0	0	2
Myo5c	myosin VC	Sum of Sum_pept_count	0	0	0	2
		Max of Max_Dif_pept	0	0	0	2
Dnajb6	DnaJ (Hsp40) homolog, subfamily B, member 6	Sum of Sum_pept_count	2	0	1	2
		Max of Max_Dif_pept	2	0	1	2
Sept8	septin 8	Sum of Sum_pept_count	0	1	1	4
		Max of Max_Dif_pept	0	1	1	2
B2m	beta-2 microglobulin	Sum of Sum_pept_count	5	4	4	4
		Max of Max_Dif_pept	2	1	1	2

Appendix B: Sample comparison by precursor ion signal intensity

B1. Filenames convention

Technical step	Band No	Given Name	Original study name
Injection replicates	Band 2	Pool_wBBM_1_2_a Pool_wBBM_1_2_b Pool_wBBM_1_2_c	ET72_sameGel_BM_mix_2_3_a ET72_sameGel_BM_mix_2_3_b ET72_sameGel_BM_mix_2_3_c
	Band 9	Pool_wBBM_1_9_a Pool_wBBM_1_9_b Pool_wBBM_1_9_c	ET72_sameGel_BM_mix_9_3_a ET72_sameGel_BM_mix_9_3_b ET72_sameGel_BM_mix_9_3_c
Same gel variation	Band 2	SameGel_pool_wBBM_1_2 SameGel_pool_wBBM_2_2 SameGel_pool_wBBM_3_2	ET72_sameGel_BM_mix_2_1 ET72_sameGel_BM_mix_2_2 ET72_sameGel_BM_mix_2_3_a
	Band 9	SameGel_pool_wBBM_1_9 SameGel_pool_wBBM_2_9 SameGel_pool_wBBM_3_9	ET72_sameGel_BM_mix_9_1 ET72_sameGel_BM_mix_9_2 ET72_sameGel_BM_mix_9_3_a
	Band 11	SameGel_pool_wBBM_1_11 SameGel_pool_wBBM_2_11 SameGel_pool_wBBM_3_11	ET72_sameGel_BM_mix_11_1 ET72_sameGel_BM_mix_11_2 ET72_sameGel_BM_mix_11_3_a
Different gel variation	Band 2	Difgel_pool_wBBM_A_2 Difgel_pool_wBBM_B_2 Difgel_pool_wBBM_C_2	ET72_difgel_BM_mix_2_1 ET72_difGel_BM_mix_2_2 ET72_sameGel_BM_mix_2_2
	Band 9	Difgel_pool_wBBM_A_9 Difgel_pool_wBBM_B_9 Difgel_pool_wBBM_C_9	ET72_difgel_BM_mix_9_1 ET72_difGel_BM_mix_9_2 ET72_sameGel_BM_mix_9_2
	Band 11	Difgel_pool_wBBM_A_11 Difgel_pool_wBBM_B_11 Difgel_pool_wBBM_C_11	ET72_difgel_BM_mix_11_1 ET72_difGel_BM_mix_11_2 ET72_sameGel_BM_mix_11_2
Preparation variation	Band 2	PV_wBBM_1_2 PV_wBBM_2_2 PV_wBBM_3_2	ET72_PV_BM_1_2 ET72_PV_BM_2_2 ET72_PV_BM_3_2
	Band 9	PV_wBBM_1_9 PV_wBBM_2_9 PV_wBBM_3_9	ET72_PV_BM_1_9 ET72_PV_BM_2_9 ET72_PV_BM_3_9
	Band 11	PV_wBBM_1_11 PV_wBBM_2_11 PV_wBBM_3_11	ET72_PV_BM_1_11 ET72_PV_BM_2_11 ET72_PV_BM_3_11
Total variation	Band 2	TV_wBBM_1_2 TV_wBBM_2_2 TV_wBBM_3_2	ET65_TV_BM_1_2 ET65_TV_BM_2_2 ET65_TV_BM_3_2
	Band 9	TV_wBBM_1_9 TV_wBBM_2_9 TV_wBBM_3_9	ET65_TV_BM_1_9 ET65_TV_BM_2_9 ET65_TV_BM_3_9
	Band 11	TV_wBBM_1_11 TV_wBBM_2_11 TV_wBBM_3_11	ET65_TV_BM_1_11 ET65_TV_BM_2_11 ET65_TV_BM_3_11

B2. Analytical time scale for the sample measurement used in the process variation analysis

Process variation measurement times

For naming convention, please refer to appendix B1.

B3. Scatter plots and Spearmann correlation values for all the sample comparisons described in the section 4.5.3.2.2

B.3.1 Injection replicates

Injection replicates for band 2

A

B
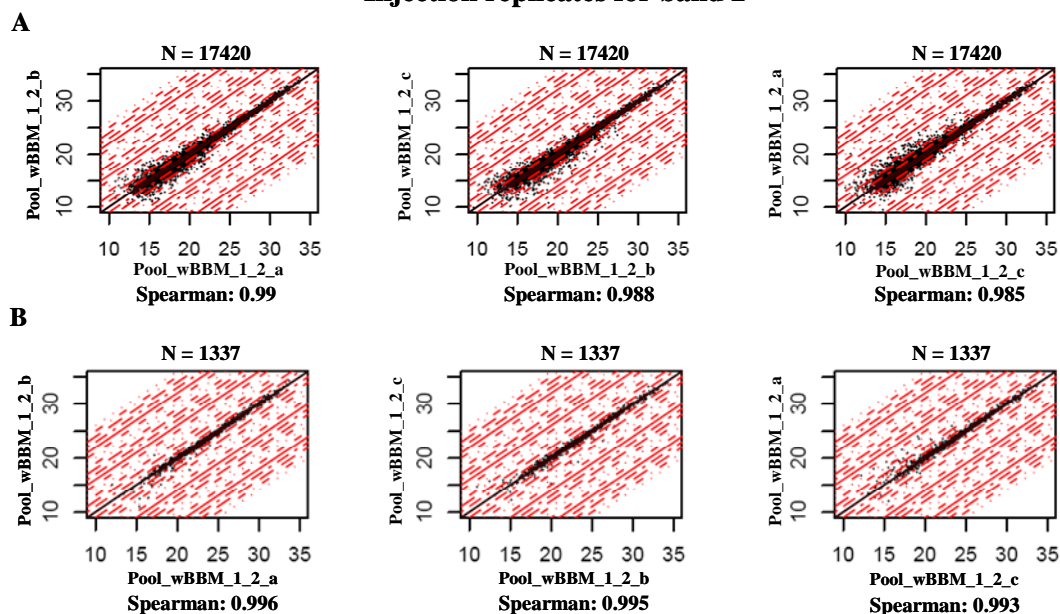
Injection replicates for band 9

A

B
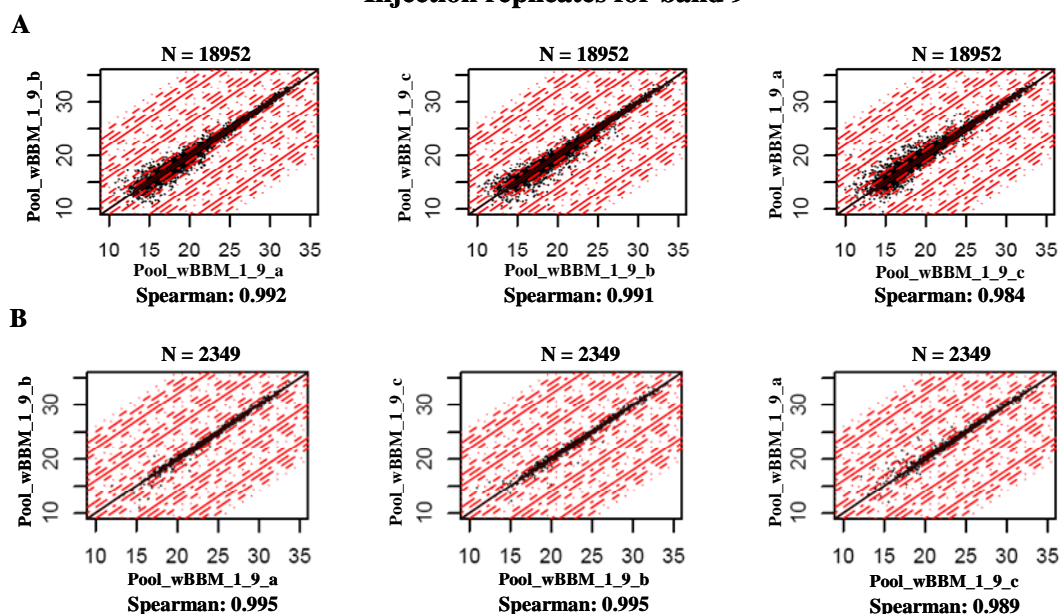
Scatter plot representations of the injection replicates of the bands 2 and 9. Panel A: Scatter plots of all the common MS signals between the replicates. Panel B: Scatter plots of the MS signals that correspond to successful MS/MS measurements. N is the number of common signals. The Spearman value reflects the similarity of the signal intensities between the compared samples (ideal case Spearman correlation=1). Each red dot line parallel to the diagonal represents a two fold difference, i.e. across three lines, the fold change is 2*2*2=8. Accordingly, the axes represent an arbitrary mass spectrometric intensity (in counts) in log2 units. All sample denominations are as described in Fig. 4.21.
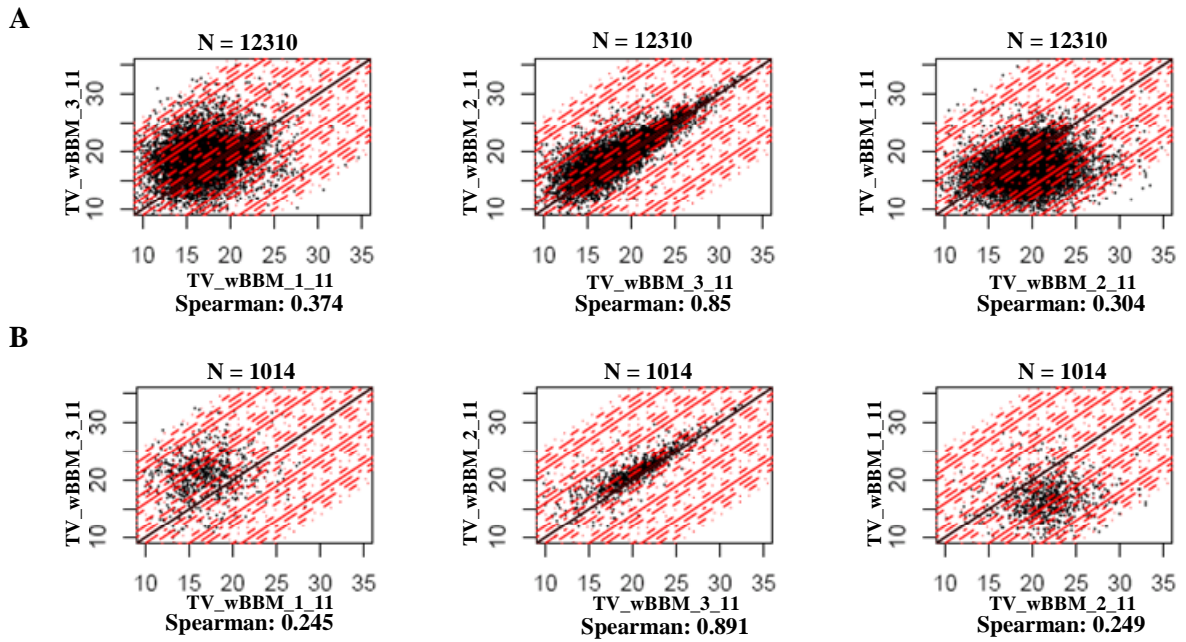
B.3.2 "Same gel variation" replicates

Same gel variation for band 2
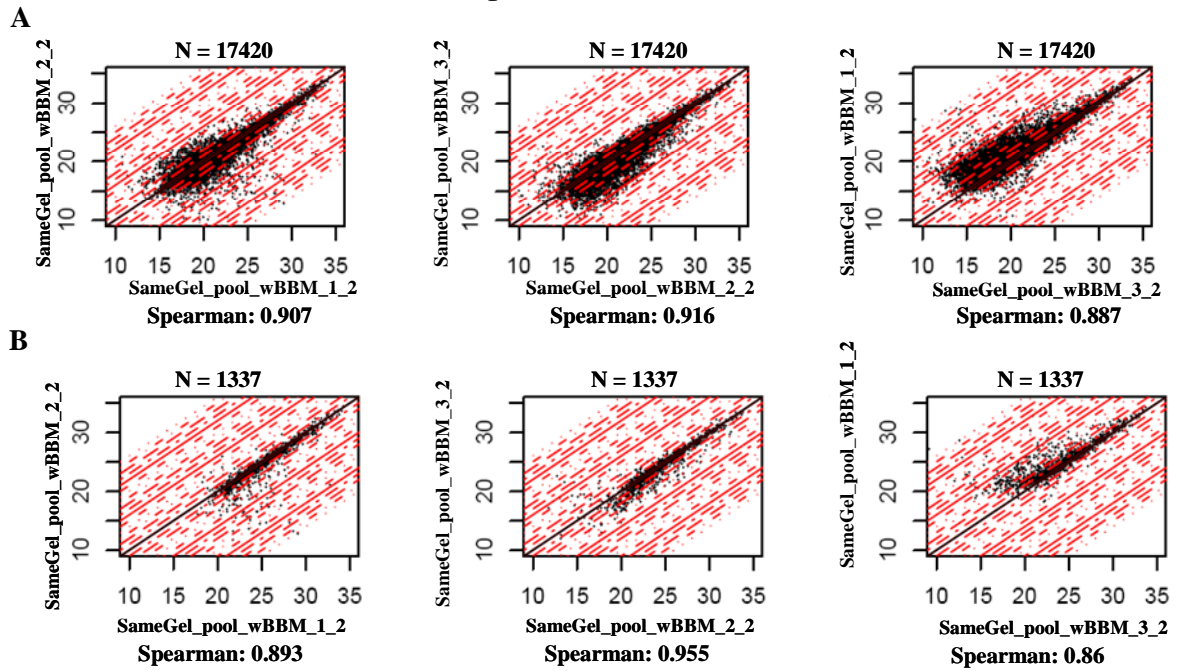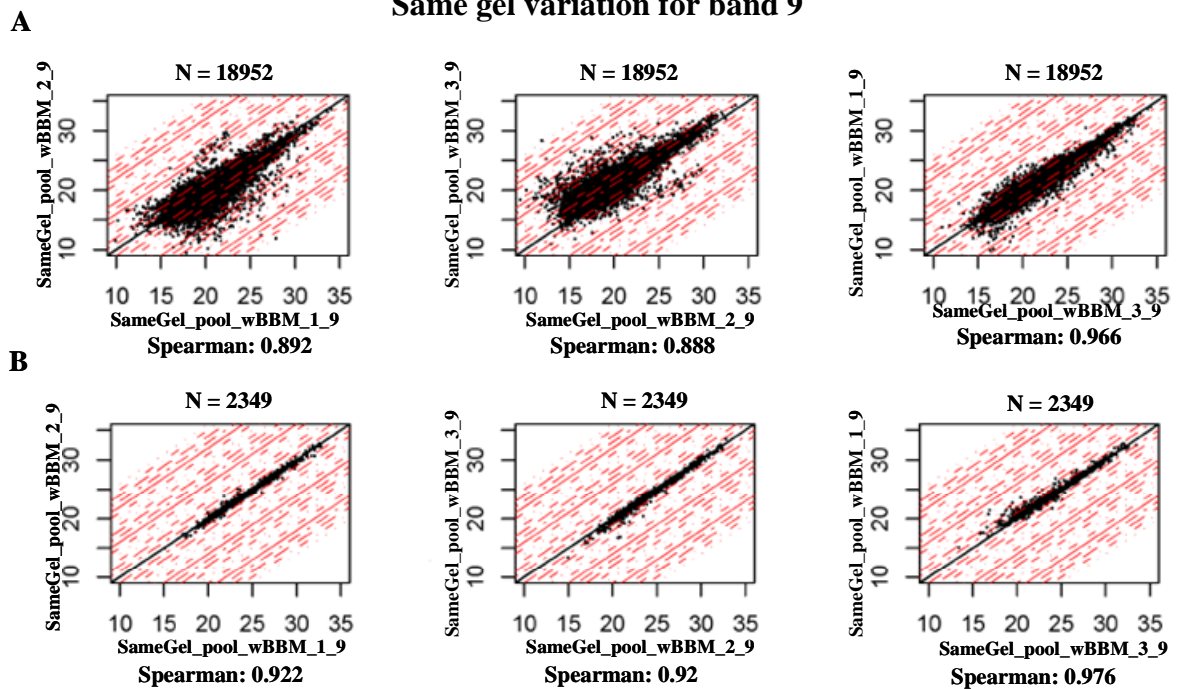
A

N = 17420
Spearman: 0.907

N = 17420
Spearman: 0.916

N = 17420
Spearman: 0.887

B

N = 1337
Spearman: 0.893

N = 1337
Spearman: 0.955

N = 1337
Spearman: 0.86

Same gel variation for band 9

A

N = 18952
Spearman: 0.892

N = 18952
Spearman: 0.888

N = 18952
Spearman: 0.966

B

N = 2349
Spearman: 0.922

N = 2349
Spearman: 0.92

N = 2349
Spearman: 0.976

Same gel variation for band 11

A

B
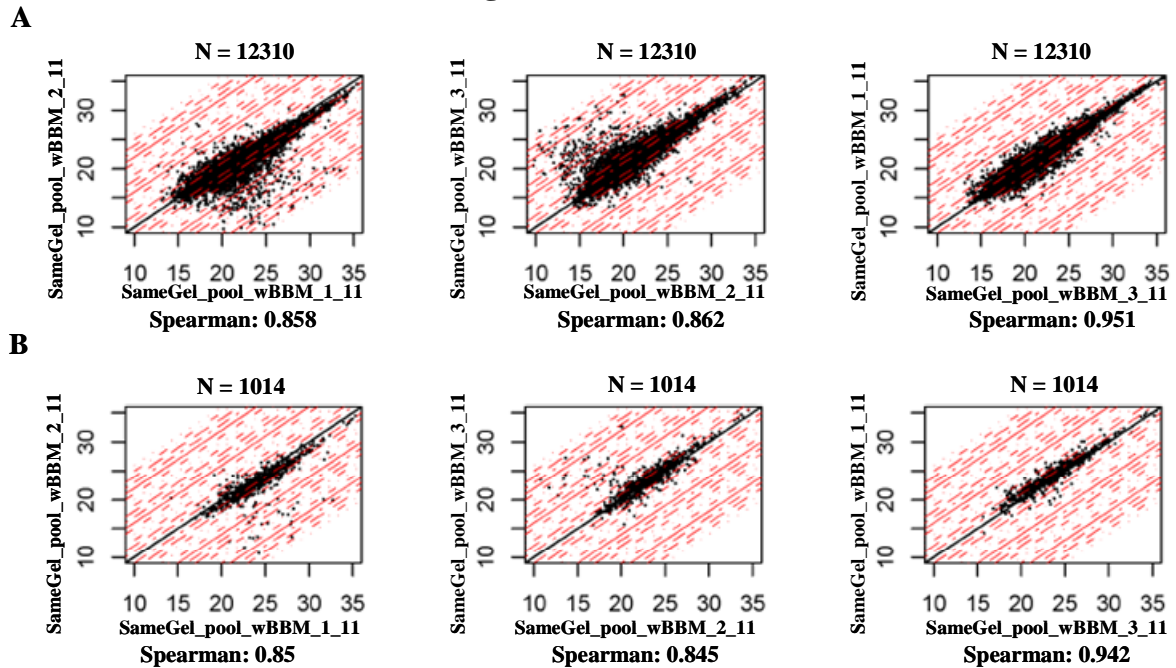
*Scatter plot representations of the variability for the band 2, 9 and 11, cut horizontally from adjacent identical lanes. **Panel A:** Scatter plots of all the common MS signals between the replicates. **Panel B:** Scatter plots of the MS signals that correspond to successful MS/MS measurements. N is the number of common signals. The Spearman value reflects the similarity of the signal intensities between the compared samples (ideal case Spearman correlation=1). Each red dot line parallel to the diagonal represents a two fold difference, i.e. across three lines, the fold change is 2*2*2=8. Accordingly, the axes represent an arbitrary mass spectrometric intensity (in counts) in log2 units. All sample denominations are as described in Fig. 4.21.*

Different gel variation for band 2

A

N = 17420 | N = 17420 | N = 17420
Spearman: 0.854 | Spearman: 0.945 | Spearman: 0841

B

N = 1337 | N = 1337 | N = 1337
Spearman: 0817 | Spearman: 0.964 | Spearman: 0.814

Different gel variation for band 9

A

N = 18952 | N = 18952 | N = 18952
Spearman: 0.955 | Spearman: 0.966 | Spearman: 0.954

B

N = 2349 | N = 2349 | N = 2349
Spearman: 0.955 | Spearman: 0.964 | Spearman: 0.954

Different gel variation for band 11

*Scatter plot representations of the band 2, 9 and 11 from identical BBM samples, loaded in three different gels. **Panel A:** Scatter plots of all the common MS signals between the replicates. **Panel B:** Scatter plots of the MS signals that correspond to successful MS/MS measurements. N is the number of common signals. The Spearman value reflects the similarity of the signal intensities between the compared samples (ideal case Spearman correlation=1). Each red dot line parallel to the diagonal represents a two fold difference, i.e. across three lines, the fold change is 2*2*2=8. Accordingly, the axes represent an arbitrary mass spectrometric intensity (in counts) in log2 units. All sample denominations are as described in Fig. 4.21.*

B.3.4 "Preparation variation"replicates

Preparation variation for band 2
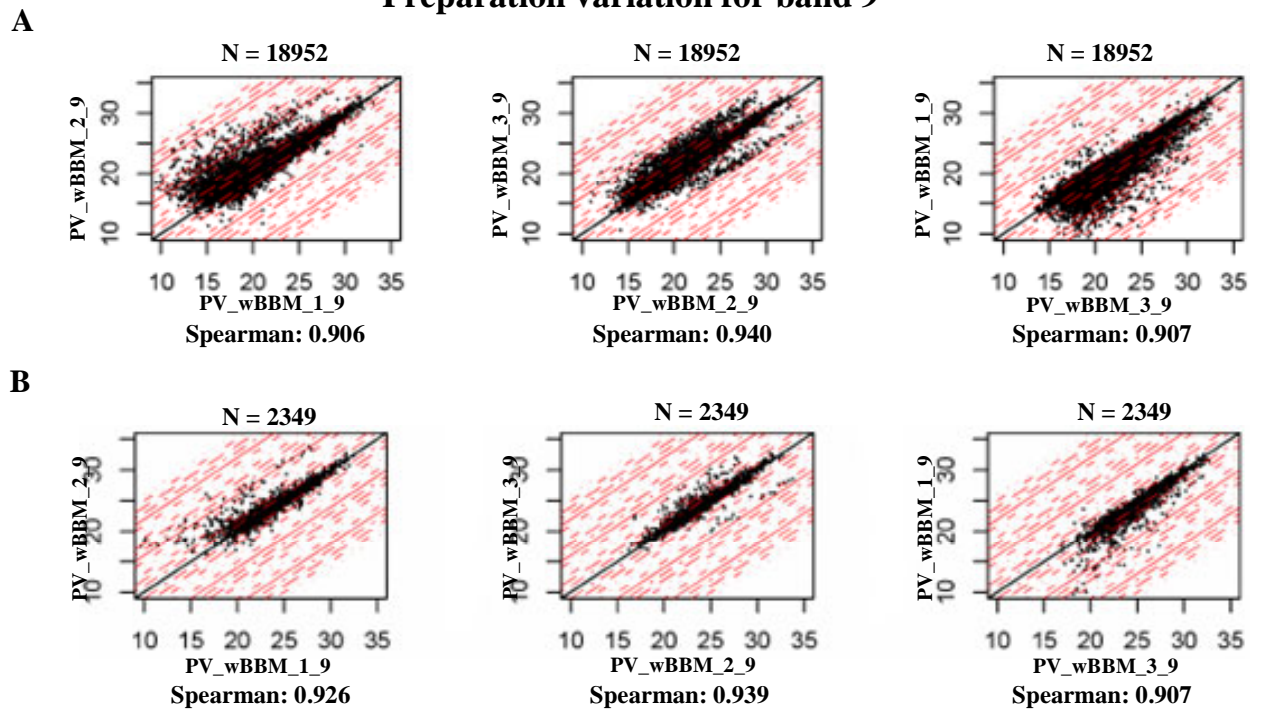
A

N = 17420

Spearman: 0.964

N = 17420

Spearman: 0.960

N = 17420

Spearman: 0.941

B

N = 1337

Spearman: 0.971

N = 1337

Spearman: 0.960

N = 1337

Spearman: 0.956

Preparation variation for band 9

A

N = 18952

Spearman: 0.906

N = 18952

Spearman: 0.940

N = 18952

Spearman: 0.907

B

N = 2349

Spearman: 0.926

N = 2349

Spearman: 0.939

N = 2349

Spearman: 0.907

Preparation variation for band 11

A

N = 12310   N = 12310   N = 12310

Spearman: 0.946   Spearman: 0.887   Spearman: 0.853

B

N = 1014   N = 1014   N = 1014

Spearman: 0.951   Spearman: 0.819   Spearman: 0.794
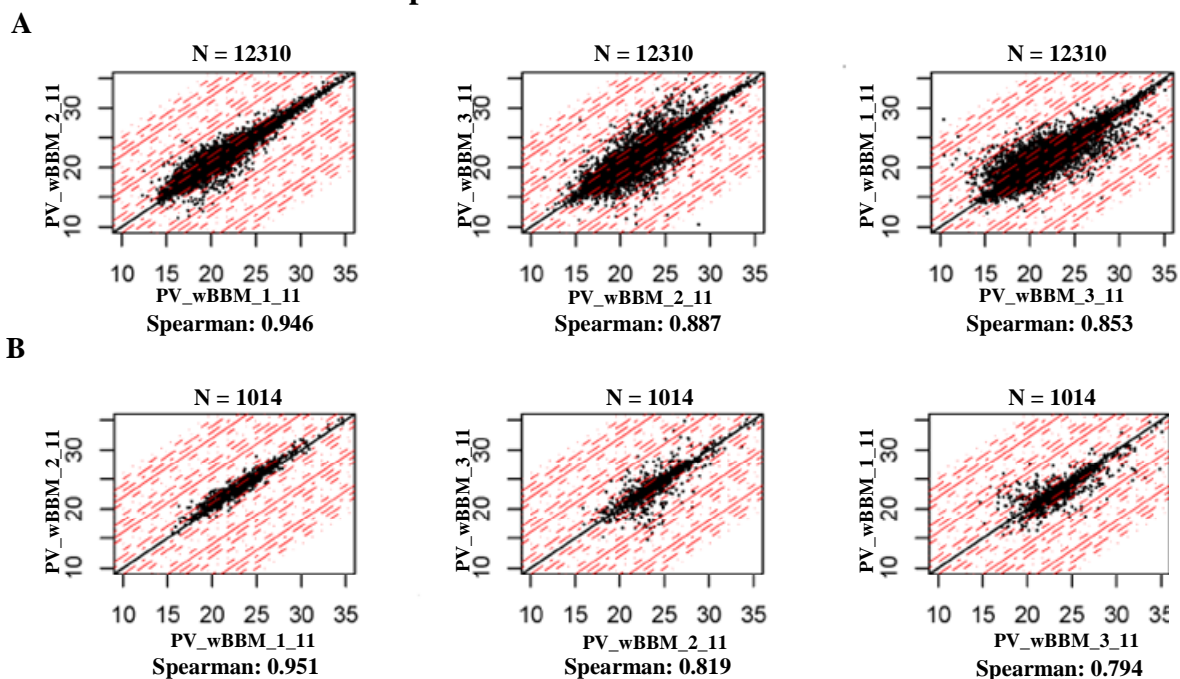
*Scatter plot representations of the preparation variability for the band 2, 9 and 11 from three BBM technical replicate preparations. **Panel A:** Scatter plots of all the common MS signals between the replicates. **Panel B:** Scatter plots of the MS signals that correspond to successful MS/MS measurements. N is the number of common signals. The Spearman value reflects the similarity of the signal intensities between the compared samples (ideal case Spearman correlation=1). Each red dot line parallel to the diagonal represents a two fold difference, i.e. across three lines, the fold change is 2*2*2=8. Accordingly, the axis represents an arbitrary mass spectrometric intensity (in counts) in log2 units. All sample denominations are as described in Fig. 4.21.*

B.3.5 "Total variation" replicates

Total variation for band 2
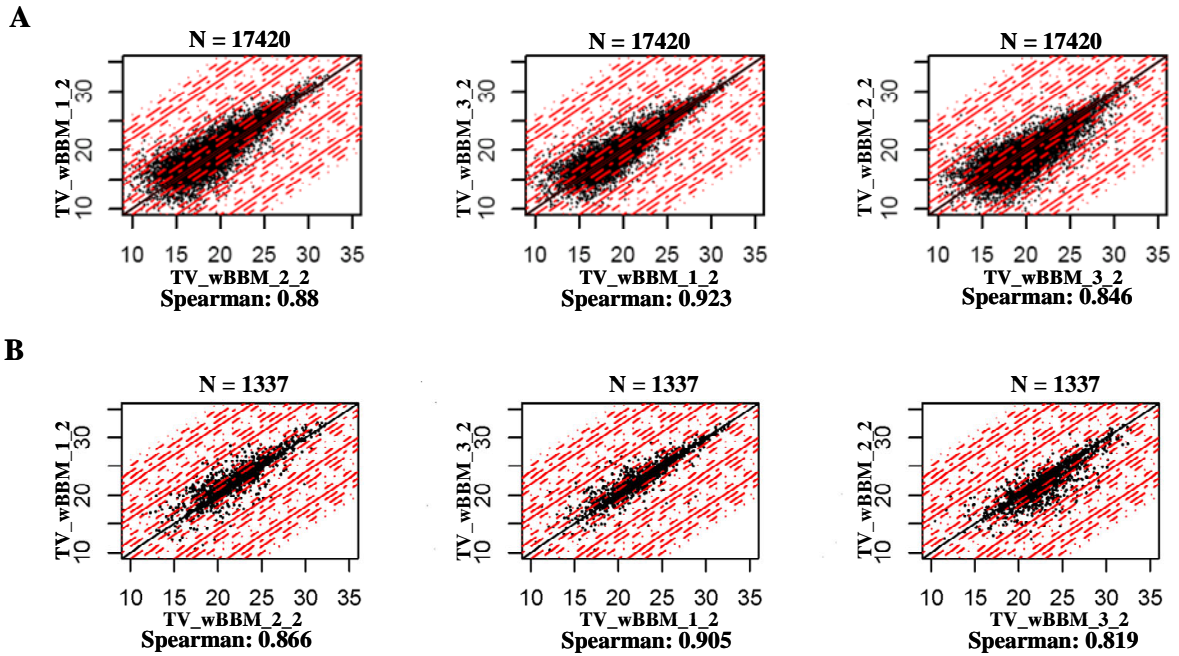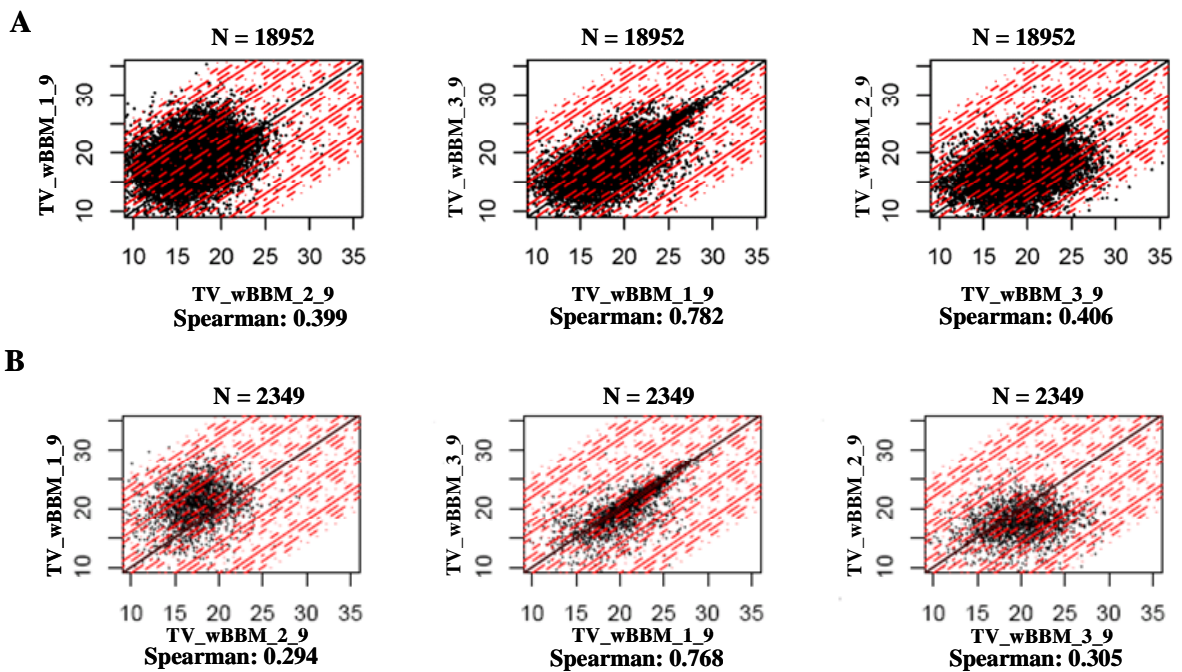
A

N = 17420
Spearman: 0.88

N = 17420
Spearman: 0.923

N = 17420
Spearman: 0.846

B

N = 1337
Spearman: 0.866

N = 1337
Spearman: 0.905

N = 1337
Spearman: 0.819

Total variation for band 9

A

N = 18952
Spearman: 0.399

N = 18952
Spearman: 0.782

N = 18952
Spearman: 0.406

B

N = 2349
Spearman: 0.294

N = 2349
Spearman: 0.768

N = 2349
Spearman: 0.305

Total variation for band 11

Scatter plot representations of the total variability for the bands 2, 9 and 11 from three BBM technical replicate preparations, randomly analyzed. Panel A: *Scatter plots of all the common MS signals between the replicates.* ***Panel B:*** *Scatter plots of the MS signals that correspond to successful MS/MS measurements. N is the number of common signals. The Spearman value reflects the similarity of the signal intensities between the compared samples (ideal case Spearman correlation=1). Each red dot line parallel to the diagonal represents a two fold difference, i.e. across three lines, the fold change is 2*2*2=8. Accordingly, the axes represent an arbitrary mass spectrometric intensity (in counts) in log2 units. All sample denominations are as described in Fig. 4.21*

Appendix C: Separation of hydrophobic peptides by hydrophilic-interaction liquid chromatography

C.1 Introduction

Reversed-Phase Liquid Chromatography (RPLC) has become an indispensable technique for the separation and purification of peptides using hydrophobic stationary phases, such as octyl- or octadecyl-derivatized silica beads. In RPLC, a solute molecule binds to an immobilized hydrophobic molecule in a polar solvent. This partitioning occurs as a result of the solute molecule tending to have hydrophobic patches at its surface and binding occuring via those patches to the matrix. A buffer of increasing hydrophobicity is used to dissociate the bound molecule at a point at which the hydrophobic interaction between the exposed patches and the immobilized matrix is less favorable than the interaction between the bound molecule and the solvent. The molecule releases from the matrix and elutes. One of the major advantages of the RPLC technique, apart from its extensive resolving power, is the availability of volatile mobile phases, of which aqueous trifluoroacetic acid (TFA), acetic acid or formic acid–acetonitrile (ACN) are the most frequently employed (123). Thus, RPLC can be directly coupled to an electrospray ionization interface for the online analysis of molecules using mass spectrometry.

Although a powerful separation mode, a major limitation of RPLC is the lack of adequate retention of polar molecules. In 1990, the name Hydrophilic-Interaction Chromatography (HILIC) was coined by Alpert (120) for the separation of hydrophilic substances such as proteins, peptides and nucleic acids using polar stationary phase, such as polyhydroxyethyl aspartamide or polyhydroxysulfotamide aspartamide. In HILIC, polar peptides are retained by the polar stationary phase in a hydrophobic solvent. The solute molecule is then eluted from the chromatographic beads by increasing the polarity of the mobile phase, for example by increasing the proportion of water or by adding salt. Similarly to normal phase liquid chromatography (or in opposite to the trends observed in RPLC) peptide retention time increases with the polarity (hydrophilicity) of the peptide and critically depends on the polarity of the stationary phase and on the initial polarity of the mobile phase (124-126). It has been considered that the interaction forces governing selectivity in HILIC are polar in origin, encompassing both hydrogen bonding, the extent of which depends upon the acidity or basicity (in the Lewis sense) of the solutes, and dipole–dipole interaction, which is dependent upon the dipole moments and polarizability of molecules (124). Similarly to RPLC, HILIC can be directly coupled to an electrospray ionization interface if the salt concentration

is kept to a minimum to avoid interference with the ionization process. However, so far, HILIC has been suggested as an alternative to cation-exchange chromatography for the separation of polar peptides, such as glycopeptides, phosphopeptides and modified histone peptides (121). It is then typically coupled off-line with a RPLC-MS for peptide characterization.

In contrast to most studies, where HILIC has been used to chromatography very polar peptides, the main point of interest in this work was to investigate the potential of HILIC to cope with the separation of very hydrophobic peptides, such as typically found in transmembrane helices of membrane proteins. Such peptides, which consist of long stretches of aliphatic and hydrophobic amino acids, and which might count an average of 40-50 amino acids in a tryptic digest, are typically not observed in a classical RPLC-MS separation. One likely explanation is that they might not survive the initial buffer conditions of a RPLC separation: those peptides cannot be kept soluble in aqueous solutions as they form secondary structure and precipitate before being injected into the RPLC system. In addition, their hydrophobicity might cause them to stick so strongly to the stationary phase that their elution might not be possible with conventional buffer systems. In contrast, the highly organic character of the initial buffer composition of a HILIC separation would favor the stability of those kinds of peptides while the polarity of the carbonyl backbone should ensure their retention onto the polar chromatographic medium.

The purpose of this work was to compare side-by-side the behavior of commercially available RPLC and HILIC chromatographic media in the separation of peptides of various hydrophobicities and to derive a first impression whether the premises delineated above could be verified.

C.2 Materials and Methods

C.2.1 Peptide synthesis

Selected peptides from bacteriorhodopsin were synthesized using standard peptide synthesis chemistry. Peptides 1-3 were synthesized at Peptide Specialty Laboratory (Heidelberg, Germany) while peptides 1-43 and 129 were synthesized at the peptide synthesis group at Roche Penzberg.

Peptide	Peptide sequence	Number of AA	Theoretical Mass (m/z)	GRAVY
Peptide 1-43	MLELLPTAVEGVSQAQITGRPE WIWLALGTALMGLGTLYFLVK	43	4686.5445	0.74
Peptide 129	FVWWAISTAAMLYILYVLFFGF TSK	25	2974.562	1.424
Peptide 1	VGFGLILLR	9	987.6354	1.956
Peptide 2	AESMRPEVASTFK	13	1452.7156	-0.523
Peptide 3	AIFGEAEAPEPSAGDGAAATSD	22	2032.886	-0.195

Table A.I: *List of synthetic tryptic and C-terminal (peptide3) peptides from bacteriorhodopsin. The GRAVY score (*grand average of hydropathicity, see Kyte, J. and Doolittle, R.F. (1982) A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. 157, 105-132) *was calculated using the ProtParam tool on the Expasy web site (www.expasy.ch). A positive GRAVY score indicates an hydrophobic sequence while a negative GRAVY score points to hydrophilic sequence. The theoretical mass of the peptide was calculated for a charge state of 1+.*

C.2.2 Capillary RPLC

Peptide mixtures were analyzed by capillary reverse phase liquid chromatography using an Agilent 1100 microflow chromatographic system using a 0.3 mm i.d x 15 cm length C18 PepMap100 (3 µm particle, 100 Å pore size, Dionex) column.

The peptide mixture of each sample was dissolved in buffer A (5% ACN, 0.1% FA). 20 µl of the sample was transferred to a glass Micro-V vial (Agilent) of which 10 µl were injected into the system at a flow rate of 5 µl/min at 5% buffer B (90% ACN, 0.5% FA) for 9 min. The peptides were eluted from the reverse phase column by increasing the ACN concentration in a linear gradient as follows: 9—48 min, 5—60% buffer B. The column was then washed for 10 min with 90% buffer B after which it was re-equilibrated for 28 min in 5% buffer B.

C.2.3 Capillary HILIC

Peptide mixtures were analyzed by HILIC chromatography with the same Agilent 1100 HPLC system that was used for the capillary RPLC method using a 0.3 µm i.d. x 15 cm length PolyHYDROXYETHYL-A (5 µm particle, 300 Å pore size, PolyLC Inc) column.

The peptide mixture of each sample was dissolved in buffer A (90% ACN, 0.1% FA, 5 mM ammonium formate). In some cases samples were dissolved in buffer A with the addition of either 0.05% SDS, 0.05% Triton X-100 or 10% tetrahydrofurane. 20 µl of the sample was transferred to a glass Micro-V vial (Agilent) of which 10 µl were injected into the system at a flow rate of 5 µl/min at 100% buffer A for 9 min. The peptides were eluted from the HILIC column by decreasing the ACN concentration in a linear gradient as follows: 9—48 min, 0—61% buffer B (5 mM ammonium formate, 0.1% FA). The column was then washed for 10 min with 88.5% buffer B after which it was re-equilibrated for 28 min in 100 % buffer B.

C.2.4 MALDI MS analysis, data processing and analysis

Peptides eluted between 10 and 70 min retention time were directly spotted onto the 384 positions of an Anchorchip target plate (Bruker Daltonics) pre-coated with matrix. Shortly, each position of the 384 anchorchip target plate was pre-spoted with 1 µl of α-cyano-4-hydroxycinnamic acid (0.25 mg/ml in 0.2 %TFA, 65 % Ethanol and 32% ACN containing 20 fmol/µl ACTH and bradykinin as peptide standards). After spotting, spots were re-crystallized at 4 $^{\circ}$C using 1 µl of 65% Ethanol, 32% ACN. Peptides were analyzed in a LC-MALDI modus with an Ultraflex I MALDI mass spectrometer (Bruker Daltonics) in reflector mode using standard operating procedures.

The LC-MALDI raw data of each fraction-spot were processed and filtered using an in-house software (MEDUSA). The algorithm (developed in the Proteomics group by Peter Berndt) removes most of the chemical noise introduced by the MALDI process and group related MS signal to a single variable (a chain). All spectral features (i.e., a peptide mass spectrometric signal) detected over a minimal signal over noise were extracted and internally calibrated. All those masses obeying a chromatographic behavior (i.e., the mass of an eluting peptide is expected to be observed in several consecutive fractions) were then clustered together using a star-shaped tolerance region and a nearest neighbor connected graph is build. For each connected component of the graph, the weighted mass and rf average and its error is estimated The algorithm removes most of the chemical noise, i.e. chemical components that are either continuously observed over the whole LC-MS procedure or that are randomly measured in

one or the other fraction, while keeping real features. The filtered signals were then graphically displayed using the Spotfire software package (TIBCO).

Protein identification from LC-MALDI analyses was performed as follows: The compiled peptide m/z masses obtained from the MEDUSA software package were analyzed using the MASCOT search algorithm (MASCOT version 2.1.04, Matrix Science). Searches were performed against the Human SwissProt database (version 49.1 February 2006, 13488 entries) with a mass tolerance of 50 ppm .Methionines (reduced/oxidized; +15.9949 Da) were considered as differential modifications while cysteines were considered as fully carbamidomethylated (+57.0199 Da). Only fully tryptic peptides with no more than two miscleavages were considered for data analysis.

C.3 Results

We examined in this study the chromatographic behavior of model peptides in respect to their retention time using either a RPLC or a HILIC modus. We were particularly interested in investigating the influence of peptide length and hydrophobicity onto peptide separation. We further examined the generality of both chromatographic systems by analyzing a simple protein digest by LC-MALDI MS and by identifying the population of peptides preferentially separated by a given separation modus.

C.3.1 Model peptides separation

A peptide mixture consisting of five synthetic peptides from bacteriorhodopsin (see Table A.1) and of ACTH was dissolved in the appropriate buffer and 10 µl of the mixture was then injected in the Agilent 1100 LC system. The final amount of each peptide in the injected sample was 1 pmol. Samples were analyzed either freshly after dilution in the initial buffer condition or after overnight incubation at 4 °C so that the peptide stability in the sample could be monitored.

The elution profiles of the peptide mixture using either RPLC or HILIC are shown in figure C.I.

Using the RPLC modus, all peptides could be chromatographically resolved except P129 (25 AA length and Gravy score of 1.424). The peptides' retention time followed roughly their hydrophobicity reverse order, with P2 (Gravy: -0.523) eluted first followed by P3, ACTH, P1 and P1-43. The elution of the very hydrophobic P1 peptide (according to the Gravy score) prior to P1-43 shows that the length of the peptide plays an important role. In particular, it can be inferred that the C-terminal Arg might significantly affect the chromatographic behavior of this otherwise rather short peptide. Interestingly, P1-43 (43 AA length and Gravy: 0.74) was only observed in the elution profile when the sample was analyzed freshly after dilution. These data suggest that long and/or hydrophobic peptides might not remain in solution using the polar solvent system required by the RPLC chromatography. In addition, such peptides will tend to stick on the surface of the vial or will be irreversibly bound to the C_{18} column.

Using the HILIC modus, all peptides including P129 could be chromatographically resolved. As expected, the peptides' retention time followed roughly their hydrophobicity order, with P129 (Gravy: -1.424) eluted first followed by P1, P2, P1-43, P3 and ACTH. The early elution of the hydrophilic peptide P2 (according to the Gravy score) prior to P3 and P1-43 also
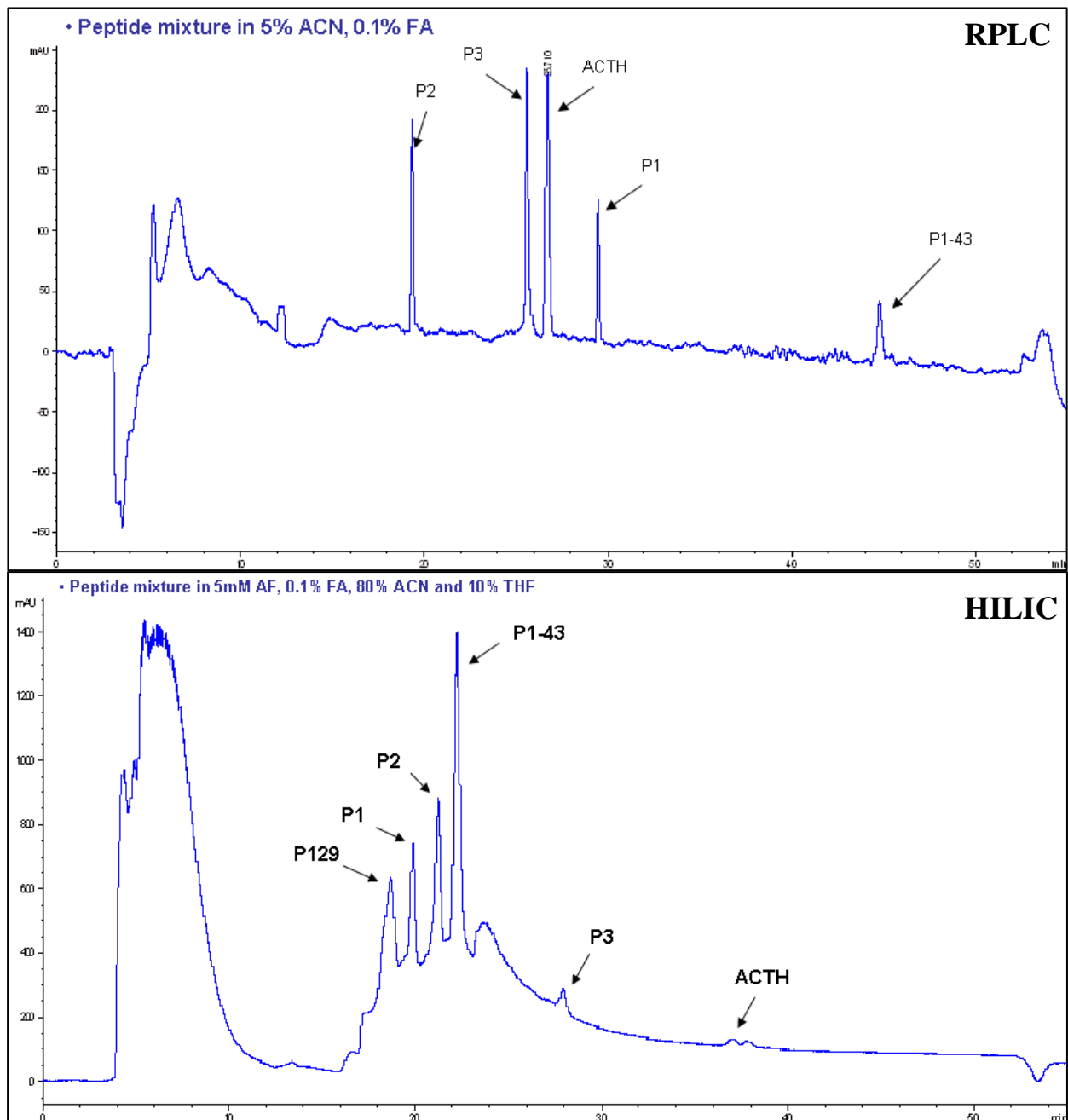
Fig.C.I: *Elution profiles of the Bacteriorhodopsin peptide mix plus ACTH. Ppeptide elution was monitored online at 214 nm. **RPLC:** 10 μl of the peptide mixture dissolved in 5% ACN and 0.1% FA were injected onto a C$_{18}$ column and eluted using an ACN linear gradient (5– 56% ACN). **HILIC**: 10 μl of the peptide mixture dissolved in 80% ACN, 0.1% FA, 5 mM AF and 10% THF were injected onto a HILIC PolyHYDROXYETHYL-A column. The peptides were eluted by lowering the ACN concentration in a linear gradient (90–10.35% ACN).*

indicates that hydrophilic interaction might not explain all the principles behind retention and separation in HILIC. An interesting aspect of HILIC is its compatibility with the use of very hydrophobic organic solvent, such as tetrahyrofurane (THF), and with the use of detergents, such as SDS or Triton-X100. In this experiment, the peptide mixture was either dissolved in buffer A (90% ACN, 0.1 % FA, 5 mM ammonium formate), or with the addition of 0.05%

SDS, 0.05% Triton X100 or using 10% THF. The presence of those modifiers didn't affect the chromatographic behavior of the HILIC column (as they were eluted in the injection peak) and all the peptide were eluted with the same order in all the condition described above (data not shown). The addition of THF in the sample buffer was observed to slightly increase the signal intensity in the elution chromatogram. The addition of those modifiers, combined with the organic character of the sample buffer, had a beneficial effect on the peptide's stability: all peptides could be observed in the elution chromatogram even after an overnight incubation at 4 $^{\circ}$C.

C.3.2 Separation of a peptide digest

In a subsequent experiment an HSA digest was analyzed using the Agilent 1100 HPLC system comparing the chromatographic separation of the C_{18} and the HILIC stationary phases. The HSA digest was prepared in the appropriate buffer and 10 µl (corresponding to 20 pmol of the digest) were injected into the system. LC fractions were collected every 9.4 sec and directly spotted onto the pre-coated 384 anchorchip target plate for subsequent MALDI MS analysis.

The elution profile of the 20 pmol HSA tryptic digest using either the RPLC or the HILIC chromatographic system is represented in the figure C.II.

Using the RPLC modus, the chromatographic separation of the HSA digest resulted in well resolved, sharp, intense peaks. Peak width was around 20-30 sec, ensuring maximum sensitivity for UV and MALDI MS measurement. In contrast, using the HILIC modus, the separation of the HSA digest resulted in an apparently poorly resolved elution chromatogram with peak elution time in the min range. As a consequence, typical peak only reached 100-200 mOD sensitivity at 214 nm absorption compared to 600-800 mOD reached with the RPLC modus. The presence of SDS, Triton X-100 or THF in the starting buffer did not have any noticeable effect on the chromatographic behavior of the HILIC column, except to cause a rather large injection peak. Interestingly, none of the HSA tryptic peptides eluted as early in the HILIC mode as the very hydrophobic rhodopsin peptides, indicating that the HSA digest contains mostly hydrophilic peptides. This is consistent with the elution chromatogram of the same digest in RPLC modus, where most peptides were eluted rather early in the chromatogram.
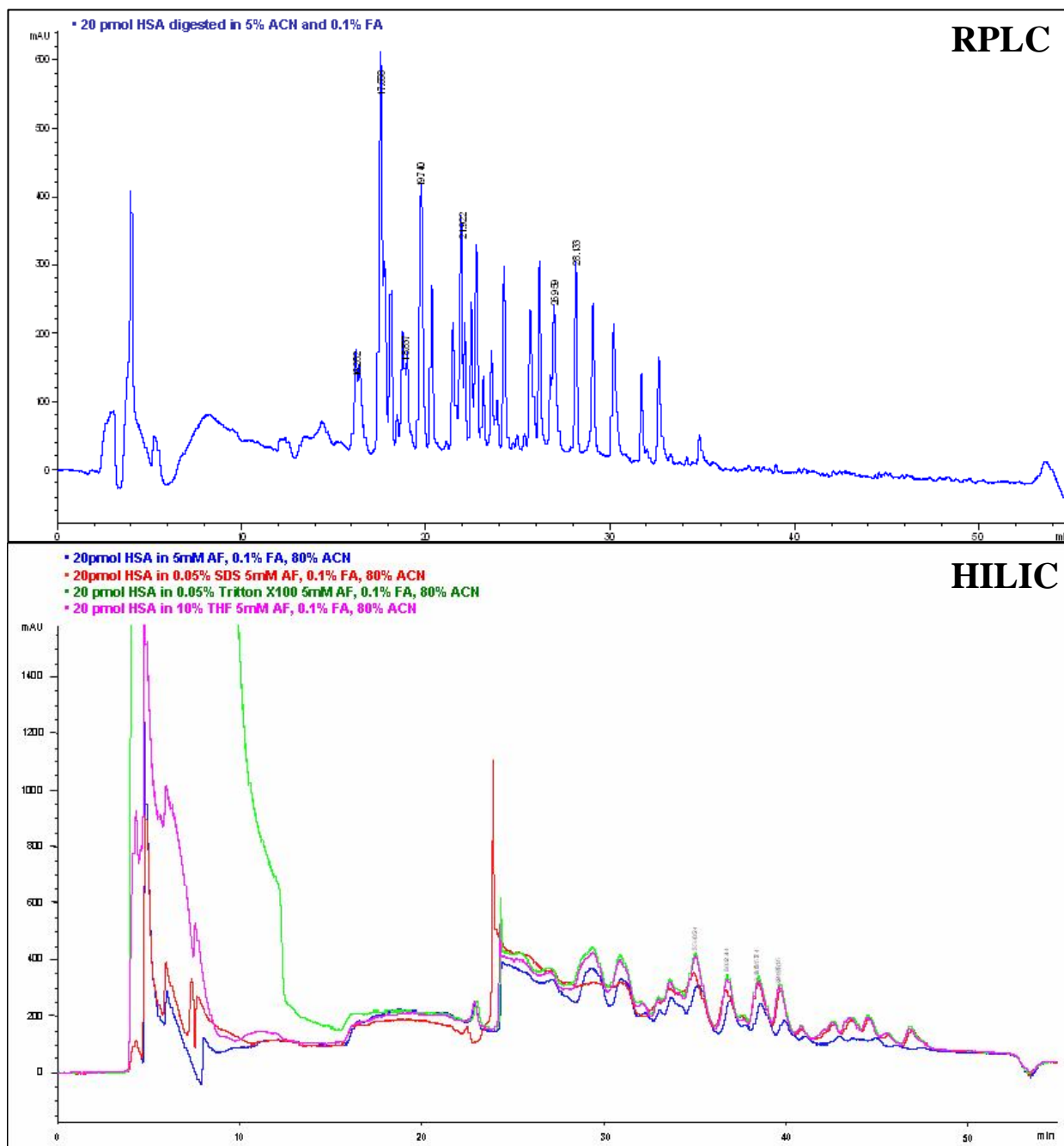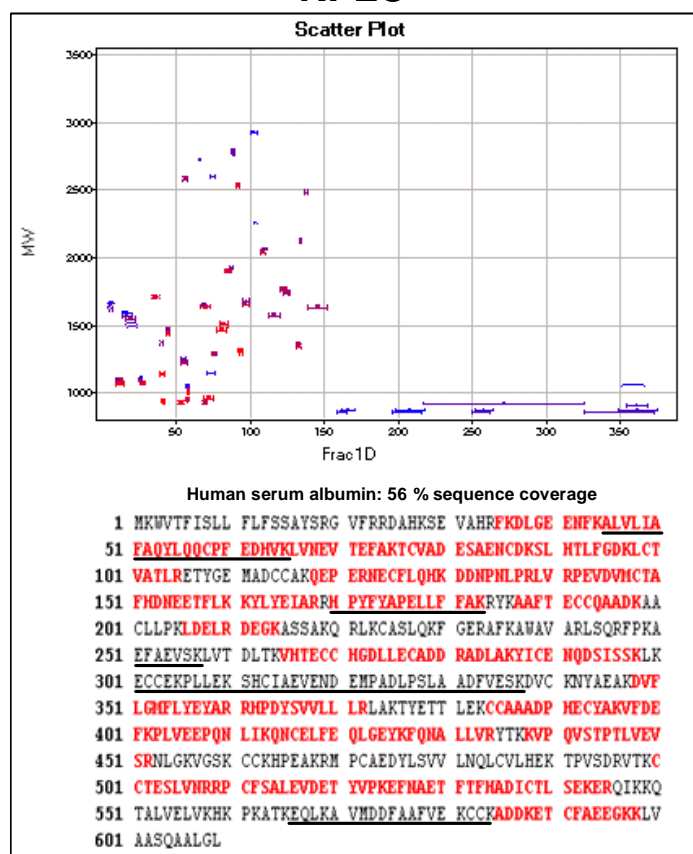
Fig.C.II: *Elution profiles of a 20 pmol HSA digest. Peptide elution was monitored online at 214 nm.* **RPLC:** *10 µl of the digest dissolved in 5% ACN and 0.1% FA were injected onto a C_{18} column and eluted using an ACN linear gradient (5– 56% ACN).* **HILIC**: *10 µl of the digest were injected onto a HILIC PolyHYDROXYETHYL-A column. The peptides were eluted by lowering the ACN concentration in a linear gradient (90–10.35% ACN).* **Blue**, *HSA digest in 90% ACN, 0.1 % FA, 5 mM AF;* **Red**, *HSA digest in 90% ACN, 0.1 % FA, 5 mM AF plus 0.005 % SDS.* **Green**, *HSA digest in 90% ACN, 0.1 % FA, 5 mM AF plus 0.05 % Triton X-100;* **Pink**, *HSA digest in 80% ACN, 0.1 % FA, 5 mM AF plus 10 % THF.*

The LC-fractions spotted onto the MALDI target were measured using an Ultraflex I MALDI mass spectrometer operated in the reflector mode. The raw data files of each fraction were filtered by the in-house MEDUSA software (Signal-to-noise ratio:12, chain length: 4) and then searched using the MASCOT search engine against the human SwissProt database at 50 ppm mass accuracy . The identified HSA tryptc peptides and their elution pattern in the RPLC or in the HILIC chromatography are schematically represented in Fig. C.III.

A

RPLC

Scatter Plot

Human serum albumin: 56 % sequence coverage

```
  1 MKWVTFISLL FLFSSAYSRG VFRRDAHKSE VAHRFKDLGE ENFKALVLIA
 51 FAQYLQQCPF EDHVKLVNEV TEFAKTCVAD ESAENCDKSL HTLFGDKLCT
101 VATLRETYGE MADCCAKQEP ERNECFLQHK DDNPNLPRLV RPEVDVMCTA
151 FHDNEETFLK KYLYEIARRH PYFYAPELLF FAKRYKAAFT ECCQAADKAA
201 CLLPKLDELR DEGKASSAKQ RLKCASLQKF GERAFKAWAV ARLSQRFPKA
251 EFAEVSKLVT DLTKVHTECC HGDLLECADD RADLAKYICE NQDSISSKLK
301 ECCEKPLLEK SHCIAEVEND EMPADLPSLA ADFVESKDVC KNYAEAKDVF
351 LGMFLYEYAR RHPDYSVVLL LRLAKTYETT LEKCCAAADP HECYAKVFDE
401 FKPLVEEPQN LIKQNCELFE QLGEYKFQNA LLVRYTKKVP QVSTPTLVEV
451 SRNLGKVGSK CCKHPEAKRM PCAEDYLSVV LNQLCVLHEK TPVSDRVTKC
501 CTESLVNRRP CFSALEVDET YVPKEFNAET FTFHADICTL SEKERQIKKQ
551 TALVELVKHK PKATKEQLKA VMDDFAAFVE KCCKADDKET CFAEEGKKLV
601 AASQAALGL
```

B   **HILIC**

Fig. CIII: *Scatter plot and sequence coverage of the 20 pmol HSA digest analyzed in the RPLC (panel A) or in the HILIC (panel B) modus. The scatter plot represents the LC-MALDI MS signals observed above the set threshold and chainlength (LoN: 12, chainlength: 4). The x axis represents the 384 fractions (the retention time) that have been collected during the RPLC run while the y axis stands for the molecular mass of the signals. The MALDI MS signal intensity is denoted by a color gradient from blue (low) to red (high). The number of fractions for which a signal has been observed is represented by the length of the horizontal line. The HSA amino acid sequence coverage for each chromatographic modi is schematically represented by highlighting the identified peptides in red. Peptides that were identified in only one of the two chromatographic separations are underlined.*

The peptides that were exclusively identified in one of the two chromatographic systems are listed in the table C.II

Peptide sequence	position	# AA	mass	GRAVY	RPLC	HILIC
ALVLIAFAQYLQQCPFEDHVK	45-65	21	2433.8	0.49	X	
HPYFYAPELLFFAK	170-183	14	1743	0.229	X	
AEFAEVSK	250-257	8	879.9	-0.138		X
ECCEKPLLEKSHCIAEVENDE MPADLPSLAADFESK	301-337	37	4091.6	0.354		X
EQLKAVMDDFAAFVEKCCK	566-584	19	2175	0.047		X

Table C.II: *Amino acid sequences and characteristics of the tryptic peptides identified exclusively through the HILIC or the RPLC system*

C.4 Discussion

Reverse-phase chromatography is at present the most commonly used chromatographic separation method for the analysis of peptides and proteins by mass spectrometry. Since its first use in the late 70' (127), 30 years of intense and systematic research has resulted in the development of stationary phases and columns with excellent resolution and sensitivity, for example, for the separation of small molecules and of peptides in complex biological samples. In particular, the decreasing size of the chromatographic beads (from 8 to 5, then 3 to today 1.8 μm diameter) enables the packing of nanoLC column (typically from 100 to 50 μm inner diameter) with unmatched resolution and sensitivity. Their coupling to the latest generation of mass spectrometers (e.g. Orbitrap or Q-ToF) allows the characterization of peptides in complex mixtures at the attomole level.

The optimal use of RPLC in the separation and analysis of peptides by mass spectrometry requires the analyzed biological samples to be void of salt (which would significant impair the electrospray process) and to be dissolved in an almost 100% aqueous solution before separation. While most peptides (and in particular tryptic peptides) withstand such conditions without problems, long, hydrophobic peptides and some proteins might stick to vial surfaces or precipitate out of solution. In addition, the hydrophobic nature of the RPLC chromatographic material will cause very hydrophilic peptides (such as phosphopeptides) to wash through the column at injection. On the other hand, very hydrophobic peptides (such as transmembrane peptides) will tend to bind strongly to a RP stationary phase and might not be eluted from the column even at high organic content of the mobile phase. This last problem can be partially alleviated by using less hydrophobic column material, such as C_8 or C_4, however at the cost of lower chromatographic resolution.

Since its description (120) in the early 90's, hydrophilic interaction liquid chromatography (HILIC) has been mainly used to chromatography and separate very polar analytes. Its mode of separation, which appears to be complementary to the RPLC chromatographic modus, has been best characterized using aspartimide-derivatized silica beads available commercially under the brand name PolyHYDROXYETHYL-A. Several studies demonstrated that mechanisms for retention and selectivity could mainly be explained by polar interactions (hydrogen bonding, dipole-dipole moments) and ion-exchange mechanisms, while the residual charges of the column's silanol groups might also provide selectivity for polar molecules (124, 128).

In this report our interest was mainly focused on the chromatographic behavior of long, hydrophobic peptides (such as generated from a tryptic digestion from membrane proteins) using HILIC. The main hypothesis behind this study relies on the observation that even long transmembrane peptides display a significant share of polar functions in the form of the carbonyl groups distributed on the peptide backbone, which should enable their retention on a HILIC stationary phase. In addition, the organic conditions typically used in HILIC and the compatibility of this chromatographic modus with various non-ionic detergents and modifiers should stabilize hydrophobic peptides better than in aqueous conditions prior to chromatographic separation. The elution profile of synthetic peptides from bacteriorhodopsin and ACTH (table A.I) using a PolyHYDROXYETHYL-A column confirmed the above idea. All the investigated peptides could be retained and separated by HILIC even if the sample was stored overnight at 4 ^{o}C to test peptide stability in those conditions. In contrast, only five of the six investigated peptides could be investigated using the C_{18} RPLC column. Peptide p1-43 could only be analyzed if the freshly-made sample was immediately injected into the HPLC system while peptide p129 was never detected, either because it precipitated out in aqueous conditions and/or because it could not be eluted from the column. Most interestingly, HILIC is compatible with organic modifiers, such as 10% THF, or detergents, such as 0.05% Triton-X100 or 0.05% SDS, as these chemicals are eluted with the injection peak and don't appear to interfere with the chromatographic separation. These reagents also contribute to the stability and the solubility of long, hydrophobic peptides or proteins in aqueous solution. Their use might open the possibility to directly analyze by HILIC-MS intact hydrophobic proteins or to digest hydrophobic proteins in presence of detergents or hydrophobic modifiers using experimental conditions that would be otherwise incompatible with RPLC.

The separation of a HSA tryptic digest using either RPLC or HILIC essentially confirmed the observations made with the bacteriorhodopsin and ACTH peptides, with the note that the chromatographic separation and resolution of the C_{18} column was considerably higher than those of the HILIC material. The eluent of the LC columns was directly spotted onto the 384-positions MALDI target, which allowed the analysis of the peptide mixture by MALDI-ToF MS. Based on sequence coverage, both chromatographic methods roughly equaled each other (61% vs. 56% for HILIC vs. RPLC). As expected, peptides commonly analyzed in both chromatographic modi were eluted in a "complementary" order (that is, peptides eluted early in the RPLC column tended to be eluted late in the HILIC column, and vice versa). Only five peptides were found in a single chromatographic modus (table A.II) with a trend for short, hydrophilic (probably not retained in a RPLC column) and long, hydrophobic (probably not

eluted from the RPLC column) peptides preferably observed on the HILIC column. The reason for observing $p_{45\text{-}65}$ and $p_{170\text{-}183}$ only in RPLC modus is not clear and might be due to the inferior performance of the HILIC column. A more systematic study using protein digests of diverse origins (especially from membrane preparation, see also below) should clarify in which conditions HILIC chromatography might provide advantages compared to RPLC.

This study summarizes a first set of experiments that was performed in our group to assess the use of HILIC for the chromatographic separation of peptides, and its potential for direct coupling to mass spectrometers. The ability to analyze peptide and protein samples in presence of high content of organic solvents (that is, for an enzymatic digestion, directly after digestion without a need for evaporation and reconstitution in aqueous conditions) leads to higher sensitivity. In particular, these conditions are beneficial to keep hydrophobic, long peptides in solution. Concomitantly, a chromatography performed with organic, volatile mobile phases is ideal for efficient desolvation and peptide ionization, in effect enhancing signal response in MS and leading to higher limits of detections (129). Interestingly, we have now several lines of evidence that even highly hydrophobic peptides (such as p1-43 or p129, table A.I) are amenable to mass spectrometric detection using electrospray ionization, provided they survive the reconstitution and the chromatographic steps. These peptides typically exhibit lower charge state as their hydrophilic counterpart and might not be detected in the "typical" mass range window scanned by the mass spectrometer.

In this study, due to time constrains, only one HILIC stationary phase (polyhydroxyethyl aspartamide) could be tested among many others that are potentially available, such as silica, amino, diol, or cyano. All these phases share roughly the same mode of operation but with a large range of specificity. In particular, the amino stationary phase appears to be particularly suited for peptide separation using solvent conditions close to RPLC as it does not require salt to minimize ion-exchange interactions. However, optimal separation conditions were achieved using TFA as an organic modifier (to shield from uncapped silanol groups), compromising thereby sensitivity using an electrospray ionization interface (125). One of the major difficulty performing this study was to find columns in the sub-microbore range (<500 µm i.d.) and/or bulk chromatographic media compatible with the packing of nanoLC column. Since the first report in 1990, HILIC has remained a niche chromatographic method for the separation of very polar compound and has lacked the development that has witnessed RPLC. However, recently, several groups have started to use HILIC for peptide separation (in particular, for the analysis of post-translational modifications, such as phosphorylation or glycosylation of peptides) and several chromatographic companies are now offering

stationary phases that could be used for the packing of nanoLC HILIC columns. A further improvement in separation could be obtained by monolithic structures instead of porous material, enabling the use of longer columns with improved peak capacity (130).

While this study provides some directions on how to analyze very hydrophobic peptides by LC-MS, the problem on how to generate such peptides and on how to keep them amenable to LC-MS analysis has not been considered here. One obvious limitation of the in-gel digest procedure remains the conditions in which the enzymatic digestion step has to be performed to avoid denaturing the proteolytic enzyme. There is at the moment no evidence that such hydrophobic peptides are generated in this procedure, and whether these peptides can be extracted efficiently from the gel matrix before they precipitate or aggregate. In the future, it will be of great importance to study the problematic in its entirety to understand whether an improved digestion protocol for membrane proteins, which should generate a better sequence coverage, in particular for transmembrane domains, can be expected at all.

PUBLICATIONS

The work presented on this thesis will contribute to the following publications:

E. Tsirogianni, M. Karas, H. Langen & A. Ducret "Characterization of Brush Border membrane proteins of mice intestinal mucosa. Emphasis to the study of cholesterol absorption", *to be submitted Jan 09*

E. Tsirogianni, F. Roos, Jens Lamerz, M. Karas, H. Langen & A. Ducret "Brush Border Membrane preparation and protein identification. Monitoring the reproducibility and the stability of a proteomics workflow based on LC-MS and LC MS/MS signal" *In preparation*

E. Tsirogianni, E. Chaput, E. Niesor, M. Karas, H. Langen & A. Ducret. "Protein characterization of Brush Border membranes of ApoE knockout: Is IBAT the missing link to hypercholesterolemia?" *In preparation*

Patents:

E. Tsirogianni & A. Ducret (2007) "Method for purifying brush border membrane proteins" submitted for patent application, number EP 07108055.0.

Oral Presentations:

E. Tsirogianni, M. Karas, H. Langen & A. Ducret "Characterization of Brush Border membrane proteins of mice intestinal mucosa. Emphasis to the study of cholesterol absorption", oral presentation at the 6th international congress of the Swiss proteomic Society, Lausanne Switzerland, 3-5 Dec. 2007.

E. Tsirogianni, N. Berntenis, M. Karas, H. Langen & A. Ducret "Identification of Brush Border membrane proteins of mice intestinal mucosa. Emphasis to the study of cholesterol absorption", oral presentation at the 2nd international conference of the Hellenic Proteomic Society, Chania, Greece, 23-25 May. 2007.

CURRICULUM VITAE

Eirini Tsirogianni

F. Hoffmann-La Roche Ltd
Pharmaceutical Division   Phone:+41-61-688-2048 (B)
Pharma Development Project Management (PDP)   +41-61-311-0508 (H)
Bld.74/4O.406   email.eirini.tsirogianni@roche.com
CH-4070 Basel

Personal information

Date of birth: October 15, 1978
Citizenship: Greek
Language skills: native Greek speaker
                        Fluent in English, oral and writing
                        Basic knowledge in German and French

Education

Johann Wolfgang Goethe University, Frankfurt-am-Main, Germany   **2004-present**

Institute of Pharmaceutical Chemistry under the joint of supervision of Prof. Dr. M. Karas (University of Frankfurt) and PD Dr. H. Langen (F. Hoffmann-La Roche Ltd).
PhD. thesis in the field of Proteomics "Membrane Proteomics: Characterization of Brush Border membrane proteins of mice intestinal mucosa. Case study: cholesterol absorption".

University of Crete, Heraklion, Greece   **2001-2003**

Chemistry Department under the supervision of Prof. Dr. G. Tsiotis
Masters of Science in Biochemistry "Study of the metabolic pathway of phenol degradation by a new strain of pseudomonas PhDV1"

University of Crete, Heraklion, Greece   **1997-2001**

Chemistry Department under the supervision of Prof. Dr. G. Tsiotis
Diploma work in BioAnalytical Chemistry: "Isolation and biochemical characterization of vitellogenin from the organism *Sparidae aurata*"
Overall GPA: 8.66/10 with mention "excellent".

Research and Working experience

F. Hoffmann-La Roche Ltd, PDP, Basel, Switzerland   **June 2008-present**

Trainee Global Project Manager in Pharma Development. Coordinator and facilitator of a Life Cycle Team.

F.Hoffmann-La Roche Ltd, PRG-RCMG, Basel, Switzerland   **2007-May 2008**

Coordinator for a joint project between the RCMG-Biology Research group and the Center of Biotechnology group of DSM Nutritional Products Ltd (in parallel to my PhD. work). Evaluation and development of quantitative methods using stable isotope labeling for the comparison of protein expression levels in bacterial strains using mass spectrometry.

F. Hoffmann-La Roche Ltd, PRG-RCMG, Basel, Switzerland   **2003-2004**

Trainee in research and technology development in the field of mass spectrometry. Evaluation and automatisation of the LIFT process in the Bruker Ultraflex II MALDI TOF mass spectrometer.
Supervisor: Dr A. Ducret

Novartis AG, Department of Immunology and Transplantation, Basel, Switzerland **2003**

Trainee in the field of immunology and transplantation. Evaluation of O-GlcNAc versus phosphorylation of proteins in activated Jurkat T-cells.
Supervisor: Dr. A. Katopodis

University of Crete, Department of Chemistry, Crete, Greece   **2002**

Preparation and set-up of the undergraduate biochemistry laboratory.
Supervisor: Prof. Dr. D. Ganotakis

University Hospital of Heraklion, Department of Clinical Pathology, Crete, Greece **2001**

Internship in the field of clinical chemistry. Research in liver carcinogenesis from hepatitis B patients.
Supervisor: Prof. Dr. I. Diamantis

University of Crete, Department of Chemistry, Crete, Greece   **2000-2001**

Researcher in the frame of the research program PENNED 99 of the Greek Secretariat for Research and Technology. Subject: "Regulatory mechanism of association and disassociation of the light harvesting complex LHCII of the photosystem II"
Supervisor: Prof. Dr. G. Tsiotis

Teaching experience

University of Crete, Department of Chemistry, Crete, Greece   **2001-2003**

Supervision of three undergraduate students during their Diploma thesis   2001-2003
Laboratory assistant for the undergraduate Biochemistry laboratory (2 semesters) 2002-2003
Teaching assistant in the undergraduate biochemistry course (2 semesters)   2001

Additional Publications

- E. Tsirogianni, M. Aivaliotis, D.G. Papasotiriou, M. Karas & G. Tsiotis "Identification of inducible protein complexes in the phenol degrader *Pseudomonas sp.* strain phDV1 by blue native gel electrophoresis and mass spectrometry", *Amino Acids*, (2006) 30, 1, 63-72.

- I. Tsirogianni, M. Aivaliotis, M. Karas & G. Tsiotis "Detection and characterization of catechol 2,3-dioxygenase in an indigenous soil *Pseudomonad* by MALDI-TOF MS using a column separation", *Biodegradation*, (2005) 16, 181-186.

- M. Aivaliotis, C. Corvey, I. Tsirogianni, M. Karas & G. Tsiotis "Membrane proteome analysis of the green sulfur bacterium *Chlorobium tepidum*", *Electrophoresis*, (2004) 25, 3468-3474.

- I. Tsirogianni, M. Aivaliotis, M. Karas & G. Tsiotis "Mass spectrometric mapping of the enzymes involved in the phenol degradation of an indigenous soil *pseudomonad*", *Biochim Biophys Acta.* (2004) 1700 (1), 117-123.

- I. Tsirogianni, M. Aivaliotis, & G. Tsiotis "Protein and lipid composition of a vitellin isolated from eggs of *Sparus aurata*", *Z. Naturforsch.* (2004) 59c, 132-134.

- Tsirogianni I. & Tsiotis, G. Preparative isoelectric focusing. In "*A Practical Guide to Membrane Protein Purification*" (Hunte, C., von Jagow, G. & Schaegger, G. eds) Academic Pres, New York, (2003) p. 131-142.

Additional Conferences

- A. Ducret; E. Kuehn; S. Kux Van Geijtenbeek; D. Röder; E. Tsirogianni & P. Berndt "Peptide Profiling by Liquid Chromatography Coupled off-line to MALDI-TOF Mass Spectrometry" poster presented at the 53rd American Society for Mass Spectrometry in San Antonia, TX, USA, 5-9 Jun 2005.

- I. Tsirogianni, M. Aivaliotis, M. Karas, & G. Tsiotis, "Investigation of the Metabolic Pathway of Phenol Degradation of an Indigenous Soil Pseudomonad by MALDI TOF-MS", poster presented at the Proteomic Forum at the International Meeting on Proteome Analysis, Technical University Munich Germany, 14-17 Sep 2003.

- M. Aivaliotis, C. Corvey, I. Tsirogianni, M. Karas & G. Tsiotis, "Investigation of the Metabolic Pathway of Phenol Degradation of an Indigenous Soil Pseudomonad by MALDI TOF-MS" poster presented at the Proteomic Forum at the International Meeting on Proteome Analysis, Technical University Munich Germany, 14-17 Sep 2003.

- I. Diamantis, E. Karamitopoulou, G. Christodoulopoulos, V. Valatas, G. Notas & I. Tsirogianni, "HNF-1alpha and VHNF genes are up regulated in livers of CCl4 treated rats", poster presented at the 38th Annual Meeting of the European Association for the study of the Liver, Geneva, Switzerland, 3-4 Jul 2003.

- I. Tsirogianni, M. Aivaliotis, M. Karas, & G. Tsiotis, "Detection and characterization of catechol-2,3-dioxygenase in an indigenous soil Pseudomonad by MALDI-TOF MS using a column separation", poster presented at the 2nd European Bioremediation conference, Chania, Crete, Greece, 30 Jun-4 Jul 2003.

- I. Tsirogianni, M. Aivaliotis & G. Tsiotis, "Research of enzymes involved in the phenol degradation of a bacterium belongs to genus *Pseudomonas*" poster presented at the 19th Panhellenic Conference of Chemistry, Heraklion, Greece. 6-10 Nov 2002.

Schlussgedanken

Eidesstattliche Versicherung

Ich erkläre hiermit an Eides hiermit an Eides Statt, dass ich die vorgelegte Dissertation über "Membrane Proteomics: Characterization of Brush Border membrane proteins of mice intestinal mucosa. Case study: cholesterol absorption" selbständig angefertigt und anderer Hilfsmittel als der in ihr angegebenen nicht bedient habe, insbesondere, dass aus Schriften Entlehnungen, soweit sie in der Dissertation nicht ausdrücklich als solche mit Angabe der betreffenden Schrift beyeichnet sind, nicht stattgefunden haben.

Frankfurt am Main, den ....

....
Eirini Tsirogianni