“Next Generation Sequencing“ - basierende genomische Biomarker zur Risikodetektion von chronischem Schmerz

Dissertation

zur Erlangung des Doktorgrades

der Naturwissenschaften

vorgelegt beim Fachbereich Biowissenschaften

der Johann Wolfgang Goethe - Universität

in Frankfurt am Main

von

Dario Kringel

aus Frankfurt am Main

Frankfurt (2019)

(D30)

Vom Fachbereich Biowissenschaften der Goethe Universität Frankfurt am Main
als Dissertation angenommen.

Dekan:	Prof. Dr. Klimpel
Gutachter:	Prof. Dr. Helge Bode
	Prof. Dr. Dr. Jörn Lötsch

Datum der Disputation:

1. ZUSAMMENFASSUNG 1

2. EINLEITUNG 5

2.1 Allgemeine Mechanismen der Nozizeption 6

2.1.1 Opioid-Rezeptor System 7

2.1.2 TRP-Kanal-Rezeptor System 8

2.2 Chronifizierung von Schmerz 9

2.2.1 Molekulare Grundlagen der Schmerz-Chronifizierung 9

2.2.2 Chronischer neuropathischer Schmerz 10

2.2.3 Gliazellen 11

2.3 Entwicklung der Systemmedizin 12

2.3.1 Das Zeitalter der “Omics“-Technologien 12

2.3.2 Funktionelle Genomik 12

2.4 DNA-Sequenzierung 14

2.4.1 Genetische Grundlagen der DNA-Sequenzierung 14

2.4.2 Anwendungsbereiche von DNA-Sequenzierungen 15

2.4.3 “Next-Generation-Sequencing“ 16

2.4.4 Ion Torrent™ Technologie 18

2.4.5 Ion Torrent™ Workflow 19

2.4.5.1 Template Konstruktion 19

2.4.5.2 Datenanalyse 20

2.4.5.3 Genetische Variationen 21

2.5 Computionale Wissensgenerierung 22

2.5.1 Maschinelles Lernen 23

2.5.2 Maschinelles Lernen in der Schmerzforschung 24

2.6 Wissenschaftliche Hintergründe dieser Arbeit 25

2.6.1 Etablierung der Ion Torrent™ Plattform 25

2.6.2 Etablierung bioinformatischer Methoden zur Auswertung von “Big Data“ 27

2.6.3 Charakterisierung und Identifizierung von Schmerzphänotypen 28

2.6.4 Funktionelle genomische Charakterisierung der Schmerz-Chronifizierung 30

2.6.5 Charakterisierung chronischer Schmerzgenotypen 31

2.6.6 Genotyp-Phänotyp Assoziation bei chronischen Schmerzpatienten 33

2.7 Ziel dieser Arbeit 34

3. DISKUSSION 35

3.1 Ion Torrent™ Etablierung und Auswertung von NGS-"Big Data" 36

3.2 Identifizierung von schmerzbasierten Phänotypen 38

3.3 Charakterisierung von Genotypen durch funktionelle Genomik 40

3.4 Identifizierung genetischer Prädiktoren für chronische Schmerzen 42

3.5 Komplexe Genotyp-Phänotyp Assoziationen 45

3.6 Ausblick 49

3.6.1 "Big Data" in der biomedizinischen Forschung 50

4. LITERATURVERZEICHNIS 52

5. ANHANG 61

5.1 Abkürzungen 61

5.2 Eidesstattliche Erklärung 62

5.3 Danksagung 63

5.4 Curriculum Vitae 64

6. MANUSKRIPTE 65

1. Zusammenfassung

Die Physiologie des Schmerzes umfasst komplexe immunologische, sensorische und inflammatorische Prozesse im Rückenmark, im Gehirn und in der Peripherie. Wiederholte nozizeptive Stimulation induziert pathophysiologische Veränderungen bei der Schmerzweiterleitung, aus denen eine periphere oder zentrale Sensibilisierung resultiert. Diese kann bei dafür anfälligen Patienten zu der Ausbildung von chronischen Schmerzzuständen führen. Obwohl das Wissen über die genauen molekularen Vorgänge der Schmerz-Chronifizierung noch immer unvollständig ist, sind die Identifizierung von Risikofaktoren vernünftige Schritte, um die individuelle Anfälligkeit für die Entwicklung chronischer Schmerzen zu bestimmen. Das Hauptziel dieser Doktorarbeit bestand daher in der Identifikation humaner genetischer Biomarker für chronische Schmerzzustände.

Das biologische Ausgangsmaterial stammte von gesunden Probanden sowie von Schmerzpatienten, basierend auf einer Kohorte von 1.000 Frauen kaukasischer Herkunft im Alter von 28 bis 75 Jahren, welche an nicht metastasierendem Brustkrebs litten. Nach der operativen Behandlung wurde die Intensität der postoperativen Schmerzen in mehreren Abständen für eine Dauer von 3 Jahren anhand von Fragebögen erfasst. Aus dieser Kohorte wurden zwei Gruppen von Schmerzphänotypen identifiziert, bei der jeweils 70 Patienten zu einer “Schmerz-Chronifizierer“ oder einer “Nicht-Schmerz-Chronifizierer“ Gruppe zugeordnet wurden, sich sonst aber in anderen Parametern wie Gewicht und Alter nicht unterschieden.

Die Existenz einer genetischen Komponente, welche auf die individuelle Wahrnehmung von Schmerz und den pathophysiologischen Prozess seiner Chronifizierung funktionell einwirkt, wird durch das Wissen über viele Schmerz-modulierende Mutationen deutlich, welche durch die wissenschaftliche Forschung im Bereich der Humangenetik innerhalb der letzten Jahrzehnte identifiziert wurden. Da viele dieser Mutationen als Einzelnukleotid-Abweichungen auftreten, welche erst im Zusammenspiel eine erkennbare Auswirkung auf den Phänotyp haben, ist es schwierig, eine einzelne Mutation mit der veränderten Reaktion auf bestimmte Medikamente oder dem Risiko einer Ausbildung bestimmter Krankheitsbilder zu assoziieren.

Während häufige Einzelnukleotid-Varianten seit mehreren Jahren mit großem Erfolg durch genomweite Assoziationsstudien identifiziert werden konnten, ist die Identifizierung seltener Varianten, insbesondere aufgrund der großen Zahl der beitragenden Gene, nur begrenzt erfolgreich. Im Laufe der letzten Jahre konnten mehr als 500 Gene identifiziert werden, die an der Modulation schmerzbezogener Phänotypen beteiligt sind. Durch die rasante Entwicklung im Bereichen der DNA-Sequenzierung, insbesondere dem "Next Generation Sequencing" (NGS) und den daraus resultierenden Sequenzdaten, welche das gesamte Exom umfassen und dem parallelen Fortschritt im Bereich der "Data Science", vor allem in den Teilbereichen des maschinellen Lernens und der künstlichen Intelligenz, ist es möglich, dem klinischen Problem der Schmerz-Chronifizierung auf einer genomweiten Ebene zu begegnen.

Aufgrund von technisch bedingten Limitierungen innerhalb der NGS-Anwendungen war es notwendig die Anzahl der Gene, welche für die Genotypisierung in Frage kommen, zu reduzieren. Die Auswahl der vorliegenden Kandidaten-Gene erfolgte mittels systematischer "Pathway-Analysen", funktioneller Genomik und systembiologischer sowie bioinformatischer Methoden. Durch innovative Verbesserungen im Bereich der Systemmedizin, welche die verschiedenen Fachrichtungen wie Biologie, Medizin, und Informatik in sich vereint, sind mittlerweile genauere Charakterisierungen der biochemischen, zellulären und physiologischen Eigenschaften von Genprodukten möglich, um die Beziehung zwischen dem Genom und dem Phänotyp besser zu verstehen.

Dadurch kann das aus klinischen Studien generierte Wissen über genetische Modulationen mit aktuellen Kenntnissen über Gen-Funktionen kombiniert werden. Es wurde mit Hilfe dieser Methoden ein Set von 77 sogenannten "Schmerz-Chronifizierungs-Genen" zusammengestellt, welches als Grundlage für die NGS-Genotypisierung, basierend auf der Ion Torrent™ Technologie, diente.

Das Set enthielt Gene, welche (i) für die Steuerung biologischer Lernprozesse und neuronale Plastizität verantwortlich sind, basierend auf der Sichtweise, dass sich Schmerz als eine Art Deregulierung bestimmter biologischer Prozesse abbilden lässt, welche in der“ Gene Ontologie-Database“ unter den Termen “Learning or Memory“ und “Nervous System Development“ zusammengefasst werden. Zudem enthielt es Gene, die (ii) funktionelle Varianten enthalten, welche die Schmerz-Chronifizierung in mindestens zwei unterschiedlichen klinischen Schmerzzuständen modulieren. Außerdem wurden Gene miteingeschlossen, welche (iii) in den letzten Jahren im Bereich der Schmerzforschung wiederholt eine wichtige Rolle bei der Modulation von langanhaltenden Schmerzen spielten oder teilweise auch erst kürzlich als neuartige und vielversprechende genetische Modulatoren gehandelt wurden.

Die Hauptidee bei der Auswertung der NGS-Daten war das Trainieren einer künstlichen Intelligenz (KI), implementiert als verschiedene Arten des maschinellen Lernens, um ihr Assoziationen der genetischen Informationen mit dem Schmerz-Phänotyp anzulernen und anschließend die trainierte KI zu verwenden, um in neuen Daten einen Phänotyp vorherzusagen. Dabei wurde der innovative Ansatz verfolgt, bei dem eine durch NGS-basierte genetische Informationen trainierte KI, den entsprechenden Phänotyp besser vorhersagen sollte, als dass es durch Raten möglich wäre. Dazu wurden die ausgewählten 140 Probanden basierend auf der Ion Torrent™ Technologie genotypisiert.

Aufgrund der zu erwartenden hohen Dimensionalität der komplexen, durch NGS gewonnenen genetischen Informationen, wurden Methoden des maschinellen Lernens von vornherein traditionellen statistischen Methoden zur Datenanalyse vorgezogen. Die Datenauswertung umfasste eine Merkmals-Selektion zur Identifizierung der informativsten genetischen Varianten mit anschließender Genotyp-Phänotyp-Assoziation unter Verwendung verschiedener Arten von Algorithmen basierend auf Methoden des maschinellen Lernens. Im Zuge der Datenanalyse wurde die genetische Information von den ursprünglichen 4.748 Varianten, die in den 77 Genen gefunden wurden, auf 21 relevante Varianten reduziert, die in 13 verschiedenen Genen lokalisiert waren. Das gefundene Muster stützte eine Assoziation der Genotypen mit den Schmerzphänotypen. Die angewendete KI, die mit 2/3 der Daten trainiert wurde, identifizierte die richtige Schmerzphänotyp-Gruppe bei den verbleibenden 1/3 der Patienten mit einer Genauigkeit von etwa 70 Prozent.

Obwohl die Auswahl der Kandidaten-Gene größtenteils auf der Grundlage früherer wissenschaftlicher Studien basierte, welche größtenteils keine Assoziationen zu chronischen Schmerzen nach Brustkrebs-Operationen enthielten, lieferten die vorliegenden Ergebnisse eine Gen-Rangfolge, welche sich mit ihrer besonderen Bedeutung ihres klinischen Kontextes vereinbaren lässt. Die Analyse identifizierte die genetischen Unterschiede zweier Phänotyp-Gruppen im Hinblick auf die Entwicklung einer Schmerz-Chronifizierung zusätzlich zu einer Modulation des klinischen Verlaufs von Brustkrebs. Die Schmerz-Modulation kann bei dieser Analyse anhand bestimmter Nukleotid-Abweichungen charakterisiert werden.

Außerdem unterstützt die Tatsache, dass mehr als 75 Prozent der bei dieser Analyse identifizierten Gene, Targets von bereits zugelassenen oder sich noch in der Entwicklung befindenden Schmerzmitteln sind, die Wichtigkeit dieser Ergebnisse. Der vorliegende Satz von Gen-Varianten kann als ein wichtiger Baustein dienen, um den Weg für die Etablierung einer personalisierten Schmerz-Therapie zu ebnen, wenn einige der sich noch in der Entwicklung befindlichen Medikamente klinisch verfügbar werden.

2. Einleitung

“Ein unangenehmes Sinnes- und Gefühlserlebnis, das mit aktueller oder potentieller Gewebeschädigung verknüpft ist oder mit Begriffen einer solchen Schädigung beschrieben wird"; so wird von der Weltschmerzorganisation “International Association for the Study of Pain“ (IASP) der Begriff “Schmerz“ definiert [1]. Schmerz ist überlebenswichtig, er dient als ein Warnsignal indem er den Körper auf eine drohende Gewebeschädigung hinweist und in der Regel geht dem Schmerz ein ihn auslösender Reiz voraus [2]. Ist der Schmerz lokal und zeitlich begrenzt, spricht man von einem akuten Schmerz. Bleibt der Schmerz aber über einen längeren Zeitraum von mehreren Monaten bestehen und man kann ihn nicht mehr mit einem bestimmten Auslöser in Verbindung bringen, spricht man von chronischem Schmerz [3].

Während akuter Schmerz ein wichtiges Symptom ist, verliert der Schmerzreiz durch die Chronifizierung seinen schützenden Charakter und somit auch seine Relevanz [4]. Chronische Schmerzen sind ein weitverbreitetes Gesundheitsproblem. Sie verringern die Lebensqualität der Patienten und verursachen erhebliche Kosten im Gesundheitswesen aufgrund von Frühverrentung und Arbeitsunfähigkeit [5]. Schätzungsweise leidet jeder fünfte Europäer und jeder dritte Europäer über 70 Jahren täglich an chronischen Schmerzen, die so stark sind, dass sie eine medizinische Behandlung erfordern [6].

Diese hohe Prävalenz ist ein eindeutiges Zeichen für momentan noch vorherrschende unzureichende Behandlungsmöglichkeiten. In der Tat belegen wissenschaftliche Studien, dass die aktuell verfügbaren Analgetika nur bei einer Minderheit der Patienten eine zufriedenstellende Schmerzlinderung bewirken [7, 8]. Die verfügbaren Schmerzmittel sind entweder nicht effektiv genug oder die Patienten leiden an Unverträglichkeiten und können die Medikamente aufgrund zu starker Nebenwirkungen nicht einnehmen [9]. Daher ist es erforderlich sich auf die Herstellung von Analgetika mit neuartigen Wirkungsmechanismen zu konzentrieren, deren Entwicklung auf der Grundlage eines besseren Verständnisses der komplexen Pathophysiologie des Schmerzes [10] und dessen klinischer Darstellung beruht [11].

Diese Wissensgrundlage wurde innerhalb der letzten 50 Jahre im Bereich der Schmerzforschung durch die Kenntnis über viele Schmerz-modulierende humane Polymorphismen erschaffen. Lötsch und Kollegen haben zu diesem Thema eine ausführliche Übersichtsarbeit verfasst [12]. Dadurch wird die Existenz einer genetischen Komponente deutlich, welche funktionell auf die Schmerz-Chronifizierung einwirkt. Obwohl mittlerweile mehr als 500 Gene identifiziert werden konnten, die an dieser Modulation beteiligt sind, ist das Verständnis um aller dazugehörigen molekularen Vorgänge noch immer unvollständig [13]. Dieser aktuelle Zustand verdeutlicht die Notwendigkeit einer Etablierung von datenwissenschaftlich basierten Methoden, wie maschinellem Lernen und künstlicher Intelligenz, in Kombination mit modernsten Techniken der Hochdurchsatz-Sequenzierung ("Next Generation Sequencing"), um genetische Prädiktoren für das Risiko einer Schmerz-Chronifizierung zu charakterisieren und diesem klinischen Problem auf einer genomweiten Ebene zu begegnen.

2.1 Allgemeine Mechanismen der Nozizeption

Die einzelnen Stimuli, welche eine nozizeptive Reaktion hervorbringen können, sind sehr unterschiedlich, aber die Rezeptoren und endogenen Mechanismen in der Peripherie des Nervensystems interagieren immer auf die gleiche Weise. Chemisch-, mechanisch- und thermisch-reizbare Rezeptoren erkennen und definieren die Intensität, Dauer und Lokalität von nozizeptiven Stimuli. Sie leiten die Schmerzreize weiter an das Hinterhorn des Rückenmarks, wo sekundäre Neurone, aktiviert von Neurotransmittern, das Signal mit Hilfe von aufsteigenden Nervenbahnen ans Gehirn weitergeben [10]. Daraufhin werden sensorische oder modulatorische Prozesse eingeleitet, um den Vorgang der Schmerzempfindung zu stoppen. Im Idealfall nehmen die Schmerzreize mit einer fortschreitenden Heilung ab und das Schmerzempfinden lässt nach, bis keine Schmerzen mehr festgestellt werden können.

2.1.1 Opioid-Rezeptor System

Das körpereigene Opioid-Rezeptor-System besteht aus einem weitverbreiteten, aber strategisch sinnvoll konzentriertem Netz aus Opioid-Rezeptoren und deren endogenen Agonisten. Dieses System spielt eine maßgebliche Rolle bei anti-nozizeptiven Prozessen und ist deshalb vor allem in Körperregionen, die mit Schmerz und Nozizeption in Verbindung stehen, repräsentiert [14]. Das endogene Opioid-System spielt jedoch auch bei der peripheren Analgesie und der Modulation von Stimmungslage und Stress eine wichtige Rolle, außerdem vermittelt es die Wirkungen der exogenen Opioid-Liganden [15]. Der Bereich der endogenen Opioide umfasst die Peptid-Gruppen der Endorphine, Enkephaline und Dynorphine [16] und hat differenzielle Affinitäten zu den Opioid-Rezeptoren [17]. Vier verschiedene Opioid-Rezeptoren: Mu (μ), Delta (δ), Kappa (κ) und Opioid-Rezeptor-Like (ORL) wurden bisher identifiziert. Alle vier Opioid-Rezeptoren gehören zur Familie der G-Protein-gekoppelten Rezeptoren und werden von den Genen *OPRM1*, *OPRK1*, *OPRD1* und *OPRL1* codiert [18].

Opioid-induzierte Analgesie wird durch Aktivierung des endogenen Opioid-Systems ausgelöst. Der Prozess der Schmerzlinderung erfolgt dann durch verschiedene molekulare Mechanismen, die zur analgetischen Wirkung der Opioide beitragen und sich über das gesamt nozizeptive System erstreckt. Sie sorgen für eine Hemmung der Reizweiterleitung in das Rückenmark und aktivieren das absteigende Hemmsystem [19]. Opioide sind aus der heutigen Schmerztherapie nicht mehr wegzudenken, die Verabreichung kann jedoch auch zu unerwünschten Nebenwirkungen führen, was ihre Dosierung und somit auch ihre Wirksamkeit einschränkt. Zu den häufigsten Nebenwirkungen zählen Verstopfung, Übelkeit und Müdigkeit [20], es kann jedoch auch zu schwerwiegenderen Komplikationen, wie zum Beispiel einer Atemwegsdepression kommen [21]. Es existieren zwei weitere klinisch relevante Phänomene, die man als eine Art unerwünschten Nebeneffekt bezeichnen kann. Die “Opioid-induzierte Toleranz“, bei der nach wiederholter Applikation die Opioide ihre schmerzstillende Wirkung auf Grund einer Toleranzentwicklung verlieren, und die “Opioid-induzierte Hyperalgesie“, eine Absenkung der Schmerzschwelle, die gleichzeitig mit der Opioid-Gabe beginnt, und plötzlich einsetzt [22]. Beiden Phänomenen liegen ähnliche molekulare Mechanismen zugrunde, wobei die Aktivierung des exitatorischen glutaminergen Systems als einer der wichtigsten Prozesse zu nennen ist [23], da gezeigt werden konnte, dass Antagonisten von Glutamat-sensitiven NMDA-Rezeptoren Opioid-induzierte Toleranz sowie Opioid-induzierte Hyperalgesie abschwächen können [24, 25].

2.1.2 TRP-Kanal-Rezeptor System

Die TRP-Kanäle (englisch: “Transient Receptor Potential Channels“) sind eine Familie zellulärer nicht-selektiver Ionenkanäle, die sich in sieben Unterklassen aufteilen lassen. Die Homologie zwischen den einzelnen Klassen ist nur geringfügig ausgeprägt, sie alle haben jedoch gemeinsam, dass sie 6 Transmembranregionen besitzen und den Ionentransport nicht nur ermöglichen, sondern auch hemmen können. TRP-Kanäle sind entwicklungsgeschichtlich sehr alt und konnten bereits in Insekten, aber auch in höher entwickelten Säugetieren, nachgewiesen werden. Sie können durch harmlose sowohl als auch schädliche äußere Reize aktiviert werden, welche mechanischen, chemischen und osmotischen Ursprungs sein können. Mehrere klinische Studien haben gezeigt, dass TRP-Kanäle eine wichtige Rolle bei der Weiterleitung thermischer Schmerzen einnehmen, da einige TRP-Kanäle durch bestimmte Temperaturen aktiviert werden können. TRPM8 wird durch Kälte (≤25 °C) und Menthol aktiviert. TRPV1 wird durch Hitze (≥42 °C) und Capsaicin aktiviert, während TRPA1 ebenfalls durch Kälte (<18 °C) und durch eine Vielzahl weiterer chemischer Verbindungen wie Zimtaldehyd und Allicin aktiviert werden kann [26-28]. TRP-Kanäle gelten in der Arzneimittelforschung als vielversprechende Targets für die Entwicklung von neuen Analgetika, wobei sich mehrere pharmazeutische Substanzen derzeit in der klinischen Testphase befinden. Als wichtigste pharmazeutische Kandidaten aus der Gruppe der thermosensitiver TRP-Kanäle sind TRPV1 und TRPA1 zu nennen [29]. Eine Abfrage des Thomson Reuters “Drugs and Biologics Search Tool“ (http://integrity.thomsonpharma.com) im Juni 2016 ergab, dass TRPV1 mit mehr als 200 Agonisten oder Antagonisten in der Klasse der Analgetika, welche sich aktuell in der Wirkstoffentwicklung befinden, der TRP-Kanal ist, auf dem das meiste pharmazeutische Interesse liegt [30], dicht gefolgt von TRPA1 und TRPM8 [29]. Wissenschaftliche Studien konnten außerdem eine Co-Expression von TRPV1 und TRPM8 sowie die Co-Expression von TRPV1 und TRPA1 nachweisen [31].

2.2 Chronifizierung von Schmerz

Vereinfacht gesagt sind chronische Schmerzen ein paradoxes Phänomen, welches im Vergleich zum akuten Schmerz keinem adaptiven Zweck dient [32]. Akute Schmerzen lassen sich experimentell leicht als physiologische oder pathophysiologische Reaktionen des Körpers auf zugeführte Reize oder Gewebeschäden darstellen, während chronische Schmerzen durch verschiedene Arten von Gewebeschäden und auch durch Schädigung der peripheren oder zentralen Nerven des Schmerzleitungssystems verursacht werden können [33] [34]. Chronische Schmerzen können aber auch spontan auftreten, ohne dass irgendeine Gewebeschädigung vorliegt, oder sie können paradoxerweise in Bereichen auftreten, in denen gar kein geschädigtes Gewebe mehr vorhanden ist [35, 36]. Chronische Schmerzen können die Schwellenwerte für das, was wir normalerweise als Schmerz definieren würden, verschieben und zu einer “Hyperalgesie“ (erhöhte Schmerzempfindlichkeit gegenüber einem schmerzhaften Reiz) oder einer “Allodynie“ (Schmerz als Reaktion auf Reize, die normalerweise keine Schmerzen verursachen) führen [37, 38]. Das Verständnis um die Notwendigkeit unterschiedlicher Behandlungsansätze bei akuten und chronischen Schmerzen steht seit Jahren im Mittelpunkt von intensiver Forschungsarbeit. Die Schwerpunkte auf diesem Gebiet liegen auf einer verbesserten Interpretation und einem detaillierteren Verständnis der einzelnen Prozesse, die bei der Chronifizierung von Schmerzen im menschlichen Körper ablaufen [32].

2.2.1 Molekulare Grundlagen der Schmerz-Chronifizierung

Länger anhaltende, intensivere Schmerzen, bei denen die Gefahr einer Chronifizierung besteht, aktivieren sekundäre Mechanismen im Zentralnervensystem, sowohl als auch in der Peripherie, was Allodynie, Hyperalgesie und Hyperpathie verursachen kann, und die normale Funktion der Schmerzempfindung beeinträchtigt [35]. Diese Veränderungen beginnen in der Peripherie mit der Hochregulierung des Enzyms Cyclooxygenase 2 [39] und im Zentralnervensystem mit einer verstärkten Aktivierung von NMDA-Rezeptoren [40], einer daraus folgenden Sensibilisierung von Nervenzellen-Kaskaden und einer Aktivierung der Mikroglia [41], welche die neuronale Architektur beeinflussen und verändern können [42, 43]. Bei diesen komplexen Vorgängen, die während der Umwandlung von akutem zu chronischem Schmerz ablaufen, spielen außerdem noch Prostaglandine [44], Endocannabinoide [45] und

ionenspezifische Kanäle [46] wichtige Schlüsselrollen, deren genaues Zusammenspiel bisher jedoch noch nicht vollständig aufgeklärt werden konnte [35]. Die Entstehung und Aufrechterhaltung chronischer Schmerzen erfolgt unter anderem durch den Prozess der zentralen Sensibilisierung [47], ein Phänomen basierend auf der Grundlage synaptischer Plastizität [48] und neuronaler Reorganisation des zentralen Nervensystems [49]. Einige wissenschaftliche Arbeiten weisen darauf hin, dass die zentrale Sensibilisierung auch durch Neuroinflammation im peripheren sowohl auch im zentralen Nervensystem ausgelöst werden kann [39, 50, 51]. Unabhängig von ihrer Ursache umfasst der Bereich der Neuroinflammation immer die Interaktion von Nervenzellen, Gliazellen und Immunzellen [52].

2.2.2 Chronischer neuropathischer Schmerz

Je nach Art der Schmerzentstehung existiert außer dem physiologischen oder pathophysiologischen Nozizeptor-Schmerz noch der neuropathische Schmerz, der durch Schädigungen der Nervenfaser entsteht, und somit seine natürliche Warnfunktion verloren hat. Während bei nozizeptiven Schmerzen die Nervenbahnen nur als Übermittler von Schmerzreizen dienen, ist bei den neuropathischen Schmerzen das Nervensystem selbst der Schmerzverursacher [53]. Diese Nervenschädigung kann durch eine Verletzung oder Infektion entstehen und im Zentralnervensystem sowie in der Peripherie auftreten [54].

Neuropathischer Schmerz betrifft bis zu zehn Prozent der gesamten Weltbevölkerung [55] und gilt immer noch als eine der größten Herausforderungen in der Schmerztherapie, da moderne Behandlungsmethoden oft nur in geringem Maße verfügbar sind und hauptsächlich aus einer Kombination von Opioiden mit neu entwickelten Antidepressiva und Antiepileptika bestehen, die den Schmerz zwar lindern, aber in einem Großteil der Patienten ihn weder komplett beseitigen, noch verhindern können, dass es zu einer Chronifizierung kommt [11, 56]. Dies zeigt sich auch in der aktuellen Förderpolitik der Europäischen Union, welche im Rahmen ihres siebten Forschungsrahmenprogramms (FP7) den Bereich der Schmerzforschung mit mehr als 46 Millionen Euro unterstützt hat, wobei von den acht geförderten Konsortien sich sechs Forschungsgruppierungen mit neuropathischen Schmerzen beschäftigt haben [11]. Es wird angenommen, dass Neuroinflammation und die Aktivierung von Gliazellen bei neuropathischen Schmerzen, aber auch bei anderen chronischen Schmerzzuständen, eine entscheidende Rolle spielen [52].

2.2.3 Gliazellen

Gliazellen (Mikroglia und Astrozyten) sind hochreaktive immunkompetente Zellen des Zentralnervensystems, die durch Entzündung, Infektion, Nerven- und Gewebeschädigung sowie Stress aktiviert werden können [57]. Sie haben sich innerhalb der letzten Jahre als wichtige Schlüsselfigur beim Verständnis der komplexen Mechanismen zur Entwicklung einer zentralen Schmerzsensibilisierung durch Neuroinflammation etabliert [58]. Ein charakteristisches Merkmal der Neuroinflammation ist nämlich die Aktivierung von Gliazellen im Bereich des Rückenmarks und Gehirns, was eine Freisetzung von pro-inflammatorischen Zytokinen und Chemokinen zur Folge hat [59]. Jüngste Studien legen außerdem nahe, dass Zytokine und Chemokine als starke Neuromodulatoren wirken können, und somit eine bedeutsame Rolle bei der Auslösung von Hyperalgesie und Allodynie spielen [52, 60, 61].

Ein weiterer Prozess, der mit der Aktivierung von Gliazellen in Verbindung gebracht wird, ist die Opioid-induzierte Hyperalgesie [62]. Die genauen komplexen molekularen Vorgänge und die Frage, ob die Mobilisierung des opiodergischen Systems in Folge chronischer Schmerzen die Gliazellen sowohl über endogene als auch über exogene Opioid-Rezeptor-Liganden aktivieren kann, ist bisher noch nicht vollständig geklärt [63].

2.3 Entwicklung der Systemmedizin

Lange Zeit war in der Schmerzforschung und auch im Bereich der molekularen Biowissenschaften eine Arbeitsweise vorherrschend, die den Fokus auf die Erforschung einzelner Gene oder Proteine gelegt hat. Es konnte somit ein umfangreiches Faktenwissen über einzelne Komponenten angesammelt werden. Das Zusammenspiel der ganzen Faktoren, die für die Abläufe komplexer zellulärer Prozesse in verschiedenen Gewebestrukturen und in komplexen Organismen notwendig sind, wurden dabei vernachlässigt [64]. Für eine ganzheitliche Betrachtung eines Organismus ist jedoch ein Zusammenspiel verschiedenster Fachbereiche wie der Biologie, Medizin, und Informatik notwendig. Den Kern dieser Systemmedizin bildet das ständige Wechselspiel zwischen Modellierung und Laborarbeit. Auf der Basis dessen, was im Labor entdeckt wird, können dann zu bestimmten Aspekten Computermodelle entworfen werden. Die wiederum werden im Labor erneut überprüft. Auf diese Weise kommt man Schritt für Schritt zu einem präziseren Resultat [65].

2.3.1 Das Zeitalter der “Omics“-Technologien

Insbesondere im Zeitalter der “Omics“-Technologien (Genomik, Proteomik, und Metabolomik) hat sich die biologische Forschung tiefgreifend verändert. Die Anwendung von modernen Methoden wie den Hochdurchsatz-Sequenzierungen (NGS), kombiniert mit interdisziplinärer Kooperation in den verschiedenen Teilbereichen der Biologie, führt dazu, dass wir in einem noch nie da gewesenen Maße komplexe Zusammenhänge in unserer Umwelt besser verstehen und erforschen können. Dies bedarf jedoch einer effektiven und effizienten Vernetzung von biologischen Daten [66].

2.3.2 Funktionelle Genomik

Das Forschungsgebiet der funktionellen Genomik untersucht die biochemischen, zellulären und physiologischen Eigenschaften von Genprodukten, mit dem Ziel die Beziehung zwischen dem Genom und dem Phänotyp besser zu verstehen. Dieser Prozess hat den Bereich der Arzneimittelforschung, gerade bei der Entwicklung neuer Analgetika, zu einer datenlastigen Disziplin gemacht, deren Anwendung erheblich vom parallel verlaufenden Fortschritt im Bereich der Informatik, vor allem in den Teilbereichen des maschinellen Lernens und der

künstlichen Intelligenz, abhängt [67]. Durch die Untersuchung funktioneller Einzelnukleotid-Variationen, welche Einfluss auf die Schmerzwahrnehmung haben oder die Reaktion auf Analgetika modulieren, und durch die tierexperimentelle Forschung an Modellorganismen, insbesondere an transgenen Mäuse, wurden mittlerweile mehr als 500 Gene identifiziert, die an der Modulation schmerzbezogener Phänotypen beteiligt sind [13].

Um Schmerzen auf genomweiter Ebene zu analysieren, werden mithilfe der funktionellen Genomik Daten kombiniert, die aus verschiedenen Prozessen im Zusammenhang mit DNA-Sequenz, Genexpression und Proteinfunktion stammen. Die meisten wissenschaftlichen Studien zu diesem Thema sind jedoch genomweite Expressionsanalysen oder genetische Assoziationsstudien, die nur eine kleine Anzahl von Genen beinhalten, während die Verwendung der funktionellen Genomik mit dem expliziten Ziel der Entdeckung neuer Wirkstoffe und Targets im Bereich der Analgetika vergleichsweise selten vorkommt [13].

Die Arzneimittelforschung konzentriert sich bei der Wirkstoffentdeckung oft auf Target-basierte Ansätze. Insbesondere die gängigen Arzneimittel-Targets sind in weltweit verfügbaren Datenbanken, wie zum Beispiel in der “DrugBank-Database“ [68], für jedermann zugänglich. Bei einer alternativen Herangehensweise in dieser Thematik liegt der Fokus nicht mehr nur auf dem Wirkstoff-Target, sondern konzentriert sich auf die Krankheit an sich und ihre zugrundeliegenden biologischen Prozesse. Dieser eher phänotypisch basierte Ansatz versucht die Komplexität von Krankheiten zu begreifen, ohne dabei das Wissen von Wirkstoffspektren einzelner Medikamente und deren Targets zu benutzen [69]. Die biologischen Prozesse, in denen die mit der Krankheit assoziierten Gene involviert sind, können ebenfalls aus öffentlich zugänglichen Datenbanken ausgelesen werden, allen voran die “Gene Ontology (GO) Database“ [70]. Dies hat den Vorteil, dass dieser funktionell genomische Ansatz sich nicht nur auf einzelne Arzneimittel-Targets stützt, sondern versucht die Krankheit an sich in ihrer ganzen Komplexität zu betrachten, was bei der Schmerzwahrnehmung, welche im Allgemeinen als ein sehr umfassender und komplexer molekularbiologischer und biochemischer Prozess angesehen wird, durchaus Sinn macht [71].

2.4 DNA-Sequenzierung

Der Begriff “DNA-Sequenzierung“ beschreibt die Bestimmung der genauen Nukleotid-Abfolge in einem DNA-Molekül. Dieser Prozess hat die molekularen Biowissenschaften revolutioniert und die Ära der Genomik eingeleitet. Um die Terminologie der DNA-Sequenzierung zu verstehen, ist es wichtig die molekulargenetischen Vorgänge zu kennen, die diesem Prozess zugrunde liegen.

2.4.1 Genetische Grundlagen der DNA-Sequenzierung

Desoxyribonukleinsäure (DNA) ist die Grundeinheit, deren Abfolge der einzelnen Basenpaare alle Informationen eines Organismus codiert, die für das Funktionieren seiner gesamten biologischen und biochemischen Prozesse erforderlich sind. DNA liegt in Form einer doppelsträngigen Helix vor, die aus vier Nukleotiden besteht, die wiederum unterschiedliche Basen enthalten: Adenin (A), Guanin (G), Cytosin (C) und Thymin (T). Diese DNA-Stränge sind in Chromosomen organisiert. Die Doppelstrangstruktur basiert auf der Komplementarität der einzelnen Basen. Adenin paart mit Thymin und Guanin paart mit Cytosin. Die DNA liefert auch die Vorlage, mit der die Messenger-Ribonukleinsäure (mRNA), durch einen Prozess namens “Transkription“ erzeugt wird. Die mRNA ist einzelsträngig und somit klein genug, um durch den Zellkern zu gelangen und wird anschließend durch einen als “Translation“ bezeichneten Prozess im Cytoplasma in eine Kette von Aminosäuren übersetzt, um ein funktionierendes Protein zu bilden. Ein Codon ist ein Satz von drei aufeinanderfolgenden Basen und jedes Codon kann in eine bestimmte Aminosäure übersetzt werden oder zeigt das Ende des Proteins an (zum Beispiel entspricht das Codon GTC der Aminosäure Valin und TAG ist eines der drei Stopp-Codons). Als menschliches “Genom“ bezeichnet man den kompletten Satz chromosomaler DNA, der beim Menschen ungefähr 3 Milliarden Basenpaare umfasst, die in 23 Chromosomenpaare unterteilt sind. Allerdings sind nicht alle dieser Basenpaare an der Codierung von Proteinen beteiligt, da das Genom aus Protein-codierenden Regionen (Exons) und nicht-codierenden Regionen (Introns) besteht. Der vollständige Satz von Protein-codierenden Regionen wird als "Exom" bezeichnet und stellt ungefähr ein bis zwei Prozent des Genoms dar [72].

2.4.2 Anwendungsbereiche von DNA-Sequenzierungen

Eine genetische Analyse kann zu medizinischen oder nicht medizinischen Zwecken (wie zum Beispiel bei Abstammungsanalysen) durchgeführt werden. Im klinisch-medizinischen Bereich zielt die DNA-Sequenzierung unmittelbar auf eine Diagnostik hin, wie zum Beispiel bei der Fragestellung nach dem Vorliegen einer vererbten oder erworbenen genetischen Abweichung [73]. Die jeweiligen Ansätze einer genetischen Analyse können sehr unterschiedlich sein und sind vom Umfang her von ihrem jeweiligen Anliegen abhängig. Es können einzelne, fest definierte Gen-Stellen im Fokus liegen oder auch umfangreiche genomweite Ansätze durchgeführt werden. Die Exom-Sequenzierung hat die Erfassung aller für die Proteinmoleküle codierenden Genabschnitte zum Ziel [74], bei einer Genom-Sequenzierung wird die Gesamtheit der Nukleotid-Sequenzen in allen 23 Chromosomenpaaren untersucht [75]. Eine eher punktuelle Analyse wird bei einer zielgerichteten Gen-Sequenzierung, oder auch Panel-Diagnostik, durchgeführt, bei der nur Gen-Stellen, die für eine bestimmte Diagnose relevant sind, untersucht werden [76]. Dies wird oft im Bereich der Pharmakogenetik durchgeführt, wo sich DNA-Sequenzierungen zur Feststellung von Genvarianten auf spezielle Gene fokussieren, die für die Wirkung oder den Abbau von bestimmten Medikamenten zuständig sind, oder im Bereich der Krebsdiagnostik, wo sich Analysen auf Veränderungen in Genen richten, deren Varianten sich krebsfördernd (Onkogene) oder krebshemmend (Tumorsuppressorgene) auswirken können [73]. Das Ergebnis einer genetischen Analyse erhält jedoch erst durch das Wissen um seine phänotypische Bedeutung für das untersuchte Individuum oder die untersuchte Gruppe eine Aussagekraft. Bei prädiktiver Diagnostik liegt hingegen noch kein relevanter Phänotyp vor. Die genetische Untersuchung hat hierbei vielmehr eine Voraussage zum Ziel, nämlich ob und mit welcher Wahrscheinlichkeit und unter welchen sonstigen Bedingungen bei einer Person ein bestimmter Phänotyp entstehen könnte [77].

2.4.3 “Next-Generation-Sequencing“

Der Begriff “Next-Generation-Sequencing“ (NGS), auf Deutsch “Gensequenzierung der nächsten Generation“ oder auch "Hochdurchsatz-Sequenzierung" genannt, umfasst eine Sammlung von genetischen Sequenzierungstechniken. Diese DNA- und RNA-Sequenzierungsmethoden verwenden viele parallele Prozesse, was zu einer schnelleren Verarbeitung führt, womit sie dem ursprünglichen Sanger-Sequenzierungsverfahren in den Punkten Zeitaufwand, Kosten und Genauigkeit der Sequenzierung überlegen sind [78, 79]. Beispiele hierfür sind unter anderem die Plattformen Illumina, Roche 454, Ion Torrent™ und SOLiD. In einer kürzlich veröffentlichten Studie konnte gezeigt werden, dass NGS bei der Detektion von Mutationen beim Gen *BRCA1/2* akkurater und sensitiver gewesen war, als die traditionelle Sanger-Methode [80].

Fast 30 Jahre lang war die Sanger-Sequenzierung der Goldstandart für die Mutationsdetektion in der molekularen Genetik. Diese Methode wurde 1977 von Frederick Sanger vorgestellt und 1980 mit dem Nobelpreis ausgezeichnet [81]. Sie beruht auf dem Kettenabbruch-Verfahren, bei der wie bei der Polymerase-Kettenreaktion (PCR) ein DNA-Strang von einer Polymerase abgelesen und eine komplementäre Kopie davon angefertigt wird. Der Vorgang der Sequenzierung erfolgte dann an einem einzelnen DNA-Fragment durch Kettenabbrüche und anschließender elektrophoretischer Auftrennung der unterschiedlich langen Syntheseprodukte [82]. Beim NGS erfolgt dieser Vorgang parallel, das heißt es werden Millionen von unterschiedlich langen DNA-Fragmenten gleichzeitig sequenziert. Dies wird erreicht, indem man die Einzelfragmente an einer Oberfläche fixiert, dies können beispielsweise kleine Kugeln auch “Beads“ genannt (Ion Torrent™) oder eine Glasoberfläche (Illumina) sein. Die fixierten Einzelfragmente werden dann anschließend um ein Vielfaches amplifiziert. Der Vorgang der Sequenzierung erfolgt lokal begrenzt an der jeweiligen Oberfläche, wodurch es möglich ist, parallel Millionen von Fragmente gleichzeitig und im selben Volumen zu synthetisieren [83].

Eine weitere große Erneuerung dieser neuen Technologien ist die Möglichkeit, dass die Sequenzier-Reaktion selbst jeweils Base für Base spezifisch sichtbar gemacht werden kann. Die jeweiligen Methoden, um diese Sequenzier-Reaktionen optisch aufzunehmen, sind bei jeder Plattform unterschiedlich realisiert worden. Bei der Ion Torrent™ Plattform kann der Einbau eines Nukleotids durch eine leichte Abänderung des pH-Werts gemessen werden, bei

der Illumina Plattfort ist die Nukleotid-Detektion an einen Fluoreszenzfarbstoff gekoppelt. Ist der eigentliche Sequenzier-Vorgang abgeschlossen, werden die zahlreichen sich teilweise überlappenden Einzelfragmente mit Hilfe einer speziellen und komplexen geräteinternen Analyse-Software wieder in die richtige Reihenfolge zusammengesetzt. Dies geschieht unter Zuhilfenahme des bekannten menschlichen Referenzgenoms (hg19). Weitere geräteinterne Schritte am Ende einer Sequenzierung umfassen die Zuordnung von zusammengesetzten Teilsequenzen an die korrekte Lokalisation im zutreffenden Chromosom und eine umfangreiche Fehlerkorrektur, denn erst in diesem Stadium der Erstellung einer Genomsequenz ist eine interpretierende Untersuchung von funktionell relevanten genetischen Varianten des gesamten Chromosoms sinnvoll möglich [77].

Da das Probenmaterial vervielfältigt werden muss und eine Polymerase-Kettenreaktion in der Theorie nie komplett fehlerfrei abläuft, können bei dem Sequenzier-Vorgang Unsicherheiten oder Fehler auftreten. Da das Genom aber nicht nur ein einziges Mal synthetisiert wird, sondern so viele Fragmente enthält, dass sie zusammen mindestens dem 30-fachen des ganzen Genoms entsprechen, kann ein sehr großer Anteil der auftretenden Fehler korrigiert werden. Man spricht in diesem Zusammenhang von einer “30-fachen-Coverage“, was auch in dem Begriff der “Hochdurchsatz-Sequenzierung“ wiedergegeben wird [84]. Der Arbeitsaufwand ist jedoch nach Abschluss des Sequenzier-Prozesses und dem anschließenden Sortieren der Sequenzstücke zu einem kompletten Genom noch lange nicht abgeschlossen. Nun erfolgt ein ebenso wichtiger wie komplexer Arbeitsschritt, das Auswerten der Sequenzdaten. Für diese Form der Datenanalyse wird eine komplett neue bioinformatorische Expertise benötigt, welche die herkömmliche Auswertung biomedizinischer Daten bei Weitem übersteigt [85].

2.4.4 Ion Torrent™ Technologie

Im Gegensatz zu anderen auf dem Markt existierenden NGS-Plattformen nutzt das Ion Torrent™-System für den Sequenzierungsprozess das Halbleiterverfahren, womit es möglich ist mittels integrierter Schaltkreise eine unmittelbare nicht-optische Genom-Sequenzierung durchzuführen. Dieses Prinzip basiert auf einem Halbleiter-Chip, ähnlich den Chips, die in Digitalkameras verbaut werden. Während Kamera-Chips eine Schicht mit Millionen von Pixeln besitzen, die Licht in digitale Informationen umwandeln, hat ein Ion Torrent™ Sequenzier-Chip Millionen kleiner Vertiefungen, in denen sich ionensensitive Sensoren befinden, in denen die Polymerase-Reaktion stattfindet. Die chemische Information dieses Vorgangs wird erfasst und in digitale Information umgewandelt.

Der Sequenzier-Prozess beginnt, nachdem ein DNA-Sample in Millionen einzelner Fragmente aufgetrennt wird. Jedes Fragment bindet dann an seinen eigenen “Bead“ und wird vervielfältigt, bis die gesamte Oberfläche mit Replikationen bedeckt ist. Dieser automatisierte Prozess findet an Millionen verschiedener “Beads“ und mit Millionen verschiedenen DNA-Fragmente statt. Alle 15 Sekunden werden nacheinander alle vier möglichen Nukleotide hinzugegeben, wird ein Nukleotid nach dem Prinzip der komplementären Basenpaarung in den DNA-Strang eingebaut, wird dabei ein Wasserstoff-Ion freigesetzt. Das freigelassene H+ Ion ändert den pH-Wert der Lösung und diese minimale Abänderung wird direkt auf dem Chip registriert. Direkt unter jeder Vertiefung befindet sich die Ionen-sensitive Schicht, welche die pH-Veränderung erkennt und in Strom umwandelt. Bei jeder eingebauten Base wird also ein Stromsignal abgegeben und anhand der Intensität des gemessenen Signals können Wiederholungen desselben Nukleotids identifiziert werden. Zusammengefasst gesagt: Das Ion Torrent™ System kombiniert einen einfachen chemischen Prozess mit bewährter Halbleiter-Technologie und vereinfacht, beschleunigt und vergünstigt somit den Sequenzier-Vorgang gegenüber dem traditionellen Sanger-Verfahren erheblich.

Aufgrund der Halbleiter-Technologie ist das System prinzipiell auch weniger störanfälliger als andere Sequenzier-Methoden, wie zum Beispiel das Pyrosequencing, da zum Auslesen der Informationen keine photometrischen Verfahren verwendet werden [86]. Außerdem ist zu erwarten, dass sich Sequenzier-Kosten und die Sequenzier–Geschwindigkeit durch den Verbau integrierter Schaltkreise, ähnlich wie bei Computerchips, durch immer komplexere Herstellungsverfahren laufend verbessern (Mooresches Gesetz) [87]. Die Kosten für die NGS-

Analyse eines Patienten betragen momentan ungefähr 150 Euro [88, 89], wobei das immer abhängig von der jeweiligen Größe des Gen-Panels und der Größe der gewünschten Sequenzier-Abdeckung ("Coverage") ist. Ein Großteil der Kosten fällt dabei auf die Anfertigung des Primer-Panels und die Konstruktion der Templates, nur ein kleiner Anteil von ungefähr 10 Prozent ist für den eigentlichen Sequenzier-Vorgang aufzuwenden [89].

2.4.5 Ion Torrent™ Workflow

Der Ion Torrent™ Workflow lässt sich in drei Bereiche aufteilen: (i) Anfertigung des Template, (ii) Sequenzierung und (iii) Daten-Analyse. Zunächst ist es bei einer zielgerichteten DNA-Sequenzierung jedoch erforderlich, die einzelnen Amplicons für die gewünschten Targets zu einem Primer-Panel zu synthetisieren. Bei der Ion Torrent™-Plattform geschieht das über das Online-Tool AmpliSeq™ Designer [84]. Ein AmpliSeq™-Panel besteht aus einem Pool von Oligonukleotid-Primerpaaren und bildet die Basis der Primer-Bibliothek, wobei jedes Paar zur Amplifikation einer bestimmten genomischen Region bestimmt ist. Einzigartig bei diesem Ansatz ist die Fähigkeit, dass bis zu 24,000 Primer-Paare in einer einzigen PCR-Reaktion vereint werden können, was bedeutet, dass man in einem einzelnen Sequenzier-Durchgang eine Vielzahl an Genen integrieren kann [84].

2.4.5.1 Template Konstruktion

Die Template-Konstruktion umfasst den Aufbau einer Bibliothek von Nukleinsäuren (DNA oder komplementäre DNA) und deren Amplifikation auf Basis des vorher konturierten AmpliSeq™-Pools. Diese Sequenzier-Bibliotheken werden konstruiert, indem die DNA-Probe zunächst fragmentiert wird und danach Adaptersequenzen (synthetische Oligonukleotide einer bekannten Sequenz) an die Enden der DNA-Fragmente ligiert werden. Einmal angefertigt, bilden die Bibliotheken den ersten Schritt der Sequenzierung und können für viele verschiedene Sequenzier-Vorgänge amplifiziert werden. Die Ion Torrent™ "Personal Genome Machine"-Einheit (PGM) verwendet eine Emulsions-PCR auf Basis des IonOneTouch™-Systems, um Fragmente einer einzelnen Bibliothek auf Mikrokügelchen zu amplifizieren, während das MiSeq-System der Firma Illumina eine Brückenamplifikation verwendet, um Matrizencluster auf einer Durchflusszelle zu bilden [86, 90].

Um eine Nukleinsäure-Sequenz aus den amplifizierten Bibliotheken zu erhalten, stützen sich sowohl das Ion Torrent™ PGM als auch das MiSeq auf das Prinzip “Sequenzierung durch Synthese“ (“Sequencing by Synthesis“). Es wird also die Synthese eines komplementären Stranges Nukleotid für Nukleotid verfolgt. Die Bibliotheks-Fragmente fungieren dabei als Matrize, aus der ein neues DNA-Fragment synthetisiert wird. Die Sequenzierung erfolgt durch einen Wasch- und Floating-Zyklus der einzelnen Fragmente mit den bekannten Nukleotiden in sequentieller Reihenfolge. Wenn Nukleotide in den wachsenden DNA-Strang eingebaut werden, werden sie digital als Sequenz aufgezeichnet. Das PGM und das MiSeq besitzen einen unterschiedlichen Mechanismus zum Nachweis der Sequenzinformationen. Das PGM führt eine Halbleitersequenzierung durch, die auf dem Nachweis von pH-Änderungen beruht, die durch die Freisetzung eines Wasserstoff Ions beim Einbau eines Nukleotids in einen wachsenden DNA-Strang induziert werden. Im Gegensatz dazu beruht der MiSeq auf dem Nachweis der Fluoreszenz, die durch den Einbau fluoreszenzmarkierter Nukleotide in den wachsenden DNA-Strang erzeugt wird [86].

2.4.5.2 Datenanalyse

Sobald der Sequenzier-Vorgang abgeschlossen ist, müssen die Rohsequenzdaten mehreren Analyseschritten unterzogen werden, um sie verwertbar zu machen. Eine verallgemeinerte Datenanalyse-Pipeline für NGS-Daten umfasst (i) eine Vorverarbeitung der Daten, um Adaptersequenzen und Lesevorgänge von geringer Qualität zu entfernen, (ii) die Zuordnung der Daten zu einem Referenzgenom oder die “de novo“-Ausrichtung der Lesevorgänge (Sequenzier-Vorgang ohne Referenzgenom) und (iii) die Analyse der zusammengesetzten Sequenzen. Der letzte Schritt kann eine Vielzahl von bioinformatischen Methoden umfassen. Es existieren einige freizugängliche Software-Tools um grundlegende Datenauswertungen auf der Basis von Nukleotid-Variationen, dem Nachweis neuer Gene oder regulatorischer Elemente und die Bewertung der Transkriptions-Expressionsgrade vorzunehmen [91]. NGS bietet einen umfassenden Zugang auf die Sequenzdaten des Genoms und kann auch problemlos zur Diagnose krankheitsassoziierter genetischer Merkmale eingesetzt werden [92]. Die Interpretation der Daten bleibt jedoch aufgrund der Größe und Komplexität des Genoms immer noch eine bioinformatische Herausforderung.

2.4.5.3 Genetische Variationen

Einzelnukleotid-Variationen (SNVs) treten auf, wenn eine Base durch eine andere ersetzt wird (zum Beispiel ein Guanin für ein Thymin). Dies verändert die DNA an einem einzelnen Punkt und wird daher auch als Punktmutation bezeichnet. Der Effekt dieser Punktmutation kann dramatisch sein und man klassifiziert SNVs aufgrund ihrer Auswirkungen in "Missense", "Silence" und "Nonsense". Eine “Missense”-Mutation führt zu einem Aminosäurewechsel, zum Beispiel ein Wechsel von GTA zu GAA würde bewirken, dass die Aminosäure von Valin zu Glutamin wechselt. Eine “Silence”-Mutation führt nicht zu einer Änderung der Aminosäure, da Redundanz besteht, was bedeutet, dass viele Aminosäuren von mehreren Codon-Tripletts codiert werden. Eine “Nonsense“-Mutation führt zur Einführung eines Stopp-Codons. Beispielsweise führt eine Mutation von TCA zu TAA zur Produktion des Stopp-Codons TAA anstelle der Aminosäure Serin. Dies kann dann dazu führen, dass ein kleineres Protein produziert wird, da die DNA nach dem Stopp-Codon nicht weiter translatiert wird, was die Funktion des Proteins wahrscheinlich erheblich beeinträchtigt.

Insertionen und Deletionen sind weitere Varianten von Einzelnukleotid-Abweichungen und werden auch als “Indels” bezeichnet. In diesem Fall werden eine oder mehrere Basen in die Sequenz eingefügt oder aus der Sequenz gelöscht, was zu einer Verschiebung des kompletten Leserasters führt, insofern die Anzahl der eingefügten oder gelöschten Nukleotide nicht drei oder ein Vielfaches von drei ist.

Als Polymorphismus bezeichnet man eine in einer Population häufig vorkommende Abweichung der DNA-Sequenz. Ob eine genetische Variation als Mutation oder als Polymorphismus bezeichnet wird, hängt von der Häufigkeit ab, in der sie in der Population auftritt. Hier wurde ein Grenzwert von einem Prozent festgelegt, was bedeutet, dass eine genetische Variation als Polymorphismus bezeichnet wird, wenn sie in mehr als in einem Prozent der Bevölkerung auftritt, und als Mutation, wenn sie in weniger als einem Prozent der Bevölkerung vorkommt [93].

Single Nucleotide Polymorphismen (SNPs) sind die am häufigsten vorkommenden Typen von SNVs. Viele SNPs treten in nicht-codierenden Bereichen des Genoms auf und die Mehrzahl der SNPs hat keine, zumindest bekannte, klinische Relevanz [94]. Andere SNPs sind jedoch mit einer veränderten Reaktion auf bestimmte Medikamente oder dem Risiko der Ausbildung bestimmter Krankheitsbilder assoziiert, wobei es aber schwierig ist, einen einzelnen verantwortlichen SNP zu lokalisieren, da viele SNPs korrelieren und erst gemeinsam als Masse eine erkennbare Auswirkung auf den Phänotyp haben [95].

2.5 Computionale Wissensgenerierung

In den letzten Jahren ist das Volumen von generierten Daten in den molekularen Biowissenschaften vor allem durch den massiven Fortschritt im Bereich der NGS-Technologien exponentiell angestiegen [85]. Weil die Technologie der Sequenzier-Verfahren immer besser wird und gleichzeitig auch die Kosten sinken, steigt demnach auch die Menge der generierten Daten. Zum Beispiels besitzt ein Ion Torrent^{TM} S5XL-System die Kapazität, um innerhalb eines Jahres mehr als 18 Tera-Basen (10^{12} Basen) zu produzieren [84], was bedeuten würde, das man täglich das komplette Genom eines Menschen sequenzieren könnte. Zum Vergleich, das "Human Genome Project", an dem mehr als 1.000 Wissenschaftler beteiligt gewesen sind, benötigte zur vollständigen Entschlüsselung des menschlichen Genoms über 12 Jahre [96].

Die Kosten für die Sequenzierung eines menschlichen Genoms betragen derzeit weniger als 1.000 US-Dollar [97] und sind somit auf ihrem bisher niedrigsten Wert, was es nun möglich macht, groß angelegte Sequenzier-Projekte zu finanzieren und das in einem Volumen, wie es vor einigen Jahren noch nicht denkbar gewesen wäre. Nicht enthalten sind in diesen Preisprojektionen allerdings die Kosten für die Interpretation der umfangreichen und komplexen Daten. Diese dürften trotz entsprechenden Fortschritten im Bereich der "Data Science" auch weiterhin vergleichsweise hoch bleiben [77]. Man schätzt trotzdem, dass innerhalb des nächsten Jahrzehnts bis zu zwei Milliarden menschliche Genome fertig sequenziert werden könnten, was mehr als ein Viertel der aktuellen Weltbevölkerung ausmacht [98]. Dieser rasante Fortschritt wird sich erheblich auf viele Forschungsgebiete auswirken, vor allem auf Bereiche der personalisierten Medizin. Des Weiteren sind spürbare Auswirkungen bei der Entdeckung und Erforschung neuer Wirkstoffe und Arzneimittel absehbar [99].

2.5.1 Maschinelles Lernen

Maschinelles Lernen bezeichnet eine Ansammlung informatischer Methoden, die dazu in der Lage sind automatisch Muster in Daten zu erkennen und diese dann zu nutzen, um zukünftige Daten vorherzusagen, sie zu klassifizieren, Strukturen wie Untergruppen zu identifizieren oder Informationen aus den Daten zu extrahieren, die dann wiederum dazu geeignet sind, um an ihnen neues Wissen ableiten zu können [100].

Fast jeder hat im Alltag bereits Erfahrungen mit maschinellen Lernsystemen gemacht und nutzt sie auch regelmäßig. Die Rede ist von Spracherkennungssoftware, so wie “Siri“, “Cortana“, und “Alexa“. Sprachgesteuerte Assistenzsysteme sind allerdings nur eine der vielen Anwendungsbereiche für selbständiges maschinelles Lernen. Zu den weiteren Anwendungen gehören die automatische Betrugserkennung, Empfehlungssysteme in Onlineshops oder die DNA-Analyse [101]. Dank einer rasanten Weiterentwicklung von Computersystemen hat sich auch der Bereich der “Data Science“ schnell entwickelt. Seit Kurzem stehen potenziellen Datenwissenschaftlern regelrechte Supercomputer zur Verfügung, die in der Lage sind, die für das maschinelle Lernen notwendige hohe Rechenleistung zu erbringen. Mit acht Tesla-P100-Grafikprozessoren ausgestattet leistet der Rechner des Herstellers “Nvidia“ 170 Tera-FLOPS, wobei ein Tera-FLOP 10^{12} Operationen pro Sekunde entspricht. Für diese Leistung benötigte man hierfür bislang bis zu 250 Intel-Xeon-E5-Systeme [102]. Diese Leistung wird möglich, da Grafikprozessoren dazu in der Lage sind Dutzende oder sogar Hunderte von Rechenvorgängen parallel zu verarbeiten [103].

Die Methoden des maschinellen Lernens erlauben es Muster in großen und komplexen Datenmengen zu identifizieren, ohne dass die Struktur der Muster explizit gegeben sein muss. In diesem Zusammenhang wird oft der Begriff “Big Data“ genannt. Dies bezeichnet elektronisch erzeugte und gespeicherte Datensätze, die sowohl in ihrer Größe als auch in ihrer Komplexität so umfangreich sind, dass neue algorithmische Techniken erforderlich sind, um aus ihnen nützliche Informationen zu gewinnen [100]. Diese lernenden Systeme können dabei nicht nur Gesetzmäßigkeit in großen Datenmengen erkennen und daraus Regeln zur Klassifizierung ableiten, sie treffen auch Vorhersagen über zukünftige Ereignisse, bereiten wissensbasiert Entscheidungen vor oder treffen diese selbstständig ohne menschliches Zutun [104].

Je nachdem was für Daten man vorliegen hat oder was man mit ihnen anstellen möchte, gibt es unterschiedliche algorithmische Ansätze. Grundsätzlich funktionieren diese verschiedenen Ansätze nach einem ähnlichen Schema. Das System erhält zunächst Trainingsdaten anhand derer es ein Model ableitet, dieses wird dann mit Hilfe von Testdaten überprüft und optimiert. Nach einigen Durchgängen kann das Verfahren dann auf Daten angewendet werden deren Klassifikation unbekannt ist, um Muster oder Assoziationen zu finden [103]. Die verschiedenen Verfahren des maschinellen Lernens lassen sich grob in zwei verschiedene Kategorien einteilen. Man unterscheidet zwischen dem "Unüberwachten Lernen" ("Unsupervised Learning"), also ein Lernen ohne menschliches Zutun und ohne im Voraus bekannte Zielwerte, und dem "Überwachten Lernen" ("Supervised Learning"), bei dem ein menschlicher "Trainer" die Lernergebnisse bewertet und die Ergebnisse durch Naturgesetze oder Expertenwissen bekannt sind und genutzt werden um das System anzulernen [105]. Unüberwachte Algorithmen eignen sich vor allem für die Mustererkennung in Datensätzen. Beispiele sind hierfür das "k-Means-Clustering"-Verfahren oder die Hauptkomponenten-Analyse [106]. Das überwachte Lernen findet häufig Anwendung um Vorhersagen zu treffen oder Klassifizierungen vorzunehmen. Beispiele für überwachtes Lernen sind die "k-Nearest-Neighbor"-Analyse, "Support-Vektor-Machines" oder "Random Forests".

2.5.2 Maschinelles Lernen in der Schmerzforschung

Die Schmerzforschung gehört zu den Bereichen, in denen große Mengen an multidimensionalen Daten aufgenommen werden, da sich Schmerz aufgrund seiner komplexen Pathophysiologie in komplexen und heterogenen klinischen Phänotypen ausdrückt [107]. Die Verwendung moderner DNA-Sequenzierungs-Methoden wie dem NGS führt zwar zu uneingeschränktem Zugang zum menschlichen Genom, produziert jedoch auch "Big Data", das in Menge und Komplexität die Auslese-Fähigkeit der klassischen analytischen Ansätze bei Weitem übertrifft [108]. Durch die computergestützte Verarbeitung komplexer schmerzbezogener Datensätze ist es möglich Informationen effektiver zu extrahieren. Maschinelles Lernen hat aufgrund dieser neuen Art der Wissensgenerierung die Möglichkeit das pathophysiologische Verständnis von Schmerzen zu verbessern und somit die Behandlungsmöglichkeiten von Schmerzpatienten positiv zu beeinflussen [100].

2.6 Wissenschaftliche Hintergründe dieser Arbeit

Diese Arbeit wurde im Rahmen eines, von der Europäischen Union geförderten Projekts des Forschungsrahmenprogramms 7 (FP7), durchgeführt. Das Projekt GLORIA (Glial Opioid Receptor Interface in Analgesia) beschäftigte sich mit Neuroinflammation und der Rolle der Gliazellen-Aktivierung bei chronischen Schmerzzuständen. Ziel war es, neuartige Analgetika zu entwickeln und Moleküle zu synthetisieren, die entweder auf die Gliazellen-Aktivierung über den Toll-like-Rezeptor Reaktionsweg abzielen, analog sind zu endogenen Opioid-Liganden oder an die gleichen Rezeptoren binden, so wie die Liganden aus der Biomolekül-Familie der neuroprophen Faktoren. Die Hauptaufgabe der Arbeitsgruppe aus Frankfurt am Main bestand in der Identifizierung genetischer Biomarker bei chronischen Schmerzzuständen. Dazu wurde ein sehr komplexer Datensatz von der Universität Helsinki zur Verfügung gestellt, basierend auf einer Kohorte von 1.000 Frauen kaukasischer Herkunft, die sich einer Brustkrebs Operation unterzogen hatten, und umfasste außer DNA-Proben auch umfangreiche Schmerzdaten [109]. Außerdem standen die modernen Technologien der Hochdurchsatz-Sequenzierung (Ion Torrent™ Plattform) zur Verfügung.

2.6.1 Etablierung der Ion Torrent™ Plattform

Manuskript #1 Next-generation sequencing of human opioid receptor genes based on a custom AmpliSeq™ library and ion torrent personal genome machine

Zunächst wurde die neue Methode der Ion Torrent™ Sequenzierung etabliert [89]. Da die DNA-Menge des von der Universität Helsinki zu Verfügung gestellten Datensatzes nur in begrenzter Menge vorlag, und die Kosten für eine NGS-Sequenzierung vergleichsweise hoch sind, wurde zunächst mit bereits archivierter und projektfremder humaner DNA gearbeitet, um sicher zu stellen, dass die Etablierung des Ion Torrent™-Systems mit Proben durchzuführt werden konnte, welche in ausreichender Menge zur Verfügung standen. Dazu wurde auf einen älteren Datensatz zurückgegriffen, der aus DNA-Proben von Patienten bestand, welche zum Erzielen einer analgetischen Wirkung ungewöhnlich hohe Mengen an Opioiden benötigten, obwohl dafür keine medizinische Ursache vorlag [88].

Das Opioid-System ist an der Kontrolle des Belohnungssystems, Suchtverhalten aber auch bei der Schmerzwahrnehmung beteiligt. Opioide üben ihre pharmakologische Wirkung durch die agonistische Bindung an endogene Opioid-Rezeptoren aus und Variationen in den jeweiligen codierenden Genen können die Expression oder Signalgebung der Opioid-Rezeptoren abändern. Aufgrund der Tatsache, dass die Opioid-Rezeptoren auch einen der Forschungsschwerpunkte des GLORIA-Projektes ausmachten, war der vorhandene Datensatz hervorragend für eine NGS-Analyse geeignet.

Um eine zielgerichtete exonische Genotypisierung durchzuführen, wurde mit Hilfe des Online-Tools AmpliSeq™ ein Primer-Panel für die Gene der humanen Opioid-Rezeptoren *OPRM1* *OPRD1* und *OPRK1* und der Opioid-ähnlichen Rezeptoren *OPRL1* und *SIGMA1* entworfen. Dann wurde der NGS-Workflow in der Kohorte von 79 Schmerzpatienten mit teilweise sehr hohen Opioid-Gaben durchgeführt, um die exonischen Sequenzen der codierenden Genbereiche zu analysieren und anschließend zu bewerten.

Die Sequenzier-Abdeckung ("Coverage") ist abhängig von der Panel-Größe und um die vom Hersteller empfohlene Mindestanforderung bei der "Coverage" zu erreichen, ist es bei einem großen Gen-Panel erforderlich, die Patienten auf mehrere Sequenzier-Läufe zu verteilen. Damit sichergestellt werden konnte, dass alle einzelnen Läufe eine vergleichbare Sequenzier-Qualität aufwiesen, wurde ein besonderes Augenmerk auf eine gleichbleibende Messgenauigkeit gelegt. Außerdem wurde eine Methode etabliert, um die NGS-Messergebnisse zu validieren. Dazu wurden mehrere zufällig ausgewählte DNA-Proben in einem externen und unabhängigen Labor per Sanger-Sequenzierung nochmals analysiert und anschließend miteinander verglichen [89].

2.6.2 Etablierung bioinformatischer Methoden zur Auswertung von "Big Data"

Manuskript #2 Emergent biomarker derived from next-generation sequencing to identify pain patients requiring uncommonly high opioid doses

Die NGS-Analyse der im vorangegangenen Kapitel beschriebenen 79 Schmerzpatienten konnte erfolgreich durchgeführt werden, die verschiedenen Sequenzier-Läufe entsprachen den vom Hersteller empfohlenen Qualitätsanforderungen [84] und lieferten in der gesamten Population in allen Opioid-Rezeptoren insgesamt 152 Einzelnukleotid-Variationen. Es stellte sich bei der Datenanalyse jedoch heraus, dass das klinische Problem des hohen Opioid-Bedarfs nicht anhand einzelner Varianten zu erklären war. Einzelne funktionelle genetische Variationen werden für die genetische Diagnose klinischer Phänotypen zunehmend als unzureichend betrachtet [110]. Die Akkumulation mehrerer genetischer Varianten, die einzeln nur geringen funktionelle Einflüsse haben, kann jedoch erkennbare Auswirkungen haben [88]. Deswegen war es das Ziel, die durch NGS produzierte Datenmenge in einen klassifizierenden Biomarker zu konvertieren, um die Fragestellung erörtern zu können, warum einige Patienten der sequenzierten Kohorte ohne ersichtlichen Grund extrem hohe Mengen an Opioiden benötigten. Dieser selbst-lernende und sich selbst-verbessernde Biomarker sollte in der Lage sein, mit Hilfe der NGS-Daten einzelne Patienten richtig zu klassifizieren, also die jeweiligen Patienten in die Gruppe der "Hohe-Opioid-Dosen" oder "Niedrige-Opioid-Dosen" richtig zuordnen zu können. Außerdem sollte er das Potenzial haben bisherige klassische Ansätze zur genetisch basierten Patientenklassifikation zu übertreffen, deren Funktionalität durch Komplexität der vorhandenen Menge an "Big Data" nicht mehr gegeben war [88].

2.6.3 Charakterisierung und Identifizierung von Schmerzphänotypen

Manuskript #3 Next-generation sequencing of the human TRPV1 gene and the regulating co-players LTB4R and LTB4R2 based on a custom AmpliSeq™ panel & Manuskript #4 *Machine-learned analysis of the association of next-generation sequencing–based human TRPV1 and TRPA1 genotypes with the sensitivity to heat stimuli and topically applied capsaicin*

Ein Schmerzphänotyp kann als ein Maß definiert werden, das direkt oder indirekt die gesamte und teilweise Verarbeitung des Schmerzsystems widerspiegelt. Er dient als eine Art Fenster für die der Schmerzwahrnehmung zugrundeliegenden pathophysiologischen neuronalen Mechanismen und kann somit auch als eine Art Leitfaden für die Entwicklung einer personalisierten Schmerzmedizin dienen [111]. Ebenso ist die Existenz von Untergruppen in der jeweiligen Untersuchungskohorte eine häufige und oft auch bedeutsame Fragestellung, da aus deren Charakterisierung Rückschlüsse auf den sogenannten datengenerierenden Prozess geschlossen werden kann [107].

Diese Fragestellung war bei der gewünschten Assoziation von komplexen Genotypen mit Schmerzphänotypen essenziell, erforderte aber bei der durch die NGS-Analyse des Datensatzes aus Helsinki zu erwarteten umfangreichen Menge an “Big Data“ eine Entwicklung neuer bioinformatischer Methoden. Deshalb wurde die Etablierung dieser analytischen Anwendung an einem kleineren NGS-Datensatz durchgeführt. Dazu wurde auf eine Kohorte gesunder Patienten aus einer zuvor veröffentlichten Studie zurückgriffen, bei der die Rolle von Capsaicin und UV-B als Sensibilisierungsmethode gegenüber Hitzereizen erörtert wurde [112]. Mit Hilfe von NGS wurden 75 gesunde Probanden genotypisiert und die exonischen DNA-Sequenzen der für die TRP-Kanäle codierenden Gene *TRPA1* und *TRPV1* erzeugt. Die bei dem NGS-Ansatz erzeugten genetischen Informationen enthielten 278 Gen-Loci mit Abweichungen zum Referenzgenom in Form von Einzelnukleotid-Variationen [30].

Die zellulären Ionenkanäle der TRP-Familie spielen eine wichtige Rolle in der Schmerzwahrnehmung und vermitteln nozizeptive Sensibilisierung und Hyperalgesie [113, 114]. TRPV1 sowie TRPA1 sind aktuell das Ziel neuartiger Analgetika und gelten als vielversprechende Targets bei der Entwicklung von neuen Medikamenten [30]. Von daher könnten genetische Varianten von TRPV1 und TRPA1 vielversprechende Kandidaten für pharmakogenetische Modulationen von Arzneimittelwirkungen sein [29].

Bei dieser Analyse wurde die Fragestellung verfolgt, ob es bei den genotypisierten Probanden, bei denen im Vorfeld Hitzeschmerzschwellen vor und nach Sensibilisierung mit lokal aufgetragenem Capsaicin gemessen wurden, Untergruppen von Reaktionsmustern auf Hitzeschmerz gibt und ob diese Untergruppen sich hinsichtlich des Genotyps von TRPV1 und TRPA1 unterscheiden [115].

Nach Datenvisualisierung und anschließender Modellierung mittels eines Gaußschen Mischmodells konnten zwei Untergruppen definiert werden, deren Mitglieder sich entweder durch niedrige Hitzeschmerzsensitivität, welche durch das Auftragen von Capsaicin deutlich verstärkt wurde, oder durch eine von vornherein hohe Schmerzsensitivität gegenüber Hitzereizen auszeichneten. Darüber hinaus wurde eine weitere unabhängige und computerbasierte Analyse mit dem GWAVA-Tool [116] durchgeführt um die biologische Auswirkung nicht-codierender Gen-Variationen zu ermitteln und diese Ergebnisse mit denen des maschinellen Lernens zu vergleichen [115].

2.6.4 Funktionelle genomische Charakterisierung der Schmerz-Chronifizierung

Manuskript #5 A machine-learned analysis of human gene polymorphisms modulating persisting pain points to major roles of neuroimmune processes

Die wissenschaftliche Forschung hat im Bereich der Humangenetik innerhalb der letzten 10 Jahre in mehr als hundert Genen viele Einzelnukleotid-Variationen identifiziert, welche modulierend auf den chronischen Schmerzphänotyp einwirken. Dank einer stetigen Verbesserung in vielen Bereichen der Informatik, insbesondere bei künstlicher Intelligenz, maschinellem Lernen und Wissensgenerierung aus Datenbanken, ist es möglich, komplexe molekularbiologische Zusammenhänge besser zu verstehen und das Zusammenspiel der gesamten Faktoren, die für die Abläufe komplexer zellulärer Prozesse in verschiedenen Gewebestrukturen und in komplexen Organismen notwendig sind, detaillierter zu charakterisieren. Dadurch besteht nun die Möglichkeit das aus klinischen Studien generierte Wissen über genetische Modulationen mit aktuellen Kenntnissen über Gen-Funktionen zu kombinieren und somit dem klinischen Phänomen der Schmerz-Chronifizierung auf Ebene der funktionellen Genomik zu begegnen. Viele wissenschaftliche Analysen konzentrieren sich im Bereich der Schmerzforschung auf einzelne bestimmte Gene und meistens sind das oftmals die gleichen Gene, die schon seit Jahren im Fokus klinischer Forschungsarbeit liegen. Im Rahmen dieser Arbeit sollte einem innovativeren Ansatz nachgegangen werden und somit wurde die Fragestellung verfolgt, ob eine Zusammenführung des Wissens aller momentan bekannten schmerzrelevanten Gene neue Erkenntnisse und einen komplexeren Einblick in die Pathophysiologie von chronischen Schmerzen beim Menschen liefern kann [117]. Dafür wurde auf bioinformatische Methoden des maschinellen Lernens und verfügbare empirische Daten über die Modulation langanhaltender Schmerzzuständen durch humane Polymorphismen zurückgriffen, um die grundlegenden pathophysiologischen Prozesse zu identifizieren, die bei der Chronifizierung von Schmerzen beim Menschen eine Rolle spielen. Ziel dieser Auswertung war eine Charakterisierung der funktionellen Genomik des chronischen Schmerzphänotyps basierend auf der biologischen Rolle und Funktion von den Genen, die für den Prozess der Chronifizierung modulierend sind, um anschließend die wichtigsten biologischen Funktionen ihrer Genprodukte zu analysieren.

2.6.5 Charakterisierung chronischer Schmerzgenotypen

Manuskript #6 Development of an AmpliSeq™ Panel for Next-Generation Sequencing of a Set of Genetic Predictors of Persisting Pain

Eine genetische Komponente, die dazu beiträgt die individuelle Wahrnehmung von Schmerzen und den pathophysiologischen Prozess der Chronifizierung besser zu verstehen, ist seit mehr als einem halben Jahrhundert Thema molekularbiologischer Forschung. Mittlerweile weiß man, dass mehr als 500 Gene an diesen Prozessen beteiligt sind [13]. Während dieses Wissen es aktuell ermöglicht, gemeinsam mit anderen experimentellen und klinischen Forschungsstrategien potenzielle Arzneimittel-Targets zu identifizieren, um die Entwicklung neuartiger Analgetika zu intensivieren, ist unser Verständnis von der komplexen genetischen Architektur des Schmerzes und seine Entwicklung in Richtung einer Chronifizierung noch immer unvollständig.

Um die Assoziation der genetischen Muster, die den schmerzbezogenen Phänotypen zugrunde liegen, auf einer breiteren Ebene zu untersuchen und nicht nur allgemein anerkannte Kandidaten-Gene zu verwenden, die schon seit Jahren mit Schmerz-Chronifizierung in Verbindung gebracht werden, basierte die vorliegende Auswahl des Gen-Sets auf zuvor durchgeführten systembiologischen Analysen [13, 117]. Das Wissen von 535 empirisch als schmerzrelevant identifizierter humaner Gene wurde mit dem Wissen über die biologischen Funktionen Tausender humaner Gene kombiniert. Diese Analysen ergaben schließlich ein Set von 77 “Schmerz-Chronifizierungs-Genen“, bestehend aus drei Untergruppen aus verschiedener Evidenz, welches dann als Basis für NGS-Panel diente [118].

(i) Die erste Gruppe bestand aus Genen, welche für die Steuerung biologischer Lernprozesse und neuronale Plastizität verantwortlich sind [13]. Als Basis für diese Hypothese diente die bereits wissenschaftlich anerkannte Sichtweise [119, 120] dass sich Schmerz als eine Art Deregulierung bestimmter biologischer Prozesse abbilden lässt, welche in der (GO)-Datenbank unter den Termen “Learning or Memory“ und “Nervous System Development“ zusammengefasst werden. Es wurde eine funktionelle Genomanalyse durchgeführt, die eine Untergruppe von 34 Schmerzgenen identifizierte, welche mit beiden GO-Termen annotiert waren.

(ii) Die zweite Gruppe umfasste Gene, die funktionelle Polymorphismen enthielten, welche die Schmerz-Chronifizierung in mindestens zwei unterschiedlichen klinischen Schmerzzuständen modulieren. Sie setzte sich zusammen aus der Schnittmenge von 535 sogenannten “Schmerz-Genen“ mit 22 Genen, die aus einer unabhängigen funktionellen genomischen Analyse (Kapitel 2.6.4) in Kombination mit Methoden des maschinellen Lernens stammten, und bildete den zweiten Gen-Satz von 13 Genen, welche auf funktioneller Ebene hauptsächlich dem Immunsystem zugeordnet werden konnte [117].

(iii) Die dritte Untergruppe bildete sich aus Genen, welche in den letzten Jahren im Bereich der Schmerzforschung wiederholt eine wichtige Rolle bei der Modulation von langanhaltenden Schmerzen spielten, oder teilweise auch erst kürzlich als neuartige und vielversprechende genetische Modulatoren gehandelt wurden. Diese dritte Gruppe wurde erstellt, um bei der NGS-Analyse auf die altbewährten und intensiv erforschten Kandidaten-Gene nicht zu verzichten und um das Gen-Panel auch so aktuell wie möglich zu halten.

2.6.6 Genotyp-Phänotyp Assoziation bei chronischen Schmerzpatienten

Manuskript #7 Machine-learned analysis of the association of next-generation sequencing based genotypes with persistent pain after breast cancer surgery

Chronische Schmerzen sind ein ernst zu nehmendes und weitreichendes Problem, von dem etwa ein Fünftel der europäischen Bevölkerung betroffen ist [6]. Das bessere Verständnis von Schmerz und den pathophysiologischen Prozessen seiner Chronifizierung liegt im Fokus aktueller Forschungsaktivitäten und findet auf präklinischer, humanexperimenteller und klinischer Ebene statt [107]. Die Beteiligung genetischer Faktoren findet dabei immer mehr Beachtung und aktuelle wissenschaftliche Erkenntnisse deuten darauf hin, dass die individuelle Wahrnehmung von Schmerz und sein Chronifizierungs-Prozess durch eine Vielzahl von Gen-Varianten moduliert wird [121-123].

DNA-Sequenzierungen, insbesondere das NGS sind geeignete Methoden, um Wissen über funktionelle genetische Polymorphismen zu generieren. Das Erkennen von Datenstrukturen in der jeweiligen Untersuchungskohorte ist gerade bei komplexen Genotyp-Phänotyp Assoziation essenziell. Anhand maschinell erlernter Datenprojektionen können schmerzbezogene Phänotypen ermittelt werden [100], diese erfordern aber gerade bei NGS-basierten Analysen durch die produzierte komplexe und große Menge an “Big Data“ die Entwicklung neuer bioinformatischer Methoden. Derzeit richtet sich ein besonderes Interesse auf das maschinelle Lernen, welches Methoden zur Detektion interessanter, insbesondere biologisch aussagekräftiger Strukturen in hochdimensionalen Daten bereitstellt, um sogenannte Klassifikatoren zu erstellen, die dazu in der Lage sind, klinische Phänotypen aus genetischen Merkmalen vorherzusagen [107].

Die bioinformatische Methodik für eine KI-basierte Assoziation komplexer Genotypen mit schmerzbezogenen Phänotypen wurde bereits in einer vorherigen Studie erfolgreich etabliert [115] und sollte nun auf den viel umfangreicheren Datensatz angewendet werden, welcher die Basis dieser Analyse war. Dieser sehr große Datensatz basierte auf einer Kohorte von 1.000 Frauen kaukasischer Herkunft, die sich einer Brustkrebs Operation unterzogen hatten und der von der Universität Helsinki zur Verfügung gestellt wurde [109]. Er enthielt außer DNA-Proben auch umfangreiche Schmerzdaten, welche im Rahmen einer 3-jährigen Nachuntersuchung erhoben wurden. Die Hauptidee bei dieser Analyse war das Trainieren einer künstlichen Intelligenz, implementiert als verschiedene Arten des maschinellen Lernens, um Assoziationen von genetischen Informationen mit dem Schmerz-Phänotyp zu erlernen und anschließend diese trainierte Form von künstlicher Intelligenz anzuwenden, um in neuen Daten ebenfalls einen Phänotyp vorhersagen zu können [108].

2.7 Ziel dieser Arbeit

Das Ziel dieser Doktorarbeit bestand darin, genetische Biomarker für chronische Schmerzen zu identifizieren. Das biologische Ausgangsmaterial stammte von gesunden Probanden sowie von Schmerzpatienten. Die Haupthypothese betraf zunächst die Glia-Opioid-Interaktion, was auf die wichtige Rolle für die Opioid-Gene (zum Beispiel *OPRM1*) und Toll-like-Rezeptor-Gene (zum Beispiel *TLR4*) als wichtiges Targets bei chronischen Schmerzen hinweist. Das Panel sollte noch auf neue potenzielle Gene, welche im klinischen Kontext mit der Schmerz-Chronifizierung assoziiert werden können, mittels systematischer “Pathway-Analysen“, funktioneller Genomik und systembiologischer sowie bioinformatischer Methoden erweitert werden. Die vorliegenden Analysen umfassten die Auswertung genetischer, klinischer, psychologischer und demografischer Parameter, anhand derer innerhalb der Probanden Untergruppen identifiziert werden sollten, welche dieselben Phänotyp-Merkmale bezüglich chronischer Schmerzen besitzen. Zusätzlich sollten verschiedene informatische Methoden wie das maschinelle Lernen sowie NGS-Methoden angewendet werden, um große und komplexe Datenmengen (“Big Data“) zu generieren und auszuwerten. Dies sollte zu der Identifizierung von schmerzbezogenen Untergruppen innerhalb der Patienten führen, die wiederum die Assoziation mit genetischen Markern ermöglichen sollte, mit denen Patienten ermittelt werden können, die ebenfalls zu diesen Phänotyp-Gruppen gehören.

3. Diskussion

Im Rahmen dieser Doktorarbeit wurden in gesunden Probanden und in Patienten mit einem pathophysiologischen Risiko zur Schmerz-Chronifizierung relevante Genotypen sowie Phänotypen identifiziert und charakterisiert. Die Genotypisierung aller Probanden für die Generierung der exonischen Nukleotid-Sequenzen schmerzrelevanter Gene erfolgte durch Etablierung der "Next Generation Sequencing" (NGS) Methode der Ion Torrent™ Plattform. Mit Hilfe bioinformatischer Expertise konnte nach der Etablierung neuer Methoden im Bereich des maschinellen Lernens in allen analysierten Datensätzen biologisch plausible phänotypische Untergruppen definiert werden. Bioinformatische Auswertungen auf der Basis funktioneller Genomik führten zur Identifikation von Gene, welche für die Schmerz-Chronifizierung relevante Polymorphismen enthalten und bildeten eine potenzielle Basis für die Entdeckung neuer Arzneimittel-Targets. Auf dessen Grundlage wurde ein Gen-Satz bestehend aus 77 humanen "Schmerz-Chronifizierungs-Genen" zusammengestellt, welche mutmaßlich den chronischen Schmerzphänotyp modulieren. Dabei wurden verschiedene Denkansätze verfolgt und altbekannte Gene, an denen bereits jahrelang intensiv geforschte wurde, sowie Gene, welche mit Hilfe bioinformatischer Analysen basierend auf funktioneller Genomik identifiziert werden konnten, inkludiert. Es wurden in einer Kohorte, bestehend aus 1.000 Frauen kaukasischer Herkunft, nach einer Brustkrebs-Operation und anschließender dreijähriger Nachkontrolle Genotyp-Phänotyp Assoziationen durchgeführt, um zu erörtern, ob die Genotypen genug Informationen enthalten, um die Patienten entweder als "Schmerz-Chronifizierer" oder als "Nicht-Schmerz-Chronifizierer" zu kategorisieren.

3.1 Ion Torrent™ Etablierung und Auswertung von NGS-“Big Data“

Manuskript #1 und Manuskript #2

Im Rahmen dieser Doktorarbeit wurde der NGS Workflow basierend auf einer selbsterstellten AmpliSeq™ Primer-Bibliothek für die Gene der humanen Opioid-Rezeptoren etabliert und in einer Kohorte von 79 Patienten erfolgreich durchgeführt [89]. Die vorliegenden Ergebnisse legen nahe, dass es sich bei dem entwickelten NGS-Ansatz um eine hocheffiziente Methode zur Mutations-Detektion handelt. Alle wichtigen Polymorphismen, welche im Interesse aktueller Forschungen sind, wurden von dem NGS-Assay abgedeckt. Zusätzlich wurde durch die hohe Abdeckung der Amplicons über die exonischen Opioid-Rezeptor-Sequenzen eine große Menge an genetischer Information verfügbar gemacht, welche als Grundlage für weitere genetische Analysen über die funktionellen Konsequenzen der Einzelnukleotid-Variationen menschlicher Opioid-Rezeptors dienen kann.

Die Analyse durch den Ion Torrent™ Workflow ergab insgesamt 152 Einzelnukleotid-Variationen, wobei jedoch nur wenige Nukleotid-Variationen im Hinblick ihrer Allel-Frequenz statistisch signifikante Gruppenunterschiede aufwiesen. Das pathophysiologische Phänomen des extrem hohen Opioid-Bedarfs dieser Patienten lässt sich daher eher als eine Ansammlung vieler genetischer Variationen mit nur geringer Funktionalität betrachten, welche erst in ihrer Summe heraus einen Einfluss auf den Phänotypen haben. Die Auswertung der durch die NGS-Analyse erzeugten Menge an “Big Data“ erfolgte über einen “selbstlernenden subsymbolischen klassifizierenden Biomarker“, basieren auf dem “k-Nearest-Neighbor“-Algorithmus [88]. Diese Methode aus dem Bereich des maschinellen Lernens ist ein Klassifikationsverfahren, bei dem eine Klassenzuordnung unter Berücksichtigung seiner nächsten Nachbarn vorgenommen wird. Der Begriff der “Subsymbolik“ beschreibt dabei einen Klassifikator, bei dem die Tatsache, warum bestimmt Daten in bestimmte Gruppen zugeordnet wurden, nicht mehr nachvollziehbar ist. Es ist also kein direkter Einblick in die Lösungswege möglich. Auf diese Analyse bezogen würde das bedeuten, dass die genetischen Muster, welche durch die einzelnen Nukleotid-Abweichungen repräsentiert werden, nicht direkt mit einer biologischen Konsequenz assoziiert sein müssen.

Ein Klassifikator gilt als “selbstlernend“, wenn richtig zugeordnete Daten zu seinem Wissensfundus hinzugefügt werden und seine Prognosen dadurch immer besser werden, er somit “dazulernt“. Die Auswertung der NGS-Daten mit diesem Biomarker zeigte, dass die Genotypisierung der Opioid-Rezeptoren mit biologischer Plausibilität verbunden war. Nach einer Selektion der 34 Marker mit dem höchsten Informationsgehalt mit Hilfe des informatischen Tools der ABC-Analyse [124] konnte der subsymbolische Biomarker die Patienten mit einer diagnostischen Treffsicherheit von über 80 Prozent erfolgreich klassifizieren. Da seine Applikation nicht nur auf die Opioid-Rezeptoren beschränkt ist, hat seine Anwendung durch die Erfassung emergenter genetischer Strukturen im Bereich der klinischen Diagnostik großes Potenzial.

3.2 Identifizierung von schmerzbasierten Phänotypen

Manuskript #3 und Manuskript #4

Die Datengrundlage für dieses Projekt bildete eine NGS-Analyse, die im Rahmen dieser Doktorarbeit mit dem Ion Torrent™ Workflow durchgeführt wurde [89]. Dazu wurde eine AmpliSeq™ Primer-Bibliothek für Gene der humanen TRP-Kanal Rezeptoren und zugehörige Co-Rezeptoren erstellt und deren exonische Sequenzen in 75 gesunden Probanden erfasst [30]. In einer zuvor veröffentlichten Studie wurden in den gleichen Probanden Schmerzdaten über Hitze- und Capsaicin-Empfindlichkeit erhoben [112]. In den beiden Kandidaten-Genen TRPV1 und TRPA1 wurden insgesamt 278 Gen-Loci gefunden. Mit klassischer Chi-Quadrat-Statistik konnte jedoch kein signifikanter Genotyp-Unterschied zwischen den Gruppen nachgewiesen werden.

Mittels Methoden des maschinellen Lernens ist es in innovativen Ansätzen möglich, komplexe Genotypen mit schmerzbezogenen Phänotypen zu assoziieren [107]. Die durch NGS-Anwendungen produzierten großen und komplexen Datenmengen an genetischen Informationen sind jedoch gerade im Hinblick auf komplexe Genotyp-Phänotyp Assoziationen eine bioinformatische Herausforderung. Aus diesem Grund wurde mit Hilfe bioinformatischer Expertise ein neuartiger Ansatz entwickelt, um die mit Hilfe von NGS gewonnenen Mengen an "Big Data" zu verarbeiten. Die Datenanalyse wurde in mehreren Schritten durchgeführt.

(i) Der erste Schritt zielte darauf ab den "Output-Space" der Datenausgabe zu ermitteln, das heißt eine Phänotyp-Klassenstruktur zu definieren, in diesem Fall also die Verteilung der Veränderungen der Hitzeempfindlichkeit nach Capsaicin-Applikation. Eine Gaußsche Mischmodellierung [125] zeigte zwei Phänotyp-Gruppen mit hoher oder niedriger Capsaicin-induzierter Überempfindlichkeit gegen Hitze an.

(ii) Dann erfolgte eine Merkmals-Selektion basierend auf Methoden des überwachten maschinellen Lernens, um die informativsten genetischen Varianten zu identifizieren [124], da nicht jedes Merkmal eines Datensatzes den gleichen nachhaltigen Wert besitzt und ein Klassifikator mit reduziertem Merkmalsraum nicht nur schneller arbeitet, sondern in der Regel auch weitaus effizienter [126].

(iii) Nach der Datenvorverarbeitung und Reduktion auf 31 Gen-Loci erfolgte unüberwachtes maschinelles Lernen, das als “Swarm-Clustering“ [127] implementiert wurde und auf Unterschiede im genetischen Muster zwischen diesen Phänotyp-Gruppen hindeutete. Ziel dieses Analyse-Schrittes es war aus den reduzierten genetischen Daten heraus Gruppenstrukturen zu definieren, welche im Idealfall die Gruppenstruktur des Schmerzphänotyps widerspiegelt, da eine Übereinstimmung zwischen den Genotyp- und Phänotyp-Datenstrukturen eine Assoziation zwischen den beiden Datensätzen erleichtert.

(iv) Diese Genotyp-Phänotyp Assoziation wurde anschließend durch überwachtes maschinelles Lernen unter Verwendung verschiedener Arten von Algorithmen basierend auf künstlicher Intelligenz durchgeführt. Dazu wurden Klassifikations-Verfahren wie “k-nearest Neighbors“ [128], “Logistic Regressions“ [129], “Random Forests“ [130], und “Adaptive Boosting“ [131] mit einem aus dem Gesamtdatensatz zufällig ausgewählten Trainingsdatensatz, der zwei Drittel des ursprünglichen Datensatzes ausmacht, trainiert, um anhand der Genotyp-Information die Zugehörigkeit in der Phänotyp-Gruppe für das restliche Drittel des Datensatzes vorherzusagen.

Als Negativ-Kontrolle wurden die Allele der einzelnen Gen-Loci zufällig vermischt (permutiert), sodass die Geninformation “sinnlos“ wurde. Ebenso wurde ein idealer Genotyp konstruiert, bei dem fast alle Allele einer Phänotyp-Gruppe zugeordnet wurden. Dies diente als positive Kontrolle, um sicherzustellen, dass die Klassifikations-Verfahren prinzipiell in der Lage sind, solch ein Problem zu lösen. Im Ergebnis zeigte sich, dass alle Algorithmen bei Verwendung des vorliegenden Genotyps den Phänotyp besser vorhersagen konnten, als es bei der Verwendung einer zufälligen Permutation der Gensequenzen der Fall war. Somit konnte eine KI-basierte Assoziation des TRPA1/TRPV1-Genotyps mit Schmerzempfindlichkeit gegenüber Hitzereizen demonstriert werden [115].

3.3 Charakterisierung von Genotypen durch funktionelle Genomik

Manuskript #5

Die computerbasierte funktionelle genomische Analyse brachte einen Einblick in die pathophysiologischen Prozesse der Schmerz-Chronifizierung aus einer systembiologischen Perspektive [117]. Sie enthielt mehrere Methoden des maschinellen Lernens [100], die man als eine Ansammlung von bioinformatischen Tools verstehen kann, mit denen es möglich ist, Strukturen in Datensätzen aufzudecken, um dann anhand zukünftiger Daten Prognosen oder Klassifikationen zu erstellen.

Die Analyse umfasste 110 wissenschaftlich anerkannte "Schmerz-Gene". Der Fokus lag dabei im Vorfeld auf Genen, die in acht unterschiedlichen klinischen Schmerzzuständen (wie zum Beispiel Rückenschmerzen, Nervenschmerzen oder Muskelschmerzen) bereits mehrmals mit dem klinischen Aspekt der Schmerz-Chronifizierung assoziiert wurden. Die "Gene Ontology (GO) Database" beschreibt die biologische Funktion jedes Gens anhand der biologischen Prozesse, in welche die Gene involviert sind und definiert diese Rolle als eine Aneinanderreihung biologisch relevanter Ereignisse oder als molekulare Funktionen, sogenannter "GO-Terms". Die computerbasierte Auslesung der GO-Daten, basierend auf einer Überrepräsentations-Analyse [132] und mehreren Methoden des maschinellen Lernens [69, 133, 134], ergab eine starke heterogene Gruppierung (Cluster) der an der Schmerz-Chronifizierung beteiligten Gene, ohne klar erkennbaren Fokus auf die bei diesem Prozess sonst charakteristischen Merkmale.

Darum wurde eine zweite Analyse durchgeführt, in der die Hypothese verfolgt wurde, dass sich die funktionelle Genomik chronischer Schmerzen über die vorher definierten acht Schmerzzustände hinaus und somit ungeachtet von der klinisch definierten Krankheit, welche diesem Prozess ursprünglich zugrunde liegt, definiert. Das Ziel dieser zweiten Analyse war anhand der in der (GO)-Datenbank enthaltenen Informationen über chemische, zelluläre und physiologische Eigenschaften der Genprodukte ein besseres Verständnis über die Beziehung zwischen Genotyp und Phänotyp zu erhalten, um funktionelle Muster zu erkennen, die einen Einblick in die Pathophysiologie chronischer Schmerzzustände ermöglichen.

Mit Hilfe eines bioinformatischen Klassifikationsverfahrens [124] wurden genetische Untergruppen definiert, in denen die vielversprechendsten Gene enthalten waren und führte zu einer Auswahl von 22 Genen. Deren funktionelles genomisches Abbild ergab nach einer Überrepräsentations-Analyse, dass (i) das Immunsystem und (ii) die Stickoxid-Signalgebung zu den biologischen Prozessen gehören, deren Rolle bei der Schmerz-Chronifizierung von besonderer Bedeutung sind. Dieses Ergebnis ist aus biologischer Sicht sehr plausibel, da bei beiden Prozessen eine Beteiligung an der Schmerz-Chronifizierung bereits in früheren wissenschaftlichen Studien nachgewiesen werden konnte [135, 136]. Stickoxid wird bei nozizeptiven Reizen in den Neuronen des Rückenmark-Hinterhorns verstärkt produziert, wodurch dann vermehrt bestimmte Neurotransmitter freigesetzt werden, was wiederum zu einer zentralen Schmerz-Sensibilisierung und somit zu einer Chronifizierung führen kann [137]. Des Weitern kann das Immunsystem mit dem Prozess der Schmerz-Chronifizierung auf einer sich überlagernden neuroimmunen Ebene über die Glia-Opioid-Schnittstelle interagieren [136].

Die Rolle des Immunsystems bei der Entwicklung chronischer Schmerzzustände konnte nach einer Reduzierung des vorliegenden Gen-Satzes noch weiter hervorgehoben werden. Dazu wurde die Schnittmenge der 22 Gene mit einem weiteren Gen-Satz von 535 sogenannten “Schmerz-Genen“ [13], welcher in einer unabhängigen Analyse erzeugt wurde, gebildet. Das Ergebnis war eine Liste von 13 Genen, welche hauptsächlich in die Zytokin-Produktion involviert sind, was die Rolle des Immunsystems an pathophysiologischen lang anhaltenden Schmerzzuständen noch unterstreicht [117].

3.4 Identifizierung genetischer Prädiktoren für chronische Schmerzen

Manuskript #6

Die Existenz einer genetischen Komponente, welche auf die individuelle Wahrnehmung von Schmerz und den pathophysiologischen Prozess seiner Chronifizierung funktionell einwirkt, wird durch das Wissen über viele Schmerz-modulierende humane genetische Polymorphismen deutlich [121-123]. Bis vor wenigen Jahren konzentrierte sich die molekularbiologische Forschung auf die Rolle einzelner funktioneller genetischer Varianten. Die sinkenden Kosten von DNA-Sequenzierungen und die bessere Verfügbarkeit von NGS-Methoden haben diesen Trend innerhalb der letzten Jahre abgeschwächt [138] und aktuelle Analysen umfassen nun die gesamten codierenden Bereiche und regulatorischen Sequenzen schmerzrelevanter Gene [30, 89, 118].

Im Rahmen dieser Doktorarbeit wurde ein Set von 77 humanen "Schmerz-Chronifizierungs-Genen" vorgestellt, welches als Grundlage für ein NGS-Panel diente, basierend auf der Ion Torrent™ Technologie [118]. Die Genotypisierung wurde an 72 Probanden kaukasischer Herkunft durchgeführt und alle generierten Nukleotid-Sequenzen deckten sich mit den Sequenzen aus der nach dem Sanger-Verfahren durchgeführten Validierung. Die Auswahl der 77 Gene basierte auf drei Untergruppen aus verschiedener Evidenz.

Die Hypothese der ersten Gruppierung zielte auf die Annahme ab, dass sich Schmerz als eine Art Deregulierung der biologischen Prozesse abbilden lässt, welche in der (GO)-Datenbank unter den Termen "Learning or Memory" und "Nervous System Development" zusammengefasst werden [119, 120]. Sie umfasste unter anderem Gene, die mit dem mesolimbischen dopaminergen System assoziiert sind, beispielsweise das für die Tyrosinhydroxylase kodierende Gen *TH* und die für die Dopamin-Rezeptoren kodieren Gene *DRD1*, *DRD2* und *DRD3*, welche bereits mit der Ausbildung chronischer Rückenschmerzen assoziiert werden konnten [139]. Weitere 14 Gene waren an circadianen Rhythmen beteiligt, was ein wissenschaftlich anerkannter modulatorischer Faktor für verschiedene Schmerzzuständen darstellt [140, 141]. Darüber hinaus waren noch Gene enthalten, welche für metabotrope Glutamat-Rezeptoren codieren, sowie drei NMDA-Rezeptor-Gene (*GRIN1*, *GRIN2A* und *GRIN2B*), welche in vielen essentiellen physiologischen Prozessen eine Rolle spielen [142].

Die zweite Teilgruppe enthielt Gene, welche anhand einer umfangreichen Literaturrecherche identifiziert werden konnten. Eine zuvor durchgeführte funktionelle genomische Analyse (Kapitel 2.6.4) zeigte eine Beteiligung dieser Gen-Gruppe an Immunprozessen und der Stickoxid-Signalgebung [117]. Enthalten sind unter anderem Gene codierend für mehrere Interleukine (*IL1B*, *IL4*, *IL6*, *IL10*), bei denen gezeigt worden ist, dass sie an immunologischen Schmerzmechanismen beteiligt sind [143, 144]. Außerdem enthielt diese Teilgruppe die in metabolische Prozesse involvierten Gene *TNF*, *GCH1* und *ESR1* [145-147].

Die dritte Untergruppe schließlich bildeten Gene, die in den letzten Jahren im Bereich der Schmerzforschung bei der Modulation chronischer Schmerzen verstärkt im Fokus lagen, oder teilweise auch erst kürzlich als neuartige und vielversprechende genetische Modulatoren aufgekommen sind. Das inkludiert unter anderem die Gene codierend für Mitglieder der TRP-Kanal-Familie (*TRPA1*, *TRPM8* und *TRPV4*), die bekanntermaßen wichtige Akteure bei der Schmerz-Wahrnehmung über ihre Erregung durch chemische, thermische oder mechanische Reize darstellen [148]. Gleiches gilt auch für das opioiderge System, welches durch die Gene der Opioid-Rezeptoren *OPRM1*, *OPRK1* und *OPRD1* repräsentiert wird, die ebenfalls in der dritten Untergruppe enthalten sind und mit der Modulation verschiedenster Schmerz-Zustände assoziiert werden [149].

Bei der Entwicklung des Gen-Sets von 77 humanen “Schmerz-Chronifizierungs-Genen“ wurde die Strategie verfolgt, die Pathophysiologie der Schmerz-Chronifizierung eher aus Sichtweise der funktionellen Genomik abzubilden und nicht vollständig, aber jedoch teilweise von der sonst üblichen Vorgehensweise, bei der sich die Auswahl passender Kandidaten-Gene meistens auf zuvor veröffentlichte Studien konzentriert, abzuweichen. Während genomweite Assoziationsstudien mittlerweile einige vielversprechende, für Schmerz-Phänotypen relevante, Gene identifiziert haben, werden wissenschaftliche Studien schmerzrelevanter Gene immer noch von einer gleichbleibenden und vergleichsweise kleinen Anzahl von Kandidaten-Genen dominiert [150]. Mehr als die Hälfte aller schmerzrelevanter Gen-Studien entfallen auf die Analyse von nur etwa zehn Genen oder Gen-Komplexen und diese Studien werden teilweise auch noch mehrfach repliziert [122].

Ein Vergleich des ausgewählten Set von Genen mit zwei unabhängigen humanen Gen-Sätzen, welche ebenfalls mit der Modulierung von chronischem Schmerzzuständen assoziiert werden [122, 123], ergab eine nur geringfügige gegenseitige Überlappung [118]. Dies unterstreicht die Tatsache, dass das Verständnis für die genetische Architektur des Schmerzes noch immer unvollständig ist, und es von daher sinnvoll sein kann, unterschiedliche und voneinander unabhängige Forschungsansätze zu verfolgen. Interessant war auch die Tatsache, dass der hier vorgestellte Gen-Satz im Vergleich zu den beiden alternativen Vorschlägen die größte Übereinstimmung mit der Menge der 535 “Schmerz-Gene“ erzielte, welche die wohl aktuell umfangreichste Auflistung humaner Gene repräsentiert, die wissenschaftlich nachgewiesen auf den Schmerzphänotyp modulierend einwirken [13].

3.5 Komplexe Genotyp-Phänotyp Assoziationen

Manuskript #7

Die Physiologie des Schmerzes umfasst komplexe immunologische, sensorische, hormonelle und inflammatorische Prozesse im Rückenmark, im Gehirn und in der Peripherie. Wiederholte nozizeptive Stimulation induziert pathophysiologische Veränderungen bei der Schmerzweiterleitung, die zu peripherer oder zentraler Sensibilisierung führen, was bei dafür anfälligen Patienten zu der Ausbildung chronischer Schmerzen führen kann. Obwohl das Wissen über die genauen molekularen Vorgänge, die zu einer Chronifizierung führen, noch immer unvollständig ist, sind die Identifizierung von Risikofaktoren [108] sowie die psychosoziale Bewertung der Patienten [151] vernünftige Schritte, um die individuelle Anfälligkeit für die Entwicklung chronischer Schmerzen zu bestimmen.

Die Hauptidee dieser Analyse war das Trainieren einer künstlichen Intelligenz (KI), implementiert als verschiedene Arten des maschinellen Lernens, um ihr Assoziationen der genetischen Informationen mit dem Schmerz-Phänotyp anzulernen und anschließend die trainierte KI zu verwenden, um in neuen Daten einen Phänotyp vorhersagen zu können. Dabei wurde der innovative Ansatz verfolgt, bei dem eine durch NGS-basierte genetische Informationen trainierte KI, den entsprechenden Phänotyp besser vorhersagen sollte, als dass es durch Raten möglich wäre. Komplexe Genotyp-Phänotyp Assoziationen mit der jeweiligen vorhandenen Datengrundlage wären somit realisierbar.

Aufgrund der zu erwartenden hohen Dimensionalität und Kollinearität der komplexen genetischen Informationen, wurden Methoden des maschinellen Lernens von vornherein traditionellen Ansätzen, wie der logistischen Regressionsanalyse, vorgezogen. Dieser bioinformatische Ansatz für eine KI-basierte Genotyp-Phänotyp Assoziation wurde bereits in einer vorherigen Studie erfolgreich etabliert [115] und nun auf den viel umfangreicheren Datensatz angewendet, der als Datengrundlage für diese Analyse diente.

Der Datensatz basierte auf einer Kohorte von 1.000 Frauen kaukasischer Herkunft im Alter von 28 bis 75 Jahren, welche an nicht metastasierendem Brustkrebs litten. Nach der operativen Behandlung wurde die Intensität der postoperativen Schmerzen in mehreren Abständen für eine Dauer von 3 Jahren anhand von Fragebögen erfasst [109]. Aus dieser Kohorte wurden zwei Gruppen von Schmerzphänotypen identifiziert, bei der jeweils 70 Patienten zu einer "Schmerz-Chronifizierer" oder einer "Nicht-Schmerz-Chronifizierer" Gruppe zugeordnet wurden, sich sonst aber in anderen Parametern wie Gewicht und Alter nicht unterschieden [151].

Im Rahmen dieser Doktorarbeit wurden diese 140 Probanden basierend auf der Ion Torrent™ Technologie genotypisiert. Hierbei wurde das in einer zuvor veröffentlichten Studie vorgestellte NGS-Panel der 77 humanen "Schmerz-Chronifizierungs-Gene" verwendet [118]. Die Genauswahl erfolge unter der Zusammenführung von drei Evidenzlinien und wird in Kapitel 2.6.5 ausführlich erläutert. Das Panel enthielt Gene welche (i) für die Steuerung biologischer Lernprozesse und neuronale Plastizität verantwortlich sind [13], basierend auf der Sichtweise [119, 120], dass sich Schmerz als eine Art Deregulierung bestimmter biologischer Prozesse abbilden lässt, welche in der (GO)-Datenbank unter den Termen "Learning or Memory" und "Nervous System Development" zusammengefasst werden, und (ii) Gene, die funktionelle Polymorphismen enthalten, welche die Schmerz-Chronifizierung in mindestens zwei unterschiedlichen klinischen Schmerzzuständen modulieren [117], und (iii) Gene, welche in den letzten Jahren im Bereich der Schmerzforschung wiederholt eine wichtige Rolle bei der Modulation von lang anhaltenden Schmerzen spielten oder teilweise auch erst kürzlich als neuartige und vielversprechende genetische Modulatoren gehandelt wurden [118].

Nach erfolgreicher Validierung der in Kapitel 2.6.2 ausführlich erläuterten bioinformatischen Methode wurde die Analyse der NGS-Daten der beiden Schmerz-Phänotyp-Gruppen (chronischer Schmerz oder nicht-chronischer Schmerz) durchgeführt, welche aus der ursprünglichen Kohorte der 1.000 mit Brustkrebs behandelten Patienten identifiziert wurden. Während dieser Analyse wurde die genetische Information von den ursprünglichen 4.748 Varianten, die in den 77 Genen gefunden wurden, auf 21 relevante Varianten reduziert, die in 13 verschiedenen Genen lokalisiert waren. Das gefundene Muster stützte eine Assoziation der Genotypen mit den Schmerzphänotypen.

Überwachte Klassifikatoren auf der Basis maschinellen Lernens, die mit zwei Drittel der Daten trainiert wurden, identifizierten die richtige Schmerz-Phänotyp-Gruppe bei dem verbleibenden Drittel der Patienten mit einer Genauigkeit von etwa 70 Prozent [108]. Die Analyse identifizierte die genetischen Unterschiede zweier Phänotyp-Gruppen im Hinblick auf die Entwicklung einer Schmerz-Chronifizierung nach einer operativen Behandlung von Brustkrebs und obwohl die Auswahl Kandidaten-Gene größtenteils auf der Grundlage früherer wissenschaftlicher Studien basierte, lieferten die vorliegenden Ergebnisse eine Gen-Rangfolge, welche sich mit ihrer besonderen Bedeutung ihres klinischen Kontextes vereinbaren lässt.

Ein interessantes Ergebnis der vorliegenden Analyse ist, dass sich einige der wichtigsten Varianten in Genen befinden, die an dopaminergen Signalwegen beteiligt sind, so wie die Rezeptor-Gene *DRD1*, *DRD3*, *DRD4* und das Tyrosinhydroxylase-Gen *TH*. Ebenso waren Varianten in Genen, welche in katecholaminerge und serotonerge Stoffwechselwege involviert sind (*COMT*, *HTR2A*, *HTR3A*, *SLC6A2* und *SCL6A3*) hochplatzierte Kandidaten, die möglicherweise zu einer Entwicklung von chronischen Schmerzen nach einer Brustkrebs-Operation beitragen. Katecholamin und Serotonin dienen hauptsächliche als Neurotransmitter und spielen eine Rolle in der peripheren und zentralen Schmerzentwicklung [152]. Diese Neurotransmitter aktivieren bei Gewebeverletzungen periphere Nozizeptoren und tragen somit zu einer zentralen Sensibilisierung des Nervensystems bei [153] und eine Funktionsstörung bei katecholaminerger und serotonerger Neurotransmission ist mit der Ausbildung verschiedener chronischer Schmerzzustände assoziiert [154].

Trotz dieser positiven Ergebnisse offenbarte diese Analyse auch einige Einschränkungen, die an dieser Stelle auch kurz diskutiert werden sollen. Die Kohorte, welche schlussendlich einer NGS-Analyse unterzogen wurde, war verhältnismäßig klein, da nur etwa zehn Prozent der ursprünglich registrierten 1.000 Patientinnen als “Schmerz-Chronifizierer“ definiert werden konnten. Außerdem war eine Replikation der Analyse an einer unabhängigen Kontroll-Gruppe aufgrund der Einzigartigkeit der Kohorte und des Datensatzes nicht ohne weiteres möglich.

Zudem wurde bei dieser Analyse wegen der technischen Limitation eines Ampliseq™ Panels [84] ein Satz von Kandidaten-Gene verwendet, der zwar unter anderem durch eine genomweite Gen-Auswahl entstanden ist, jedoch an die Vollständigkeit einer kompletten Genom-Sequenzierung nicht herankommt. Eine weitere Einschränkung dieser Studie kam durch das Weglassen seltener Varianten bei Allel-Frequenzen unter 10 Prozent zustande. In der Tat ist es theoretisch möglich, dass die Entwicklung chronischer Schmerzen auch von einigen wenigen sehr seltenen Nukleotid-Variationen reguliert wird und das darüber hinaus bei jedem Patienten eine andere Variante vorliegt, welche somit nicht in die Auswertung mit einbezogen wurden. Eine Analyse dieser Varianten ist aber mit maschinellen Lernsystemen, so wie sie angewendet wurden, nicht möglich, bietet jedoch die Möglichkeit für die zukünftige Entwicklung alternativer Analyse-Verfahren und ist ein interessanter Aspekt, der nicht komplett außer Acht gelassen werden sollte.

Gleichwohl aller Kritikpunkte unterstützt die Tatsache, dass mehr als 75 Prozent der bei dieser Analyse identifizierten Gene Targets von bereits zugelassenen oder sich noch in der Entwicklung befindenden Schmerzmitteln sind, die Wichtigkeit dieser Ergebnisse. Der vorliegende Satz von Gen-Varianten kann als ein wichtiger Baustein dienen um den Weg zu ebnen für die Etablierung einer personalisierten Schmerz-Therapie, wenn einige der sich noch in der Entwicklung befindlichen Medikamente klinisch verfügbar werden.

3.6 Ausblick

Die rasanten Fortschritte in der wissenschaftlichen Forschung, vor allem in den Fachrichtungen “Data Science“, NGS und funktioneller Genomik werden uns in den nächsten Jahren vor große Herausforderungen stellen. Mittlerweile hat die generierte Datenmenge in diesen Bereichen eine Größenordnung erreicht, die ohne selbstlernende Computeralgorithmen (KI) nicht genutzt werden kann, da deren Stärke besonders in der Verwertung großer Datenmengen liegt. Die Erhebung von Daten und ihre Auswertung ist jedoch nur eine dieser Herausforderungen, die es zu meistern gilt. Es ist mindestens genauso wichtig bereits vorhandene Daten sinnvoll miteinander zu vernetzen, was im Idealfall bedeuten würde, dass alle verfügbaren Daten aus den unterschiedlichsten Quellen zusammengeführt und somit als Ganzes ausgewertet werden können [85].

Mittlerweile gibt es immer größere Ansammlungen von Genomdaten, auf die Wissenschaftler unter anderem zurückgreifen können. Im britischen “100.000 Genomes Project“ [155] sollen, so wie der Name bereits sagt, 100.000 Genome entschlüsselt werden. Die Niederlande haben das “GoNL-Genome of the Netherlands“ [156], Saudi-Arabien das “Saudi Human Genom Programme“ [157] und in den USA sollen mit dem im Jahr 2015 gestarteten Projekt “Precision Medicine Initiative“ [158] genetische und medizinische Daten von 1.000.000 Personen gesammelt werden. Diese Genomsequenzierungen bilden einen Datenpool, dessen Nutzung neue wissenschaftliche Entdeckungen und medizinische Erkenntnisse ermöglichen kann [65].

3.6.1 “Big Data“ in der biomedizinischen Forschung

“Big Data“ ist ein immer häufiger verwendeter Begriff, da in vielen Bereichen immer größere und komplexere Datenmengen anfallen. Allein auf der Online Streaming-Plattform “Youtube“ werden jede Minute mehr als 300 Stunden Film hochgeladen, in einem Jahr müssen bis zu einem Exabyte an generierten Daten gespeichert werden, was einer Trillion Bytes entspricht. Aber auch in der Wissenschaft liefern immer größer angelegte Experimente immer mehr Daten. Der Weg von den Daten und Informationen hin zu Erkenntnissen und Wissen benötigt jedoch auch Speicher- und Analysemöglichkeiten in einer neuen Qualität [98].

Die Genforschung könnte sich bald an die Spitze der weltweiten “Big-Data“-Produzenten setzten, wenn die durch DNA-Sequenzierungen von Pflanzen, Tieren und Menschen anfallenden Datenmengen, so wie bereits prognostiziert, weiterhin rapide ansteigen. Schon jetzt verdoppelt sich die Menge der generierten genetischen Daten alle sieben Monate und dieser Trend wird sich fortsetzen [85]. Bereits im Jahr 2025 könnten demnach in der Genomik jedes Jahr bis zu 40 Exabyte (1 Exabyte = 1 Millionen Gigabyte) an Daten generiert werden, mehr als Zwanzigmal so viel wie in den Bereichen “Youtube“, “Twitter“ und der Astronomie zusammen [98]. Ein Teilgebiet der biomedizinischen Forschung, in dem immer größere Datenmengen anfallen, ist zum Beispiel die Krebsforschung. Durch die Sequenzierung des gesamten Erbguts von Krebszellen durch NGS-Technologien oder durch neue bildgebende Verfahren können teilweise riesige Datensätze anfallen. Deren Analyse ist jedoch auch mit großen informatischen Herausforderungen verbunden.

Die Entschlüsselung der jeweiligen DNA-Sequenzen nur nämlich nur ein erster Schritt. Generiert werden zunächst reine Basen-Abfolgen und Nukleotid-Abweichungen vom Referenz-Genom. Die jeweilige Sequenzier-Plattform liefert also erst einmal nur die Rohdaten und das in riesigen Mengen. Die Sequenzdaten müssen jedoch auch ausgewertet werden und dazu benötigt man komplexe und computertechnisch anspruchsvolle Algorithmen. Erst so lassen sich die Gen-Daten zielgerichtet analysieren, biologisch relevante Muster erkennen und Vergleiche anstellen, die bei der Diagnostik von Krankheiten weiterhelfen können [66].

Die Nutzung von “Big Data“ erlaubt eine viel genauere Prognose des Krankheitsverlaufs, als es momentan noch in der Klinik möglich ist. Damit lässt sich eine Behandlungsentscheidung treffen, die viel personalisierter ist und mehr den individuellen Patientenmerkmalen entspricht [159]. Die rasant anwachsende Datenmenge stellt die Wissenschaft aber auch vor Probleme, da es nicht immer möglich ist, auf standardisierte und anwendbare Daten zugreifen und die Dateigröße vieler Datenbanken eine enorm hohe Speicherkapazität erfordert. Große Datenmengen sind oft auch noch nicht zusammengeführt und können so nicht sinnvoll genutzt werden. Gerade in dieser Integration von Daten liegt ein riesiges Potenzial, welches es ermöglich kann, neue medizinisch relevante Zusammenhänge zu entdecken.

Die Bioinformatik bildet im Bereich des NGS zu aktuellem Stand immer noch den größten Engpass und es wird empfohlen, dass für jeden Geldbetrag, der für Sequenzier-Hardware ausgegeben wird, eine gleichgroße Investition im Bereich “Data-Science“ zu tätigen [160], da man mittlerweile an einem Punkt angekommen ist, wo das Auswerten der generierten Daten mehr kostet, als der Sequenzier-Prozess selbst [85]. Ebenso ist ein Fortschritt in zukünftigen Bereichen der bioinformatischen Infrastruktur notwendig, um in eine Verbesserung neuartiger Technologien wie einer effizienteren Daten-Komprimierung und der Entwicklung neuer bioinformatischer Methoden gewährleisten zu können [98].

4. Literaturverzeichnis

1. Merskey, H., *Pain terms: a list with definitions and notes on usage. Recommended by the IASP Subcommittee on Taxonomy.* Pain, 1979. **6**(3): p. 249.

2. Bonica, J.J., *The management of pain; with special emphasis on the use of analgesic block in diagnosis, prognosis, and therapy*. 1953, Philadelphia,: Lea & Febiger. 1533 p.

3. Moriarty, O., B.E. McGuire, and D.P. Finn, *The effect of pain on cognitive function: a review of clinical and preclinical research.* Prog Neurobiol, 2011. **93**(3): p. 385-404.

4. Merskey, H. and N.E. Bogduk, *Classification of chronic pain.* Seattle: IASP Press, 1994.

5. van Hecke, O., N. Torrance, and B.H. Smith, *Chronic pain epidemiology and its clinical relevance.* Br J Anaesth, 2013. **111**(1): p. 13-8.

6. Breivik, H., et al., *Survey of chronic pain in Europe: prevalence, impact on daily life, and treatment.* Eur J Pain, 2006. **10**(4): p. 287-333.

7. Moore, R.A., et al., *Amitriptyline for neuropathic pain and fibromyalgia in adults.* Cochrane Database Syst Rev, 2012. **12**: p. CD008242.

8. Derry, S., et al., *Milnacipran for neuropathic pain and fibromyalgia in adults.* Cochrane Database Syst Rev, 2012. **3**: p. CD008244.

9. Cazacu, I., C. Mogosan, and F. Loghin, *Safety issues of current analgesics: an update.* Clujul Med, 2015. **88**(2): p. 128-36.

10. Julius, D. and A.I. Basbaum, *Molecular mechanisms of nociception.* Nature, 2001. **413**(6852): p. 203-10.

11. Kringel, D. and J. Lotsch, *Pain research funding by the European Union Seventh Framework Programme.* Eur J Pain, 2015. **19**(5): p. 595-600.

12. Lötsch, J., G. Geisslinger, and I. Tegeder, *Genetic modulation of the pharmacological treatment of pain.* Pharmacol Ther, 2009. **124**(2): p. 168-84.

13. Ultsch, A., et al., *A data science approach to candidate gene selection of pain regarded as a process of learning and neural plasticity.* Pain, 2016. **157**(12): p. 2747-2757.

14. Przewlocki, R. and B. Przewlocka, *Opioids in chronic pain.* Eur J Pharmacol, 2001. **429**(1-3): p. 79-91.

15. Vadivelu, N., S. Mitra, and R.L. Hines, *Peripheral opioid receptor agonists for analgesia: a comprehensive review.* J Opioid Manag, 2011. **7**(1): p. 55-68.

16. Benarroch, E.E., *Endogenous opioid systems: current concepts and clinical correlations.* Neurology, 2012. **79**(8): p. 807-14.

17. Kapitzke, D., I. Vetter, and P.J. Cabot, *Endogenous opioid analgesia in peripheral tissues and the clinical implications for pain control.* Ther Clin Risk Manag, 2005. **1**(4): p. 279-97.

18. Al-Hasani, R. and M.R. Bruchas, *Molecular mechanisms of opioid receptor-dependent signaling and behavior.* Anesthesiology, 2011. **115**(6): p. 1363-81.

19. Inturrisi, C.E., *Clinical pharmacology of opioids for pain.* Clin J Pain, 2002. **18**(4 Suppl): p. S3-13.

20. Kalso, E., et al., *Opioids in chronic non-cancer pain: systematic review of efficacy and safety.* Pain, 2004. **112**(3): p. 372-80.

21. White, J.M. and R.J. Irvine, *Mechanisms of fatal opioid overdose.* Addiction, 1999. **94**(7): p. 961-72.

22. Tompkins, D.A. and C.M. Campbell, *Opioid-induced hyperalgesia: clinically relevant or extraneous research phenomenon?* Curr Pain Headache Rep, 2011. **15**(2): p. 129-36.

23. Lee, M., et al., *A comprehensive review of opioid-induced hyperalgesia.* Pain Physician, 2011. **14**(2): p. 145-61.

24. Mao, J., D.D. Price, and D.J. Mayer, *Thermal hyperalgesia in association with the development of morphine tolerance in rats: roles of excitatory amino acid receptors and protein kinase C.* J Neurosci, 1994. **14**(4): p. 2301-12.

25. Mert, T., et al., *Magnesium modifies fentanyl-induced local antinociception and hyperalgesia.* Naunyn Schmiedebergs Arch Pharmacol, 2009. **380**(5): p. 415-20.

26. Story, G.M., et al., *ANKTM1, a TRP-like channel expressed in nociceptive neurons, is activated by cold temperatures.* Cell, 2003. **112**(6): p. 819-29.

27. Peier, A.M., et al., *A TRP channel that senses cold stimuli and menthol.* Cell, 2002. **108**(5): p. 705-15.

28. Caterina, M.J., et al., *The capsaicin receptor: a heat-activated ion channel in the pain pathway.* Nature, 1997. **389**(6653): p. 816-24.

29. Weyer-Menkhoff, I. and J. Lotsch, *Human pharmacological approaches to TRP-ion-channel-based analgesic drug development.* Drug Discov Today, 2018. **23**(12): p. 2003-2012.

30. Kringel, D., et al., *Next-generation sequencing of the human TRPV1 gene and the regulating co-players LTB4R and LTB4R2 based on a custom AmpliSeq panel.* PLoS One, 2017. **12**(6): p. e0180116.

31. Laing, R.J. and A. Dhaka, *ThermoTRPs and Pain.* Neuroscientist, 2016. **22**(2): p. 171-87.

32. Mifflin, K.A. and B.J. Kerr, *The transition from acute to chronic pain: understanding how different biological systems interact.* Can J Anaesth, 2014. **61**(2): p. 112-22.

33. Ready, L.B., *The Interface Between Acute and Chronic Pain Management.* The Interface Between Acute and Chronic Pain Management, 1998.

34. Grichnik, K.P. and F.M. Ferrante, *The difference between acute and chronic pain.* Mt Sinai J Med, 1991. **58**(3): p. 217-20.

35. Voscopoulos, C. and M. Lema, *When does acute pain become chronic?* Br J Anaesth, 2010. **105 Suppl 1**: p. i69-85.

36. Sherman, R.A., C.J. Sherman, and L. Parker, *Chronic phantom and stump pain among American veterans: results of a survey.* Pain, 1984. **18**(1): p. 83-95.

37. Latremoliere, A. and C.J. Woolf, *Central sensitization: a generator of pain hypersensitivity by central neural plasticity.* J Pain, 2009. **10**(9): p. 895-926.

38. Bridges, D., S.W. Thompson, and A.S. Rice, *Mechanisms of neuropathic pain.* Br J Anaesth, 2001. **87**(1): p. 12-26.

39. Samad, T.A., et al., *Interleukin-1beta-mediated induction of Cox-2 in the CNS contributes to inflammatory pain hypersensitivity.* Nature, 2001. **410**(6827): p. 471-5.

40. Price, D.D., et al., *The N-methyl-D-aspartate receptor antagonist dextromethorphan selectively reduces temporal summation of second pain in man.* Pain, 1994. **59**(2): p. 165-74.

41. Watkins, L.R., E.D. Milligan, and S.F. Maier, *Spinal cord glia: new players in pain.* Pain, 2001. **93**(3): p. 201-5.

42. Ji, R.R. and M.R. Suter, *p38 MAPK, microglial signaling, and neuropathic pain.* Mol Pain, 2007. **3**: p. 33.

43. Zhuang, Z.Y., et al., *A peptide c-Jun N-terminal kinase (JNK) inhibitor blocks mechanical allodynia after spinal nerve ligation: respective roles of JNK activation in primary sensory neurons and spinal astrocytes for neuropathic pain development and maintenance.* J Neurosci, 2006. **26**(13): p. 3551-60.

44. Telleria-Diaz, A., et al., *Spinal antinociceptive effects of cyclooxygenase inhibition during inflammation: Involvement of prostaglandins and endocannabinoids.* Pain, 2010. **148**(1): p. 26-35.

45. Jhaveri, M.D., D. Richardson, and V. Chapman, *Endocannabinoid metabolism and uptake: novel targets for neuropathic and inflammatory pain.* Br J Pharmacol, 2007. **152**(5): p. 624-32.

46. Venkatachalam, K. and C. Montell, *TRP channels.* Annu Rev Biochem, 2007. **76**: p. 387-417.

47. Maixner, W., et al., *Sensitivity of patients with painful temporomandibular disorders to experimentally evoked pain: evidence for altered temporal summation of pain.* Pain, 1998. **76**(1-2): p. 71-81.

48. Yunus, M.B., *Fibromyalgia and overlapping disorders: the unifying concept of central sensitivity syndromes.* Seminars in arthritis and rheumatism, 2007. **36**(6): p. 339-56.

49. Flor, H., et al., *Cortical reorganization and phantom phenomena in congenital and traumatic upper-extremity amputees.* Exp Brain Res, 1998. **119**(2): p. 205-12.

50. Kawasaki, Y., et al., *Cytokine mechanisms of central sensitization: distinct and overlapping role of interleukin-1beta, interleukin-6, and tumor necrosis factor-alpha in regulating synaptic and neuronal activity in the superficial spinal cord.* J Neurosci, 2008. **28**(20): p. 5189-94.

51. Guo, W., et al., *Glial-cytokine-neuronal interactions underlying the mechanisms of persistent pain.* J Neurosci, 2007. **27**(22): p. 6006-18.

52. Ji, R.R., et al., *Neuroinflammation and Central Sensitization in Chronic and Widespread Pain.* Anesthesiology, 2018. **129**(2): p. 343-366.

53. Baron, R., A. Binder, and G. Wasner, *Neuropathic pain: diagnosis, pathophysiological mechanisms, and treatment.* Lancet Neurol, 2010. **9**(8): p. 807-19.

54. Baron, R., *Neuropathic pain: a clinical perspective.* Handbook of experimental pharmacology, 2009(194): p. 3-30.

55. van Hecke, O., et al., *Neuropathic pain in the general population: a systematic review of epidemiological studies.* Pain, 2014. **155**(4): p. 654-62.

56. Cruccu, G., A. Truini, and N. Neuropathic Pain Special Interest Group of the Italian Society of, *Neuropathic Pain: The Scope of the Problem.* Pain Ther, 2017. **6**(Suppl 1): p. 1-3.

57. Ji, R.R., T. Berta, and M. Nedergaard, *Glia and pain: is chronic pain a gliopathy?* Pain, 2013. **154 Suppl 1**: p. S10-28.

58. Jha, M.K., S. Jeon, and K. Suk, *Glia as a Link between Neuroinflammation and Neuropathic Pain.* Immune Netw, 2012. **12**(2): p. 41-7.

59. Hains, B.C. and S.G. Waxman, *Activated microglia contribute to the maintenance of chronic pain after spinal cord injury.* J Neurosci, 2006. **26**(16): p. 4308-17.

60. Zhou, L.J. and X.G. Liu, *Glial Activation, A Common Mechanism Underlying Spinal Synaptic Plasticity?* Neurosci Bull, 2017. **33**(1): p. 121-123.

61. Roerink, M.E., et al., *Cytokine Inhibition in Patients With Chronic Fatigue Syndrome: A Randomized Trial.* Ann Intern Med, 2017. **166**(8): p. 557-564.

62. Ferrini, F., et al., *Morphine hyperalgesia gated through microglia-mediated disruption of neuronal Cl(-) homeostasis.* Nat Neurosci, 2013. **16**(2): p. 183-92.

63. Roeckel, L.A., et al., *Opioid-induced hyperalgesia: Cellular and molecular mechanisms.* Neuroscience, 2016. **338**: p. 160-182.

64. Hieter, P. and M. Boguski, *Functional genomics: it's all how you read it.* Science, 1997. **278**(5338): p. 601-2.

65. Warth, O., *Daten sammeln? Daten nutzen! – Vom Segen der Datenberge.* Medica Magazin, 2017.

66. Gemeinholzer, B.N., J. Tauch, A. & Goesmann, A. , *Big Data in der Biologie: kein Problem mehr!* Biologie in unserer Zeit, 2019. **49**(1): p. 58-67.

67. Lotsch, J. and D. Kringel, *Use of Computational Functional Genomics in Drug Discovery and Repurposing for Analgesic Indications.* Clin Pharmacol Ther, 2018. **103**(6): p. 975-978.

68. https://www.drugbank.ca.

69. Lotsch, J. and A. Ultsch, *Process Pharmacology: A Pharmacological Data Science Approach to Drug Development and Therapy.* CPT Pharmacometrics Syst Pharmacol, 2016. **5**(4): p. 192-200.

70. http://www.geneontology.org.

71. Lippmann, C., et al., *Computational functional genomics-based approaches in analgesic drug discovery and repurposing.* Pharmacogenomics, 2018. **19**(9): p. 783-797.

72. Lander, E.S., et al., *Initial sequencing and analysis of the human genome.* Nature, 2001. **409**(6822): p. 860-921.

73. Gonzalez-Garay, M.L., *The road from next-generation sequencing to personalized medicine.* Per Med, 2014. **11**(5): p. 523-544.

74. Yang, Y., et al., *Clinical whole-exome sequencing for the diagnosis of mendelian disorders.* N Engl J Med, 2013. **369**(16): p. 1502-11.

75. Gilissen, C., et al., *Genome sequencing identifies major causes of severe intellectual disability.* Nature, 2014. **511**(7509): p. 344-7.

76. Dilliott, A.A., et al., *Targeted Next-generation Sequencing and Bioinformatics Pipeline to Evaluate Genetic Determinants of Constitutional Disease.* J Vis Exp, 2018(134).

77. Woopen, C., *Die Zukunft der genetischen Diagnostik – von der Forschung in die klinische Anwendung.* Deutscher Ethikrat, 2013.

78. Arsenic, R., et al., *Comparison of targeted next-generation sequencing and Sanger sequencing for the detection of PIK3CA mutations in breast cancer.* BMC Clin Pathol, 2015. **15**: p. 20.

79. Di Resta, C., et al., *Next-generation sequencing approach for the diagnosis of human diseases: open challenges and new opportunities.* EJIFCC, 2018. **29**(1): p. 4-14.

80. D'Argenio, V., et al., *The molecular analysis of BRCA1 and BRCA2: Next-generation sequencing supersedes conventional approaches.* Clin Chim Acta, 2015. **446**: p. 221-5.

81. https://www.nobelprize.org.

82. Sanger, F., S. Nicklen, and A.R. Coulson, *DNA sequencing with chain-terminating inhibitors.* Proc Natl Acad Sci U S A, 1977. **74**(12): p. 5463-7.

83. Grumbt, B., et al., *Diagnostic applications of next generation sequencing in immunogenetics and molecular oncology.* Transfus Med Hemother, 2013. **40**(3): p. 196-206.

84. http://www.thermofisher.com.

85. Schmidt, B. and A. Hildebrandt, *Next-generation sequencing: big data meets high performance computing.* Drug Discov Today, 2017. **22**(4): p. 712-717.

86. Quail, M.A., et al., *A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers.* BMC Genomics, 2012. **13**: p. 341.

87. Lohmann, K. and C. Klein, *Next generation sequencing and the future of genetic diagnosis.* Neurotherapeutics, 2014. **11**(4): p. 699-707.

88. Kringel, D., et al., *Emergent biomarker derived from next-generation sequencing to identify pain patients requiring uncommonly high opioid doses.* Pharmacogenomics J, 2017. **17**(5): p. 419-426.

89. Kringel, D. and J. Lotsch, *Next-generation sequencing of human opioid receptor genes based on a custom AmpliSeq library and ion torrent personal genome machine.* Clin Chim Acta, 2016. **463**: p. 32-38.

90. Berglund, E.C., A. Kiialainen, and A.C. Syvanen, *Next-generation sequencing technologies and applications for human genetic history and forensics.* Investig Genet, 2011. **2**: p. 23.

91. Gogol-Doring, A. and W. Chen, *An overview of the analysis of next generation sequencing data.* Methods Mol Biol, 2012. **802**: p. 249-57.

92. Fox, E.J., et al., *Accuracy of Next Generation Sequencing Platforms.* Next Gener Seq Appl, 2014.

93. Erichsen, H.C. and S.J. Chanock, *SNPs in cancer research and treatment.* Br J Cancer, 2004. **90**(4): p. 747-51.

94. Venter, J.C., et al., *The sequence of the human genome.* Science, 2001. **291**(5507): p. 1304-51.

95. Crawford, D.C. and D.A. Nickerson, *Definition and clinical importance of haplotypes.* Annu Rev Med, 2005. **56**: p. 303-20.

96. https://www.genome.gov/human-genome-project.

97. http://www.genome.gov/sequencingcosts.

98. Stephens, Z.D., et al., *Big Data: Astronomical or Genomical?* PLoS Biol, 2015. **13**(7): p. e1002195.

99. McDermott, U., *Next-generation sequencing and empowering personalised cancer medicine.* Drug Discov Today, 2015. **20**(12): p. 1470-5.

100. Lotsch, J. and A. Ultsch, *Machine learning in pain research.* Pain, 2018. **159**(4): p. 623-630.

101. Kuhlmann, P., *Künstliche Intelligenz.* 2018.

102. https://www.nvidia.com.

103. Hafen, T., *Schlaue Maschienen bald überall.* Com! professional, 2017.

104. Herbrich, R., *Schlaue Maschienen bald überall.* com! professional, 2017.

105. Nguyen, C.N.Z., O., *Machine Learning.* O'Reillys, 2018.

106. Soo, K.N., A., *Data Science.* Springer, 2019.

107. Lötsch, J., G. Geisslinger, and C. Walter, *Wissensgenerierung aus komplexen Datensätzen in der humanexperimentellen Schmerzforschung.* Der Schmerz, 2019.

108. Kringel, D., et al., *Machine-learned analysis of the association of next-generation sequencing-based genotypes with persistent pain after breast cancer surgery.* Pain, 2019. **160**(10): p. 2263-2277.

109. Kaunisto, M.A., et al., *Pain in 1,000 women treated for breast cancer: a prospective study of pain sensitivity and postoperative pain.* Anesthesiology, 2013. **119**(6): p. 1410-21.

110. Lötsch, J. and G. Geisslinger, *Relevance of frequent mu-opioid receptor polymorphisms for opioid activity in healthy volunteers.* Pharmacogenomics J, 2006.

111. von Hehn, C.A., R. Baron, and C.J. Woolf, *Deconstructing the neuropathic pain phenotype to reveal neural mechanisms.* Neuron, 2012. **73**(4): p. 638-52.

112. Lotsch, J., et al., *Quantitative sensory testing response patterns to capsaicin- and ultraviolet-B-induced local skin hypersensitization in healthy subjects: a machine-learned analysis.* Pain, 2018. **159**(1): p. 11-24.

113. Albin, K.C., M.I. Carstens, and E. Carstens, *Modulation of oral heat and cold pain by irritant chemicals.* Chem Senses, 2008. **33**(1): p. 3-15.

114. Bandell, M., et al., *Noxious cold ion channel TRPA1 is activated by pungent compounds and bradykinin.* Neuron, 2004. **41**(6): p. 849-57.

115. Kringel, D., et al., *Machine-learned analysis of the association of next-generation sequencing-based human TRPV1 and TRPA1 genotypes with the sensitivity to heat stimuli and topically applied capsaicin.* Pain, 2018. **159**(7): p. 1366-1381.

116. http://www.sanger.ac.uk/sanger/StatGen_Gwava.

117. Kringel, D., et al., *A machine-learned analysis of human gene polymorphisms modulating persisting pain points to major roles of neuroimmune processes.* Eur J Pain, 2018. **22**(10): p. 1735-1756.

118. Kringel, D., et al., *Development of an AmpliSeq(TM) Panel for Next-Generation Sequencing of a Set of Genetic Predictors of Persisting Pain.* Front Pharmacol, 2018. **9**: p. 1008.

119. Mansour, A.R., et al., *Chronic pain: the role of learning and brain plasticity.* Restor Neurol Neurosci, 2014. **32**(1): p. 129-39.

120. Alvarado, S., et al., *Peripheral nerve injury is accompanied by chronic transcriptome-wide changes in the mouse prefrontal cortex.* Mol Pain, 2013. **9**: p. 21.

121. Lotsch, J., G. Geisslinger, and I. Tegeder, *Genetic modulation of the pharmacological treatment of pain.* Pharmacol Ther, 2009. **124**(2): p. 168-84.

122. Mogil, J.S., *Pain genetics: past, present and future.* Trends Genet, 2012. **28**(6): p. 258-66.

123. Zorina-Lichtenwalter, K., et al., *Genetic predictors of human chronic pain conditions.* Neuroscience, 2016. **338**: p. 36-62.

124. Ultsch, A. and J. Lotsch, *Computed ABC Analysis for Rational Selection of Most Informative Variables in Multivariate Data.* PLoS One, 2015. **10**(6): p. e0129767.

125. Ultsch, A., et al., *Identification of Molecular Fingerprints in Human Heat Pain Thresholds by Use of an Interactive Mixture Model R Toolbox (AdaptGauss).* Int J Mol Sci, 2015. **16**(10): p. 25897-911.

126. Lötsch, J.U., A., *Random forests followed by computed ABC analysis as a feature selection method for machine-learning in biomedical data.* Conference of the International Federation of Classification Societies, 2018.

127. Ultsch, A., *Visualisation and Classification with Artificial Life.* 2000.

128. Cover, T.M. and P.E. Hart, *Nearest Neighbor Pattern Classification.* Ieee Transactions on Information Theory, 1967. **13**(1): p. 21-+.

129. Walker, S.H. and D.B. Duncan, *Estimation of the probability of an event as a function of several independent variables.* Biometrika, 1967. **54**(1): p. 167-79.

130. Breiman, L., *Random forests.* Machine Learning, 2001. **45**(1): p. 5-32.

131. Schapire, R.E., *A brief introduction to boosting.* Ijcai-99: Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence, Vols 1 & 2, 1999: p. 1401-1406.

132. Backes, C., et al., *GeneTrail--advanced gene set enrichment analysis.* Nucleic Acids Res, 2007. **35**(Web Server issue): p. W186-92.

133. Ultsch, A. and J. Lotsch, *Machine-learned cluster identification in high-dimensional data.* J Biomed Inform, 2017. **66**: p. 95-104.

134. Lötsch, J., et al., *Identification of disease-distinct complex biomarker patterns by means of unsupervised machine-learning using an interactive R toolbox (Umatrix).* Big Data Analytics, 2018. **3**(1): p. 5.

135. Chung, J.M., *The role of reactive oxygen species (ROS) in persistent pain.* Mol Interv, 2004. **4**(5): p. 248-50.

136. Tian, L., et al., *Neuroimmune crosstalk in the central nervous system and its significance for neurological diseases.* J Neuroinflammation, 2012. **9**: p. 155.

137. Lin, Q., et al., *Nitric oxide mediates the central sensitization of primate spinothalamic tract neurons.* J Neurophysiol, 1999. **81**(3): p. 1075-85.

138. Metzker, M.L., *Sequencing technologies - the next generation.* Nat Rev Genet, 2010. **11**(1): p. 31-46.

139. Martikainen, I.K., et al., *Chronic Back Pain Is Associated with Alterations in Dopamine Neurotransmission in the Ventral Striatum.* J Neurosci, 2015. **35**(27): p. 9957-65.

140. Gibbs, J.E. and D.W. Ray, *The role of the circadian clock in rheumatoid arthritis.* Arthritis Res Ther, 2013. **15**(1): p. 205.

141. Gilron, I. and N. Ghasemlou, *Chronobiology of chronic pain: focus on diurnal rhythmicity of neuropathic pain.* Curr Opin Support Palliat Care, 2014. **8**(4): p. 429-36.

142. Coyle, J.T. and G. Tsai, *NMDA receptor function, neuroplasticity, and the pathophysiology of schizophrenia.* Int Rev Neurobiol, 2004. **59**: p. 491-515.

143. Nemeth, E., et al., *IL-6 mediates hypoferremia of inflammation by inducing the synthesis of the iron regulatory hormone hepcidin.* J Clin Invest, 2004. **113**(9): p. 1271-6.

144. Mocellin, S., et al., *The multifaceted relationship between IL-10 and adaptive immunity: putting together the pieces of a puzzle.* Cytokine Growth Factor Rev, 2004. **15**(1): p. 61-76.

145. Deakin, A.M., et al., *The modulation of IL-6 and TNF-alpha release by nitric oxide following stimulation of J774 cells with LPS and IFN-gamma.* Cytokine, 1995. **7**(5): p. 408-16.

146. Clapauch, R., et al., *Endothelial function and insulin resistance in early postmenopausal women with cardiovascular risk factors: importance of ESR1 and NOS3 polymorphisms.* PLoS One, 2014. **9**(7): p. e103444.

147. Zhang, L., et al., *Discovery of common human genetic variants of GTP cyclohydrolase 1 (GCH1) governing nitric oxide, autonomic activity, and cardiovascular risk.* J Clin Invest, 2007. **117**(9): p. 2658-71.

148. Clapham, D.E., *TRP channels as cellular sensors.* Nature, 2003. **426**(6966): p. 517-24.

149. Lötsch, J. and G. Geisslinger, *Are mu-opioid receptor polymorphisms important for clinical opioid therapy?* Trends Mol Med, 2005. **11**(2): p. 82-9.

150. Fillingim, R.B., et al., *Genetic contributions to pain: a review of findings in humans.* Oral Dis, 2008. **14**(8): p. 673-82.

151. Lotsch, J., et al., *Machine-learning-derived classifier predicts absence of persistent pain after breast cancer surgery with high accuracy.* Breast Cancer Res Treat, 2018. **171**(2): p. 399-411.

152. Knisely, M.R., et al., *Associations Between Catecholaminergic and Serotonergic Genes and Persistent Breast Pain Phenotypes After Breast Cancer Surgery.* J Pain, 2018. **19**(10): p. 1130-1146.

153. Meacham, K., et al., *Neuropathic Pain: Central vs. Peripheral Mechanisms.* Curr Pain Headache Rep, 2017. **21**(6): p. 28.

154. Becker, S. and P. Schweinhardt, *Dysfunctional neurotransmitter systems in fibromyalgia, their role in central stress circuitry and pharmacological actions on these systems.* Pain Res Treat, 2012. **2012**: p. 741746.

155. https://www.genomicsengland.co.uk.

156. http://www.nlgenome.nl.

157. https://www.saudigenomeprogram.org/en.

158. https://obamawhitehouse.archives.gov/precision-medicine.

159. Doostparast Torshizi, A. and K. Wang, *Next-generation sequencing in drug development: target identification and genetically stratified clinical trials.* Drug Discov Today, 2018. **23**(10): p. 1776-1783.

160. Gullapalli, R.R., et al., *Next generation sequencing in clinical medicine: Challenges and lessons for pathology and biomedical informatics.* J Pathol Inform, 2012. **3**: p. 40.

5. Anhang

5.1 Abkürzungen

°C	Grad Celsius
DNA	Desoxyribonukleinsäure
FP7	Forschungsrahmenprogramm 7
GLORIA	Glial Opioid Receptor Interface in Analgesia
GO	Gene Ontology
GWAVA	Genome-wide Annotation of Variants
hg19	Humanes Referenzgenom 19
KI	Künstliche Intelligenz
mRNA	Boten-Ribonukleinsäure
NGS	Next Generation Sequencing
NMDA	N-Methyl-D-Aspartat
PCR	Polymerase-Kettenreaktion
SNP	Einzelnukleotid Polymorphismus
SNV	Einzelnukleotid Variation
TRP	Transient Receptor Potential

5.2 Eidesstattliche Erklärung

Sehr geehrte Damen und Herren,

hiermit versichere ich, dass ich die vorliegende Dissertation selbstständig verfasst und keine anderen als die angegebenen Hilfsmittel genutzt habe. Alle wörtlichen oder inhaltlichen Entlehnungen aus anderen benutzten Druckwerken oder digitalisierten Publikationen und Quellen habe ich als solche gekennzeichnet.

Ich versichere außerdem, dass ich die beigefügte Dissertation nur in diesem und keinem anderen Promotionsverfahren eingereicht habe. Diesem Promotionsverfahren sind keine gescheiterten Promotionsverfahren vorausgegangen.

Frankfurt am Main / 1.12.2019

Ort/ Datum

Dario Kringel

5.3 Danksagung

Besonderer Dank gilt an erster Stelle Herr Professor Dr. Dr. Jörn Lötsch für die Vergabe des interessanten Promotionsthemas und vor allem für eine Betreuung, die besser nicht hätte sein können. Ich danke ihm für die vielen hilfreichen und anregenden Gespräche, die im Laufe meiner Promotionszeit zu vielen neuen und interessanten Blickwinkeln auf verschiedenste Themen in den Bereichen der molekularen Biologie, Pharmakologie und "Data-Science" geführt haben. Herr Professor Dr. Dr. Gerd Geißlinger, Direktor des Instituts für klinische Pharmakologie und der Fraunhofer IME Projektgruppe danke ich für die guten Rahmenbedingungen zur Durchführung meiner Promotion im Institut für klinische Pharmakologie und dem Industriepark Höchst. Ich danke des weiteren Herr Professor Dr. Helge Bode, Prodekan des Instituts für Biowissenschaften, für die Begutachtung dieser Arbeit. Ein ganz besonderer Dank gilt allen Mitgliedern des Forschungs-Programms GLORIA, mit denen ich in den letzten fünf Jahren einen sehr regen, interessanten und auch sehr herzlichen Austausch hatte. Ebenso bedanke ich mich bei meinen Kollegen Frau Dr. Carmen Walter, Frau Weyer-Menkhoff, Herr Onno Hansen-Goos und Herr Dr. Sebastian Malkusch ganz herzlich für die nette Zusammenarbeit sowie die gute Arbeitsatmosphäre. Herr Prof. Dr. Mike Parnham, Herr Dr. Eduard Resch und Frau Sonja Luckhardt der Fraunhofer IME Projektgruppe danke ich für die Zusammenarbeit und die Unterstützung bei fachlichen Fragen sowie den anregenden Diskussionen. Danken möchte ich außerdem allen Kollegen des Instituts für klinische Pharmakologie für die gute und freundschaftliche Zusammenarbeit, insbesondere meinem ehemaligen Kollegen und guten Freund Herr Dr. Patrick Slattery. Besonderer Dank gilt auch Frau Viviane Menges für ihre große Hilfe bei der grammatikalischen Korrektur dieser Arbeit. Mein größter Dank gilt meinen Eltern, meiner Familie und Freunden für den uneingeschränkten Glauben an mich, den mentalen Rückhalt und Zuspruch und für den starken Zusammenhalt und die Unterstützung während meiner Promotionszeit und darüber hinaus.

5.4 Curriculum Vitae

Name:	Kringel
Vorname:	Dario
Geburtsdatum:	16 Januar 1985
Geburtsort:	Frankfurt am Main
Anschrift:	Lucaestraße 15, 60433 Frankfurt

Schulbildung:	2001 – 2004	Jacob Grimm Gymnasium (Kassel) Abschluss: Abitur
Zivildienst:	2004 – 2005	Das blaue Haus (Kassel)
Studium:	2005 – 2012	Biowissenschaften an der Goethe Universität (Frankfurt am Main) Abschluss: Diplom
ERASMUS Stipendium:	2009 – 2010	Molecular Life Sciences an der Queen Mary University (London)
Berufliche Tätigkeit	2013 – 2014	Pharmaberater im Außendienst bei der Roboklon GmbH (Frankfurt am Main)
Promotion	ab 2014	Institut für Klinische Pharmakologie an der Goethe Universität (Frankfurt am Main) Anleitung durch Prof. Dr. Dr. Jörn Lötsch und Prof. Dr. Helge Bode im Rahmen des EU Forschungs-Projektes GLORIA

Erklärung zu den Autorenanteilen an der Publikation:

Next-generation sequencing of human opioid receptor genes based on a custom AmpliSeq™ library and ion torrent personal genome machine (*printed*)

Name der Zeitschrift: Clinica Chimica Acta

Beteiligte Autoren: D Kringel und J Lötsch

Was hat der Promovierende bzw. was haben die Koautoren beigetragen?

(1) zu Entwicklung und Planung

Promovierender: 50%
Autor JL: 50%

(2) zur Durchführung der einzelnen Untersuchungen und Experimente

Promovierender: 100% (DNA Extraktion, NGS-Panel Design, Sequenzierung, Varianten Annotation, Methoden Validierung)

(3) zur Erstellung der Datensammlung und Abbildungen

Promovierender: 70% (Datenerhebung, Run-Statistik, Erstellung der Abbildung Chip-Beladung und Abbildung Validierung)
Autor JL: 30% (Erstellung der Abbildung Genetische Vektoren)

(4) zur Analyse und Interpretation der Daten

Promovierender: 60% (Datenvorverarbeitung, Dateninterpretation)
Autor JL: 40% (Datenanalyse und Dateninterpretation)

(5) zum Verfassen des Manuskripts

Promovierender: 50%
Autor JL: 50%

Zustimmende Bestätigungen der oben genannten Angaben:

_____ _____

Datum/Ort Unterschrift Promovend

_____ _____

Datum/Ort Unterschrift Betreuer

Clinica Chimica Acta 463 (2016) 32–38

Contents lists available at ScienceDirect

Clinica Chimica Acta

journal homepage: www.elsevier.com/locate/clinchim

Next-generation sequencing of human opioid receptor genes based on a custom AmpliSeq™ library and ion torrent personal genome machine

Dario Kringel [a], Jörn Lötsch [a,b,*]

[a] Institute of Clinical Pharmacology, Goethe - University, Theodor-Stern-Kai 7, 60590 Frankfurt am Main, Germany
[b] Fraunhofer Institute for Molecular Biology and Applied Ecology IME, Project Group Translational Medicine and Pharmacology TMP, Theodor-Stern-Kai 7, 60590 Frankfurt am Main, Germany

ARTICLE INFO

Article history:
Received 22 April 2016
Received in revised form 12 September 2016
Accepted 7 October 2016
Available online 8 October 2016

Keywords:
Opioid system
Analgesia
Addiction
Molecular diagnosis
Mutations
NGS sequencing

ABSTRACT

Background: The opioid system is involved in the control of pain, reward, addictive behaviors and vegetative effects. Opioids exert their pharmacological actions through the agonistic binding at opioid receptors and variation in the coding genes has been found to modulate opioid receptor expression or signaling. However, a limited selection of functional opioid receptor variants is perceived as insufficient in providing a genetic diagnosis of clinical phenotypes and therefore, unrestricted access to opioid receptor genetics is required.
Methods: Next-generation sequencing (NGS) workflow was based on a custom AmpliSeq™ panel and designed for sequencing of human genes related to the opioid receptor group (*OPRM1*, *OPRD1*, *OPRK1*, *SIGMA1*, *OPRL1*) on an Ion PGM™ Sequencer. A cohort of 79 previously studied chronic pain patients was screened to evaluate and validate the detection of exomic sequences of the coding genes with 25 base pair exon padding. *In-silico* analysis was performed using SNP and Variation Suite® software.
Results: The amplicons covered approximately 90% of the target sequence. A median of 2.54×10^6 reads per run was obtained generating a total of 35,447 nucleotide reads from each DNA sample. This identified approximately 100 chromosome loci where nucleotides deviated from the reference sequence GRCh37 hg19, including functional variants such as the *OPRM1* rs1799971 SNP (118 A > G) as the most scientifically regarded variant or rs563649 SNP coding for μ-opioid receptor splice variants. Correspondence between NGS and Sanger derived nucleotide sequences was 100%.
Conclusion: Results suggested that the NGS approach based on AmpliSeq™ libraries and Ion PGM sequencing is a highly efficient mutation detection method. It is suitable for large-scale sequencing of opioid receptor genes. The method includes the variants studied so far for functional associations and adds a large amount of genetic information as a basis for complete analysis of human opioid receptor genetics and its functional consequences.

1. Introduction

Opioid signaling is triggered by endogenous opioid peptides comprising endorphins, enkephalins and dynorphins with preferential selectivities at the μ-, κ- and δ-opioid receptors, respectively [1]. The existence of opioid binding sites in the brain was established in 1973 [2] and subsequently located in various different physiological systems. In peripheral [3,4] and central [5–7] parts of the nociceptive system, activation confers an effective nocifensive mechanism. Activation at the *Area postrema* causes respiratory depression [8], activation in brain areas belonging to the reward system [9] is involved in addiction [10–12], activation in the gastrointestinal tract causes constipation [13], and activation at the hypothalamic–pituitary–adrenal (HPA) axis is involved in stress responses [14].

Genetic variation of human opioid receptors is an active research topic that has identified several possible modulators of the individual response to endogenous and exogenous opioids [15]. Variants were associated with the modulation of opioid receptor expression [16,17] or signaling [17–19] up to an almost complete functional loss [18,20]. However, the collection of certain opioid receptor variants has provided only modest explanations of clinical phenotypes [21,22]. Therefore, a limited selection of functional opioid receptor variants for which specific genetic assays had been established [23] is perceived as insufficient in providing a genetic diagnosis of the clinical phenotype [24].

Since the early 70th of the last century when the first processes of Sanger sequencing were introduced [25,26], the access to the whole genomic information has been widely accepted as a valuable method in clinical research [27]. With the recent broader availability of next generation sequencing (NGS) [28] the limitation to known functional variants has therefore fallen in favor of unrestricted access to opioid receptor genetics, which has already been shown to provide a working genetic marker for opioid-related clinical phenotypes [29]. In this report, the evaluation of a new NGS method based on a custom AmpliSeq™ library and Ion Torrent sequencing for the fast detection of genetic variations in human opioid receptor genes is described.

* Corresponding author.
E-mail address: j.loetsch@em.uni-frankfurt.de (J. Lötsch).

http://dx.doi.org/10.1016/j.cca.2016.10.009
0009-8981/

2. Methods

2.1. Gene selection

Exomic genotyping was performed for the human *OPRM1*, *OPRK1* and *OPRD1* genes (NCBI IDs 4985, 4988 and 4986), located on chromosomes 6, 8, and 1, and encoding for the three major (μ, κ, δ) opioid receptors. Their endogenous ligands are opioid neuropeptides comprising endorphins/endomorphines, dynorphins, enkephalins, respectively, and they are the main targets of opioid analgesics as listed in the DrugBank database (version 4.1; http://www.drugbank.ca [30]). In particular *OPRM1* is addressed by most exogenous opioids. In addition, the *SIGMAR1* gene (NCBI ID 10280) located on chromosome 9, was included that codes for σ_1-opioid receptors. These are not activated or blocked by opioid peptides or naloxone, respectively, however, historically belong to the opioid receptors [31], are targets of some drugs classified as opioids such as pentazocine [32] or dextromethorphan [33] and their activation modulates peripheral μ-opioid analgesia [34]. Furthermore, the gene panel was extended to the opiate receptor-like 1 gene (*OPRL1*; NCBI ID 4987), located on chromosome 20 and coding for the nociceptin receptor that shows a high degree of structural homology to the classical opioid receptors although opioid peptides or morphine-like compounds have little or no affinity for it [35]. Its endogenous ligand is nociceptin, however, exogenous opioids comprising buprenorphine [36] and its glucuronide metabolites [37] have been shown to exert agonist activity that seems to contribute to their clinical effects, and the nonselective opioid agonist etorphine also binds at the *OPRL1* gene product [38], which supports the inclusion of this gene in the present panel.

2.2. DNA template preparation and amplification

The investigation followed the Declaration of Helsinki on Biomedical Research Involving Human Subjects and was approved by the Ethics Committee of the Medical Faculty of the Goethe-University, Frankfurt, Germany. All subjects had provided informed written consent covering the genotyping. Genomic DNA was available from venous blood samples drawn from 79 pain patients in a previously reported [29] context of the pharmacogenomics of opioid dosing requirements.

DNA was extracted from 200 μl blood on a BioRobot EZ1 workstation applying the blood and body fluid spin protocol provided in the EZ1 DNA Blood 200 μl Kit (Qiagen, Hilden, Germany). A multiplex PCR amplification strategy for the coding DNA sequences was accomplished online (Ion Ampliseq™ Designer; http://www.ampliseq.com) to amplify the target region specified above (for primer sequences, see Supplementary Table 1) with 25 base pair exon padding. After a comparison of several primer design options, the design providing the maximum target sequence coverage was chosen. The ordered amplicons covered approximately 90% of the target sequence (Table 1). A total of 10 ng DNA per sample were used for the target enrichment by a multiplex PCR and each DNA pool was amplified with the Ion Ampliseq™ Library Kit in conjunction with the Ion Ampliseq™ "custom Primer Pool" - protocols according to the manufacturer procedures (Life Technologies, Darmstadt, Germany).

After each pool had undergone 17 PCR cycles, the PCR primers were removed with FuPa Reagent and the amplicons were ligated to the sequencing adaptors with short stretches of index sequences (barcodes) that enabled sample multiplexing for subsequent steps (Ion Xpress™ Barcode Adapters Kit; Life Technologies). After purification with AMPure XP beads (Beckman Coulter, Krefeld, Germany), the barcoded libraries were quantified with a Qubit® 2.0 Fluorimeter (Life Technologies, Darmstadt, Germany) and normalized for DNA concentration to a final concentration of 20 pmol/l using the Ion Library Equalizer™ Kit (Life Technologies, Darmstadt, Germany). Equalized barcoded libraries from 11 to 12 samples at a time were pooled. To clonally amplify the library DNA onto the Ion Sphere Particles (ISPs; Life Technologies, Darmstadt, Germany), the library pool was subjected to emulsion PCR by using an IT OneTouch template kit on an IT OneTouch system (Life Technologies, Darmstadt, Germany) following the manufacturer's protocol.

2.3. Sequencing

Enriched ISPs which carried many copies of the same DNA fragment were subjected to sequencing on an Ion 316 Chip to sequence pooled libraries with eleven to twelve samples. The 316 chip was chosen (instead of the low-capacity 314 or the high-capacity 318 chip) to obtain a mean sequencing depth of coverage of 50× which means that, on average, each base has been sequenced 50×, when eleven samples were loaded. A larger number of samples could be analyzed simultaneously using the 318 chip, but this would increase the turnaround time for each sample, depending on the number of samples that are received by the laboratory. Sequencing was performed using the sequencing kit (Ion PGM 200 Sequencing Kit; Life Technologies, Darmstadt, Germany) as per the manufacturer's instructions with the 200-bp single-end run configuration.

2.4. Bioinformatics generation of sequence information

The raw data (unmapped BAM-files) from the sequencing runs were processed using Torrent Suite Software (Version 4.4.2, Life Technologies, Darmstadt, Germany) to generate read alignments which are filtered by the software into mapped BAM-files using the reference genomic sequence (hg19) of target genes. Variant calling was performed with the Torrent Variant Caller Plugin using as key parameters: minimum allele frequency = 0.015, minimum quality = 10, minimum coverage = 20 and minimum coverage on either strand = 3.

The annotation of called variants was done using the Ion Reporter Software (Version 4.4; Life Technologies, Darmstadt, Germany) for the VCF files that contained the nucleotide reads. The GenomeBrowse® software (Version 2.0.4, Golden Helix, Bozeman, MT, USA) was used to map the observed sequences to the reference sequences GRCh37 hg19 (dated February 2009). The SNP and Variation Suite software (Version 8.4.4; Golden Helix, Bozeman, MT, USA) was used for the analysis of sequence quality, coverage and also for variant identification.

2.5. Method validation

For method validation, two genomic regions were chosen at random from the whole analyzed region for validation by Sanger sequencing [25,26] in an independent external laboratory (AGOWA, Berlin, Germany), which was performed in ten DNA samples randomly chosen from the n = 79 samples in the present cohort. Amplification of the respective DNA segments was done using PCR primer pairs (forward, reverse) of (i) 5´-ATGAAGACAGCAACCAACATTTAC-3´ and 5´-CCAGATGCAGATATTGATGATCTT-3´ and (ii) 5´-TTTAACCTG GCTTTGGCAGATG-3´ and 5´-ACATCGACGTCTTCCCTGACT-3´. The results of Sanger sequencing were aligned with the genomic sequence and analyzed using Chromas Lite® (Version 2.1.1, Technelysium Pty Ltd., South Brisbane, Australia) and the GenomeBrowse® (Version 2.0.4, Golden Helix, Bozeman, MT, USA) was used to compare the sequences obtained with NGS or Sanger techniques.

3. Results

The NGS assay of human opioid receptors was established on 79 genomic DNA samples [29]. As proposed previously [39], only exons and their boundary sequences for which read-depths >20 for each nucleotide could be obtained were considered as successfully analyzed. Applying this criterion, complete or nearly complete coverage of the relevant sequences was obtained (Table 1; for details on missing variants, see Supplementary Table 2).

Table 1
AmpliSeq™ amplicons and coverage details of the human opioid receptor NGS assay.

Gene	Chr	Chr start	Chr end	Amplicons	Total bases	Covered bases	Coverage	Sum (total, covered, %)
OPRD1	Chr1	29138628	29139147	1	519	94	0.181	1924, 1336, 69.4%
	Chr1	29185440	29185840	2	400	292	0.730	
	Chr1	29189228	29190233	6	1005	950	0.9453	
OPRM1	Chr6	154331605	154331719	1	114	114	1.000	23787, 21859, 91.8%
	Chr6	154331605	154332178	3	573	573	1.000	
	Chr6	154360211	154360465	2	254	254	1.000	
	Chr6	154360270	154360465	2	195	195	1.000	
	Chr6	154360520	154360994	4	474	474	1.000	
	Chr6	154360211	154360994	6	783	783	1.000	
	Chr6	154407616	154408945	8	1329	1329	1.000	
	Chr6	154407616	154408964	8	1348	1348	1.000	
	Chr6	154410935	154411338	3	403	403	1.000	
	Chr6	154410871	154411338	3	467	467	1.000	
	Chr6	154412061	154412632	4	571	571	1.000	
	Chr6	154412918	154413098	2	180	180	1.000	
	Chr6	154414379	154414680	2	301	301	1.000	
	Chr6	154428725	154428812	1	87	87	1.000	
	Chr6	154428574	154428812	2	238	238	1.000	
	Chr6	154429125	154430178	6	1053	865	0.8215	
	Chr6	154428857	154430178	9	1321	1123	0.8501	
	Chr6	154431464	154431613	1	149	149	1.000	
	Chr6	154439792	154453514	72	13722	12180	0.88763	
	Chr6	154567801	154568026	2	225	225	1.000	
OPRK1	Chr8	54138250	54142414	23	4164	3872	0.9299	5683, 5117, 90%
	Chr8	54147293	54147696	3	403	403	1.000	
	Chr8	54155273	54155497	1	224	116	0.518	
	Chr8	54163315	54163670	3	355	338	0.952	
	Chr8	54163982	54164219	1	237	194	0.819	
	Chr8	54163982	54164282	1	300	194	0.647	
SIGMAR1	Chr9	34634693	34635662	6	969	969	1.000	3520, 3427, 97.3%
	Chr9	34634693	34635880	8	1187	1187	1.000	
	Chr9	34636968	34637111	1	143	143	1.000	
	Chr9	34636968	34637134	1	166	166	1.000	
	Chr9	34637238	34637442	2	204	204	1.000	
	Chr9	34637191	34637442	2	251	251	1.000	
	Chr9	34637578	34637848	2	270	251	0.930	
	Chr9	34637518	34637848	2	330	256	0.776	
OPRL1	Chr20	62711445	62711730	2	285	230	0.807	3954, 3708, 93.7%
	Chr20	62716343	62716526	2	183	183	1.000	
	Chr20	62717832	62717977	2	145	145	1.000	
	Chr20	62723293	62723494	2	201	201	1.000	
	Chr20	62724015	62724331	4	316	316	1.000	
	Chr20	62729129	62729535	5	406	406	1.000	
	Chr20	62729603	62732021	15	2418	2227	0.921	

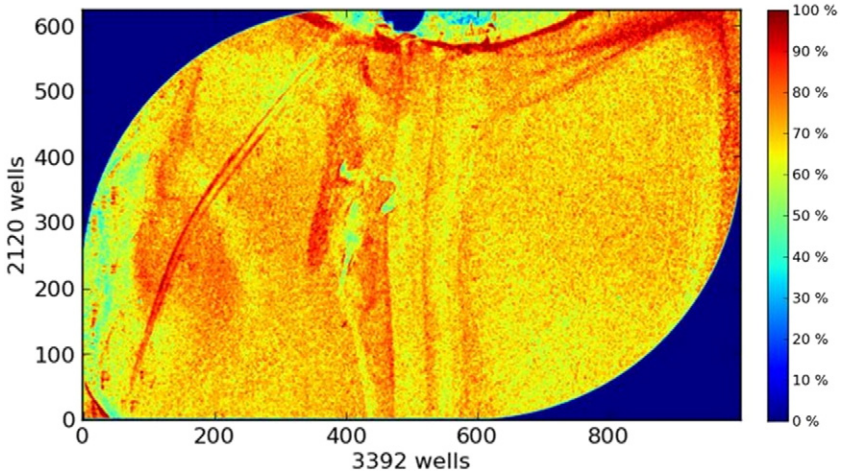


Fig. 1. Pseudo-color image of the Ion 316™ v2 Chip plate showing percent loading across the physical surface. This sequencing run had a 70% loading, which ensures a high ISP density. Every 316 chip contains more than 6 million wells and the color scale on the right side conduces as a loading indicator. Deep red coloration stays for a 100% loading which means that every well in this area contains an ISP (templated and non-templated) whereas deep blue coloration implies that the wells in this area are empty.

The sequencing of the whole cohort required seven separate runs with each 11–12 patients' samples. Coverage statistics (Table 1) were comparable among all runs and were in the range of accepted quality criteria [40–42]. During the seven runs, a median of 2.54×10^6 reads per run was generated. The mean depth was near from 200 reads, the mean read length evaluated 163 bases and average chip loading (Fig. 1) was 62%. To ensure a high density of ISPs on a chip and hence, a high sequencing output, the chip loading value should be ≥60% (Table 2). The observed NGS results matched with the results obtained with conventional sequencing of random samples (Fig. 2). In all validation samples, the correspondence between NGS and Sanger derived nucleotide sequences was 100%.

From the NGS runs, a total of 35,447 nucleotides were read from each DNA sample. Following elimination of nucleotides agreeing with the standard human genome sequence GRCh37 hg19 (dated February 2009), the final result of the NGS consisted of a vector of nucleotide information for each individual DNA sample. This vector had a length equaling the number of chromosomal positions in which a non-reference nucleotide had been found in any probe of the actual study sample. In the 28 pain patients who required standard opioid doses for analgesia (for clinical details, see [29]) and could therefore substitute for a random sample with respect to opioid receptor genetics, these vectors had a length of 91 nucleotides (Fig. 3). They included the genetic information about known functional opioid receptor variants such as about the *OPRM1* rs1799971 SNP (118 A > G; Chr6:154360797 in Fig. 3) as the so far scientifically most regarded variant possibly triggering reduced opioid efficacy or about the *OPRM1* rs563649 SNP (Chr6:154407967 in Fig. 3) coding for the μ-opioid receptor splice variant MOR-1K probably triggering opioid-induced hyperalgesia [43]. However, the full NGS information comprises a whole matrix composed of the single vectors and the evaluation and interpretation of this information requires bioinformatics approaches able to unleash the advances of NGS over single variant approaches as shown elsewhere [29].

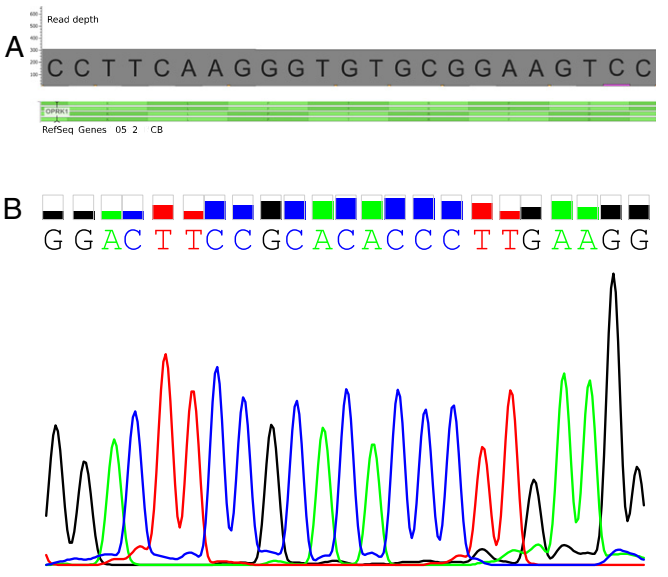


Fig. 2. The Figure shows the alignment of the ion torrent sequence of the κ-opioid receptor illustrated by Golden Helix Genome Browse® readout (A) versus the same sequence according to a Sanger electropherogram (B). The grey bars in panel A show the read depth of the single nucleotides (all nucleotide positions ≈ 300 reads). The green scale in panel A shows the amino acids from different transcripts in detail. The colored bars in Panel B indicate for a sequencing quality parameter, which means a high filled box means a trustworthy sequencing process.

4. Discussion

A comprehensive NGS assay for the exons and regulatory parts of human opioid receptors was developed that produced valid nucleotide sequences corresponding to those obtained with the classical Sanger sequencing technique. The NGS assay is suitable for small to large-scale experimental setups aiming at accessing the information about any nucleotide in a study cohort, with a selection of those that differ from the reference nucleotide.

Single opioid receptor variants have been causally involved in reduced responses to analgesic treatment [44], hyperalgesic responses to opioid analgesics [45], protection against opioid induced side effects [46,47], the individual sensitivity to painful stimuli [48,49], the risk for drug addiction [50,51], the individual response to stress [12], the success of naloxone treatment of alcoholism [52], or even the susceptibility to breast cancer [53] or the incidence of sudden infant death syndrome [54]. At the molecular level, in particular μ-opioid receptor variants have been shown to reduce opioid receptor expression [16,17], receptor signaling [17–19] up to an almost complete functional loss [18,20], and alternative splicing with the appearance of a six-transmembrane segment μ-opioid receptor [45] that triggers excitatory effects [43]. However, despite this apparently successful research on opioid receptor genetics, results have neither entered clinical practice nor have they been included in main treatment guidelines for pain or addiction [24,55]. This may have several causes such as too small effect sizes counterbalanced by concomitantly other genetic variants [56]. However, an important perception of the discrepancy between the pathophysiological importance of opioid receptors including their role as main drug targets and the modest utility of their genotyping is, that the limited selection of published functional variants with accessibility by specific genetic assays [23] is probably insufficient. Indeed, we have recently shown that NGS of human opioid receptor genes outperforms single opioid receptor variants as a genetic biomarker of opioid-analgesia related phenotypes [29].

Research interest in the complete genomic information dates back to the seventies of the last century when the selective incorporation of chain-terminating dideoxynucleotides by DNA polymerase during in vitro DNA replication had been introduced [25,26]. Techniques significantly improved during the last decades [57] with the development of contemporary machines in the late 1990s released to the market around the year 2005. The term “next generation” DNA sequencing refers to high-throughput technologies capable of parallel analyses of large numbers of different DNA sequences in a single reaction [58]. NGS has been attributed the potential to accelerate biomedical research [28,59,60]. Current leading manufacturers of commercial NGS platforms are Illumina (Genome Analyzer, HiSeq, MiSeq) and Life Technologies (SOLiD System, Ion Torrent, Proton). The systems combine conceptually similar workflows, starting with the creation of the genetic sample, which commences library preparation involving fragmentation of genomic DNA, purifying to uniform and desired fragment size and ligation to sequencing adapters specific to the platform. Differences apply to the reaction biochemistry and the way how the sequencing information is read [58]. In the present ion semiconductor sequencing method, libraries are immobilized to beads and amplified in microdroplets of aqueous solution and oil using emulsion PCR. Individual nucleotide bases are incorporated via DNA polymerase, which in the case of success triggers

Table 2
Ion Torrent PGM Run Characteristics (n = 7) of the NGS of human opioid receptor genes in n = 79 DNA samples obtained from pain patients.

Chip	Samples on chip	Loading density (%)	Total bases	Total Reads	Mean read length (bp)
316_1	12	65	502 M	2,949,673	170
316_2	12	80	220 M	1,755,623	125
316_3	11	63	339 M	2,060,050	165
316_4	11	50	406 M	2,374,049	171
316_5	11	67	540 M	3,186,425	169
316_6	11	56	477 M	2,813,274	170
316_7	11	52	495 M	2,645,442	174

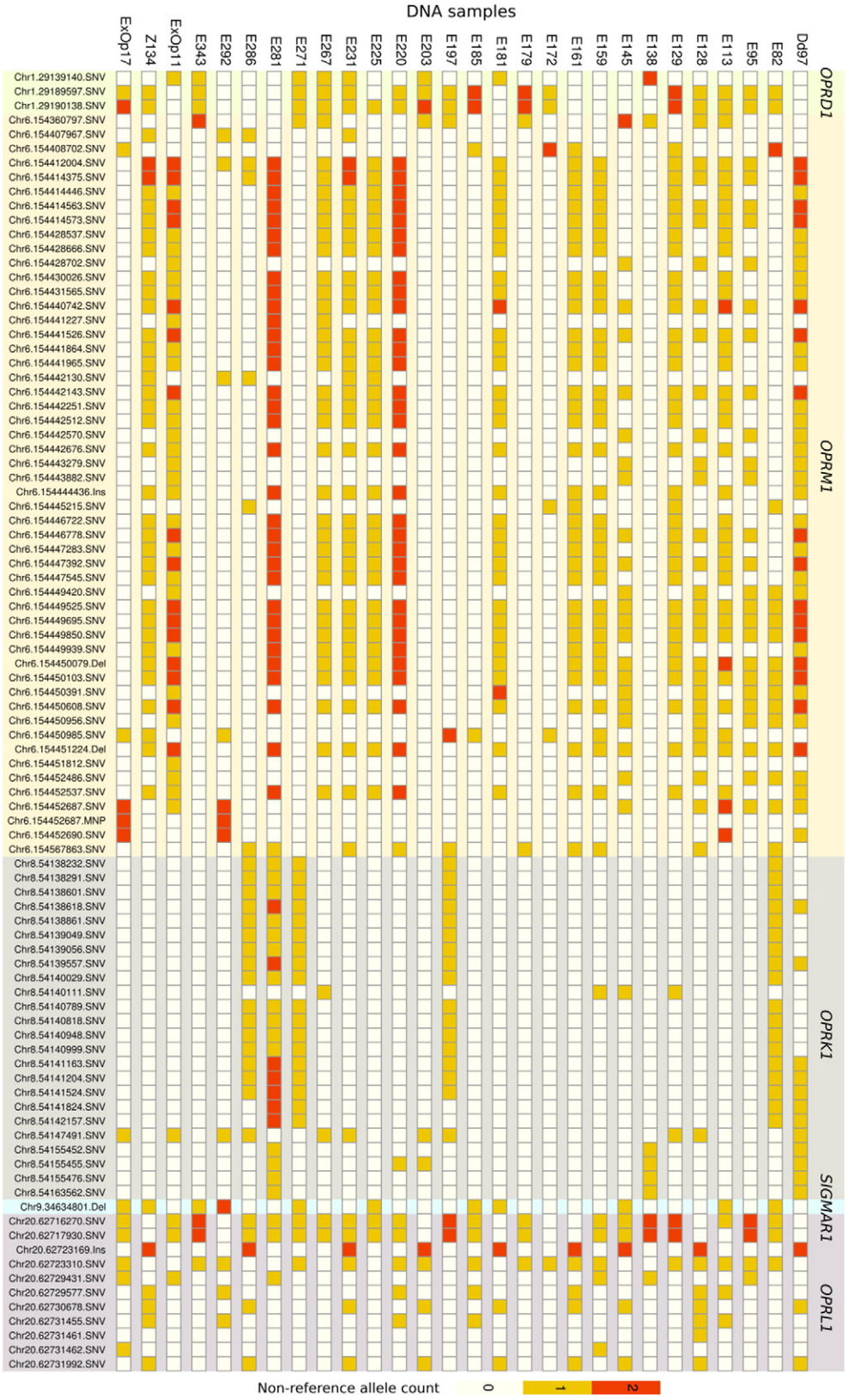


Fig. 3. Example opioid receptors' genetic pattern of 28 pain patients receiving standard opioid doses (extended from [29]). The plot shows the occurrence of variants (lines) per DNA sample (columns) as vectors of a length corresponding to the number of gene loci in which a non-reference nucleotide was found in any sample of the whole cohort. The vectors are composed of information about the number of non-reference alleles found at the respective locus in the respective sample, color codes as white, 0 non-reference alleles = wild type genotype, yellow, heterozygous, and red, 2 non-reference alleles. The opioid receptors are indicated by different color shading behind the vectors of individual genetic information.

the release of a proton. The semiconductor chip that acts as a pH meter [61] providing the final readout. Alternative techniques use the detection of light instead, i.e., from optical fluorescence signals in the case of successful nucleotide incorporation the DNA nucleotide sequence is assembled. The different techniques differ with respect to the obtained throughput and accuracy [62,63].

The high throughput and comprehensive information about DNA sequences are presently reflected in the assay costs. The sequencing of the opioid receptor genes of 79 patients required € 4500 for the AmpliSeq™ custom panel, € 5800 for library preparation, € 980 for template preparation and € 1200 for sequencing. In addition, approximately € 800 was spent for consumables and supplies. With twelve barcoded samples loaded on each chip, analysis costs for a single patient's opioid receptor gene sequence were approximately € 168. NGS costs are expected to quickly fall in near future [64]. However, despite this rapid technological progress, the analysis of the generated large data sets remains challenging [65]. As the sequencing process is only the beginning of the procedure, the analysis of the resulting "big data" requires substantial computational power, bioinformatics expertise and "up to date" databases of genomic variations. NGS technologies seem to shift the workload essentially away from the laboratory sample preparation toward various data analysis processes.

Research on the genetic variation of human opioid receptors as the main targets of endogenous and exogenous opioids is of immediate interest for assessing the clinical effects of opioid analgesics and for studying the epidemiology of substance addiction. The present NGS method is suitable for large-scale sequencing of an extended set of human genes belonging or related to opioid receptors. By covering almost the complete relevant coding and regulatory parts of the genes, the method includes all variants studied so far for functional associations and adds a large amount of genetic information as a basis for complete analysis of human opioid receptor genetics and its functional consequences.

Conflict of interest statement

The authors have declared that no further conflicts of interest exist.

Author contributions

Conceived and designed the experiments: JL, DK. Performed the experiments: DK. Analyzed the data: JL, DK. Wrote the paper: JL, DK.

Funding

This work has been funded by the European Union Seventh Framework Programme (FP7/2007–2013) under grant agreement no. 602919 ("GLORIA", JL). An additional contribution, in particular the funding of the next generation genotyping equipment, has been received from the Landesoffensive zur Entwicklung wissenschaftlich-ökonomischer Exzellenz (LOEWE), Zentrum: Translational Medicine and Pharmacology (JL). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Appendix A. Supplementary data

Supplementary information includes (i) a list of PCR primer used for the NGS assay (Supplementary Table 1) and (ii) a list of missed parts from the opioid receptor gene panel (Supplementary Table 2). Supplementary data associated with this article can be found in the online version, at doi:10.1016/j.cca.2016.10.009.

References

[1] B.N. Dhawan, F. Cesselin, R. Raghubir, et al., International Union of Pharmacology. XII. Classification of opioid receptors, Pharmacol. Rev. 48 (1996) 567–592.
[2] C.B. Pert, S.H. Snyder, Opiate receptor: demonstration in nervous tissue, Science 179 (1973) 1011–1014.
[3] C. Stein, The control of pain in peripheral tissue by opioids, N. Engl. J. Med. 332 (1995) 1685–1690.
[4] G. Kobal, Pain-related electrical potentials of the human nasal mucosa elicited by chemical stimulation, Pain 22 (1985) 151–163.
[5] M.C. Lee, V. Wanigasekera, I. Tracey, Imaging opioid analgesia in the human brain and its potential relevance for understanding opioid use in chronic pain, Neuropharmacology 84 (2014) 123–130.
[6] D.D. Price, Psychological and neural mechanisms of the affective dimension of pain, Science 288 (2000) 1769–1772.
[7] M.C. Bushnell, G.H. Duncan, R.K. Hofbauer, B. Ha, J.I. Chen, B. Carrier, Pain perception: is there a role for primary somatosensory cortex? Proc. Natl. Acad. Sci. U. S. A. 96 (1999) 7705–7709.
[8] J. Lötsch, R. Dudziak, R. Freynhagen, J. Marschner, G. Geisslinger, Fatal respiratory depression after multiple intravenous morphine injections, Clin. Pharmacokinet. 45 (2006) 1051–1060.
[9] M.C. Borras, L. Becerra, A. Ploghaus, et al., fMRI measurement of CNS responses to naloxone infusion and subsequent mild noxious thermal stimuli in healthy volunteers, J. Neurophysiol. 91 (2004) 2723–2733.
[10] N.Y. Kirson, A. Shei, A.G. White, et al., Societal economic benefits associated with an extended-release opioid with abuse-deterrent technology in the United States, Pain Med. 15 (2014) 1450–1454.
[11] C. Bond, K.S. LaForge, M. Tian, et al., Single-nucleotide polymorphism in the human mu opioid receptor gene alters beta-endorphin binding and activity: possible implications for opiate addiction, Proc. Natl. Acad. Sci. U. S. A. 95 (1998) 9608–9613.
[12] G.S. Wand, M. McCaul, X. Yang, et al., The mu-opioid receptor gene polymorphism (A118G) alters HPA axis activation induced by opioid receptor blockade, Neuropsychopharmacology 26 (2002) 106–114.
[13] P. Mosińska, M. Zielińska, J. Fichna, Expression and physiology of opioid receptors in the gastrointestinal tract, Curr. Opin. Endocrinol. Diabetes Obes. 23 (2016) 3–10.
[14] G. Drolet, E.C. Dumont, I. Gosselin, R. Kinkead, S. Laforest, J.F. Trottier, Role of endogenous opioid system in the regulation of the stress response, Prog. Neuro-Psychopharmacol. Biol. Psychiatry 25 (2001) 729–741.
[15] J. Lötsch, G. Geisslinger, Are mu-opioid receptor polymorphisms important for clinical opioid therapy? Trends Mol. Med. 11 (2005) 82–89.
[16] Y. Zhang, D. Wang, A.D. Johnson, A.C. Papp, W. Sadee, Allelic expression imbalance of human mu opioid receptor (OPRM1) caused by variant A118G, J. Biol. Chem. 280 (2005) 32618–32624.
[17] B.G. Oertel, A. Doehring, B. Roskam, et al., Genetic-epigenetic interaction modulates mu-opioid receptor regulation, Hum. Mol. Genet. 21 (2012) 4751–4760.
[18] K. Befort, D. Filliol, F.M. Decaillot, C. Gaveriaux-Ruff, M.R. Hoehe, B.L. Kieffer, A single-nucleotide polymorphic mutation in the human mu-opioid receptor severely impairs receptor signaling, J. Biol. Chem. 276 (2001) 3130–3137.
[19] B.G. Oertel, M. Kettner, K. Scholich, et al., A common human mu-opioid receptor genetic variant diminishes the receptor signaling efficacy in brain regions processing the sensory information of pain, J. Biol. Chem. 284 (2009) 6530–6535.
[20] T. Koch, T. Kroslak, M. Averbeck, et al., Allelic variation S268P of the human mu-opioid receptor affects both desensitization and G protein coupling, Mol. Pharmacol. 58 (2000) 328–334.
[21] P. Klepstad, T. Fladvad, F. Skorpen, et al., Influence from genetic variability on opioid use for cancer pain: a European genetic association study of 2294 cancer pain patients, Pain 152 (2011) 1139–1145.
[22] J. Lötsch, G. Geisslinger, A critical appraisal of human genotyping for pain therapy, Trends Pharmacol. Sci. 31 (2010) 312–317.
[23] C. Skarke, A. Kirchhof, G. Geisslinger, J. Lötsch, Comprehensive mu-opioid-receptor genotyping by pyrosequencing, Clin. Chem. 50 (2004) 640–644.
[24] J. Lötsch, G. Geisslinger, Relevance of frequent mu-opioid receptor polymorphisms for opioid activity in healthy volunteers, Pharmacogenomics J. 6 (2006) 200–210.
[25] F. Sanger, A.R. Coulson, A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase, J. Mol. Biol. 94 (1975) 441–448.
[26] F. Sanger, S. Nicklen, A.R. Coulson, DNA sequencing with chain-terminating inhibitors, Proc. Natl. Acad. Sci. U.S.A. 74 (1977) 5463–5467.
[27] M. Frank, A. Prenzler, R. Eils, G. von der Schulenburg JM, Genome sequencing: a systematic review of health economic evidence, Heal. Econ. Rev. 3 (2013) 29.
[28] M.L. Metzker, Sequencing technologies - the next generation, Nat. Rev. Genet. 11 (2010) 31–46.
[29] D. Kringel, A. Ultsch, M. Zimmermann, et al., Emergent biomarker derived from next generation sequencing to identify pain patients requiring uncommonly high opioid doses, Pharmacogenomics J. (2016).
[30] V. Law, C. Knox, Y. Djoumbou, et al., DrugBank 4.0: shedding new light on drug metabolism, Nucleic Acids Res. 42 (2014) D1091–D1097.
[31] W.R. Martin, C.G. Eades, J.A. Thompson, R.E. Huppler, P.E. Gilbert, The effects of morphine- and nalorphine- like drugs in the nondependent and morphine-dependent chronic spinal dog, J. Pharmacol. Exp. Ther. 197 (1976) 517–532.
[32] S. Narayanan, R. Bhat, C. Mesangeau, J.H. Poupaert, C.R. McCurdy, Early development of sigma-receptor ligands, Future Med. Chem. 3 (2011) 79–94.
[33] L.L. Werling, A. Keller, J.G. Frank, S.J. Nuwayhid, A comparison of the binding profiles of dextromethorphan, memantine, fluoxetine and amitriptyline: treatment of involuntary emotional expression disorder, Exp. Neurol. 207 (2007) 248–257.
[34] C. Sánchez-Fernández, Á. Montilla-García, R. González-Cano, et al., Modulation of peripheral μ-opioid analgesia by σ1 receptors, J. Pharmacol. Exp. Ther. 348 (2014) 32–45.
[35] G. Henderson, A.T. McKnight, The orphan opioid receptor and its endogenous ligand–nociceptin/orphanin FQ, Trends Pharmacol. Sci. 18 (1997) 293–300.

[36] T. Takahashi, K. Okubo, S. Kojima, et al., Antihyperalgesic effect of buprenorphine involves nociceptin/orphanin FQ peptide-receptor activation in rats with spinal nerve injury-induced neuropathy, J. Pharmacol. Sci. 122 (2013) 51–54.
[37] S.M. Brown, M. Holtzman, T. Kim, E.D. Kharasch, Buprenorphine metabolites, buprenorphine-3-glucuronide and norbuprenorphine-3-glucuronide, are biologically active, Anesthesiology 115 (2011) 1251–1260.
[38] C. Mollereau, M. Parmentier, P. Mailleux, et al., ORL1, a novel member of the opioid receptor family. Cloning, functional expression and localization, FEBS Lett. 341 (1994) 33–38.
[39] J. Tarabeux, B. Zeitouni, V. Moncoutier, et al., Streamlined Ion Torrent PGM-Based Diagnostics: BRCA1 and BRCA2 Genes as a Model, 2014.
[40] G. Millat, V. Chanavat, R. Rousson, Evaluation of a new NGS method based on a custom AmpliSeq library and ion torrent PGM sequencing for the fast detection of genetic variations in cardiomyopathies, Clin. Chim. Acta 433 (2014) 266–271.
[41] P. Concolino, A. Costella, A. Minucci, et al., A preliminary quality control (QC) for next generation sequencing (NGS) library evaluation turns out to be a very useful tool for a rapid detection of BRCA1/2 deleterious mutations, Clin. Chim. Acta 437 (2014) 72–77.
[42] A.S. Glotov, S.V. Kazakov, E.A. Zhukova, et al., Targeted next-generation sequencing (NGS) of nine candidate genes with custom AmpliSeq in patients and a cardiomyopathy risk group, Clin. Chim. Acta 446 (2015) 132–140.
[43] F.A. Oladosu, M.S. Conrad, S.C. O'Buckley, N.U. Rashid, G.D. Slade, A.G. Nackley, Mu opioid splice variant MOR-1K contributes to the development of opioid-induced hyperalgesia, PLoS One 10 (2015), e0135711.
[44] A.A. Somogyi, J.K. Coller, D.T. Barratt, Pharmacogenetics of opioid response, Clin. Pharmacol. Ther. 97 (2015) 125–127.
[45] S.A. Shabalina, D.V. Zaykin, P. Gris, et al., Expansion of the human mu-opioid receptor gene architecture: novel functional variants, Hum. Mol. Genet. (2009).
[46] J. Lötsch, M. Zimmermann, J. Darimont, et al., Does the A118G polymorphism at the mu-opioid receptor gene protect against morphine-6-glucuronide toxicity? Anesthesiology 97 (2002) 814–819.
[47] B.G. Oertel, R. Schmidt, A. Schneider, G. Geisslinger, J. Lötsch, The mu-opioid receptor gene polymorphism 118 A > G depletes alfentanil-induced analgesia and protects against respiratory depression in homozygous carriers, Pharmacogenet. Genomics 16 (2006) 625–636.
[48] R.B. Fillingim, L. Kaplan, R. Staud, et al., The A118G single nucleotide polymorphism of the mu-opioid receptor gene (OPRM1) is associated with pressure pain sensitivity in humans, J. Pain 6 (2005) 159–167.
[49] J. Lötsch, B. Stuck, T. Hummel, The human mu-opioid receptor gene polymorphism 118 A > G decreases cortical activation in response to specific nociceptive stimulation, Behav. Neurosci. 120 (2006) 1218–1224.
[50] T.-K. Clarke, R.C. Crist, K.M. Kampman, et al., Low frequency genetic variants in the μ-opioid receptor (OPRM1) affect risk for addiction to heroin and cocaine, Neurosci. Lett. 542 (2013) 71–75.
[51] I. Deb, J. Chakraborty, P.K. Gangopadhyay, S.R. Choudhury, S. Das, Single-nucleotide polymorphism (A118G) in exon 1 of OPRM1 gene causes alteration in downstream signaling by mu-opioid receptor and may contribute to the genetic risk for addiction, J. Neurochem. 112 (2010) 486–496.
[52] R. Pal, J.E. Mendelson, K. Flower, et al., Impact of prospectively determined A118G polymorphism on treatment response to injectable naltrexone among methamphetamine-dependent patients: an open-label, pilot study, J. Addict. Med. 9 (2015) 130–135.
[53] A. Cieślińska, E. Sienkiewicz-Szłapka, E. Kostyra, et al., μ-Opioid receptor gene (OPRM1) polymorphism in patients with breast cancer, Tumour Biol. 36 (2015) 4655–4660.
[54] K. Läer, T. Dörk, M. Vennemann, T. Rothämel, M. Klintschar, Polymorphisms in genes of respiratory control and sudden infant death syndrome, Int. J. Legal Med. 129 (2015) 977–984.
[55] J.S. Mogil, Are we getting anywhere in human pain genetics? Pain 146 (2009) 231–232.
[56] J. Lötsch, K. Flühr, T. Neddermayer, A. Doehring, G. Geisslinger, The consequence of concomitantly present functional genetic variants for the identification of functional genotype-phenotype associations in pain, Clin. Pharmacol. Ther. 85 (2009) 25–30.
[57] CADTH, Rapid response reports, Next Generation DNA Sequencing: A Review of the Cost Effectiveness and Guidelines, Ottawa (ON), Canadian Agency for Drugs and Technologies in Health, 2014.
[58] J.M. Rizzo, M.J. Buck, Key principles and clinical applications of "next-generation" DNA sequencing, Cancer Prev. Res. 5 (2012) 887–900.
[59] E.R. Mardis, Next-generation DNA sequencing methods, Annu. Rev. Genomics Hum. Genet. 9 (2008) 387–402.
[60] J. Shendure, H. Ji, Next-generation DNA sequencing, Nat. Biotechnol. 26 (2008) 1135–1145.
[61] J.M. Rothberg, W. Hinz, T.M. Rearick, et al., An integrated semiconductor device enabling non-optical genome sequencing, Nature 475 (2011) 348–352.
[62] N.J. Loman, R.V. Misra, T.J. Dallman, et al., Performance comparison of benchtop high-throughput sequencing platforms, Nat. Biotechnol. 30 (2012) 434–439.
[63] D. Sims, I. Sudbery, N.E. Ilott, A. Heger, C.P. Ponting, Sequencing depth and coverage: key considerations in genomic analyses, Nat. Rev. Genet. 15 (2014) 121–132.
[64] K. Lohmann, C. Klein, Next generation sequencing and the future of genetic diagnosis, Neurotherapeutics 11 (2014) 699–707.
[65] N. Norton, D. Li, R.E. Hershberger, Next-generation sequencing to identify genetic causes of cardiomyopathies, Curr. Opin. Cardiol. 27 (2012) 214–220.

Erklärung zu den Autorenanteilen an der Publikation:

Emergent biomarker derived from next-generation sequencing to identify pain patients requiring uncommonly high opioid doses (*printed*)

Name der Zeitschrift: The Pharmacogenomics Journal

Beteiligte Autoren: D Kringel, A Ultsch, M Zimmermann, J P Jansen, W Ilias, R Freynhagen, N Griessinger, A Kopf, C Stein, A Doehring, E Resch und J Lötsch

Was hat der Promovierende bzw. was haben die Koautoren beigetragen?

(1) zu Entwicklung und Planung

Promovierender: 50%
Autor JL: 50%

(2) zur Durchführung der einzelnen Untersuchungen und Experimente

Promovierender: 100% (DNA Extraktion, NGS-Panel Design, Sequenzierung, Varianten Annotation, Methoden Validierung)

(3) zur Erstellung der Datensammlung und Abbildungen

Promovierender: 40% (Datenerhebung, Run-Statistik)
Autor JL: 60% (Erstellung der Abbildungen)

(4) zur Analyse und Interpretation der Daten

Promovierender: 30% (Datenvorverarbeitung, Charakterisierung der Varianten, Dateninterpretation)
Autor JL: 70% (Datenvisualisierung, Programmierung der KI, Anwendung der Klassifikatoren, Datenanalyse, Dateninterpretation)

(5) zum Verfassen des Manuskripts

Promovierender: 25%
Autor AU: 5%
Autor MZ: 5%
Autor JPJ: 5%
Autor WI: 5%
Autor RF: 5%
Autor NG: 5%
Autor AK: 5%
Autor CS: 5%
Autor AD: 5%
Autor ER: 5%
Autor JL: 25%

Zustimmende Bestätigungen der oben genannten Angaben:

_____ _____

Datum/Ort Unterschrift Promovend

_____ _____

Datum/Ort Unterschrift Betreuer

OPEN

The Pharmacogenomics Journal (2017) 17, 419–426

www.nature.com/tpj

ORIGINAL ARTICLE

Emergent biomarker derived from next-generation sequencing to identify pain patients requiring uncommonly high opioid doses

D Kringel[1], A Ultsch[2], M Zimmermann[3], J-P Jansen[4], W Ilias[5], R Freynhagen[6,7], N Griessinger[8], A Kopf[9], C Stein[9], A Doehring[1], E Resch[10] and J Lötsch[1,10]

Next-generation sequencing (NGS) provides unrestricted access to the genome, but it produces 'big data' exceeding in amount and complexity the classical analytical approaches. We introduce a bioinformatics-based classifying biomarker that uses emergent properties in genetics to separate pain patients requiring extremely high opioid doses from controls. Following precisely calculated selection of the 34 most informative markers in the *OPRM1*, *OPRK1*, *OPRD1* and *SIGMAR1* genes, pattern of genotypes belonging to either patient group could be derived using a *k*-nearest neighbor (*k*NN) classifier that provided a diagnostic accuracy of 80.6 ± 4%. This outperformed alternative classifiers such as reportedly functional opioid receptor gene variants or complex biomarkers obtained via multiple regression or decision tree analysis. The accumulation of several genetic variants with only minor functional influences may result in a qualitative consequence affecting complex phenotypes, pointing at emergent properties in genetics.

The Pharmacogenomics Journal (2017) **17,** 419–426; doi:10.1038/tpj.2016.28; published online 3 May 2016

INTRODUCTION

Genotyping-based drug therapy decisions are increasingly desired in clinical practice; however, their introduction is still limited. The so far published single functional genetic variants are increasingly perceived as insufficient in providing a genetic diagnosis of clinical phenotypes.[1] However, exploitation of the whole genetic information becomes possible overcoming the restricted selection of known variants. Next-generation sequencing (NGS)[2] provides unrestricted access to the subjects' genome.

As the resulting 'big data' exceeds in its amount and complexity the classical approaches, the analysis of NGS derived data is an active research topic. It has already led to working solutions (for review, see Nielsen *et al.*[3] and Pabinger *et al.*[4]) and new statistical methods for analyzing NGS data are continuously emerging.[5–9] However, methods to convert NGS-derived big data into biomarkers are still sparse and solutions exploiting the whole genomic information content for patient classification are still needed. Novel types of classifying biomarkers derived from NGS information are needed, and we report a subsymbolic classifier that uses emergent properties in genetics[10] with the potential of self-learned improvement from successively addable genetic and clinical information.

The biomarker was developed based on the clinical problem of extremely high opioid demands by some pain patients without any indication of addiction. These patients were subjected to our pharmacogenetic counseling for opioid receptor genetics as the primary candidates coding for the main targets of this class of drugs. The results show that the biomarker utilizing comprehensive DNA sequence information outperforms classical approaches at genetics-based patient classification and promises the utilization of complex information in NGS-derived genotypes for successful clinical diagnostics.

MATERIALS AND METHODS

Patients

The investigation followed the Declaration of Helsinki on Biomedical Research Involving Human Subjects and was approved by the ethics committee of the Medical Faculty of the Goethe-University, Frankfurt, Germany (ethics protocol number E 195/08). Patients were included for whom pharmacogenetic counseling, in particular opioid receptor genotyping, had been requested because of the perception of uncommonly high analgesic opioid dosing requirements without any obvious clinical reason and explicit denial of an addiction background. Using conversion to oral morphine equivalents (OMEs; for details, see Supplementary Table 1), these patients were divided into two groups with high ($\geqslant$400 mg day^{-1} OME; $n = 30$; mean ± s.d.: 3.04 ± 0.34 log OME) and common ($\leqslant$100 mg day^{-1} OME; $n = 28$; 'controls'; mean ± s.d.: 1.7 ± 0.24 log OME) opioid dosing requirements for chronic pain therapy. Upon written informed consent, a venous blood sample was taken from each patient; the samples were anonymized and sent to our laboratory at the Institute of Clinical Pharmacology in Frankfurt.

Opioid receptor genotyping using NGS

DNA preparation and amplification. Genomic DNA was extracted from 200 µl venous blood on a BioRobot EZ1 workstation applying the blood and body fluid spin protocol provided in the EZ1 DNA Blood 200 µl Kit

[1]Institute of Clinical Pharmacology, Goethe-University, Frankfurt am Main, Germany; [2]DataBionics Research Group, University of Marburg, Marburg, Germany; [3]Department of Anesthesiology, Intensive Care Medicine and Pain Therapy, University Hospital Frankfurt, Frankfurt am Main, Germany; [4]Schmerzzentrum Berlin, Berlin, Germany; [5]Department of Anaesthesiology and Intensive Care Medicine, Vienna, Austria; [6]Zentrum für Anästhesiologie, Intensivmedizin, Schmerztherapie & Palliativmedizin, Benedictus Krankenhaus Tutzing, Tutzing, Germany; [7]Klinik für Anästhesiologie, Technische Universität München, München, Germany; [8]Department of Anesthesiology, University Hospital Erlangen, Erlangen, Germany; [9]Department of Anesthesiology and Critical Care Medicine, Freie Universität Berlin–Charité, Berlin, Germany and [10]Fraunhofer Institute for Molecular Biology and Applied Ecology IME, Project Group Translational Medicine and Pharmacology TMP, Frankfurt am Main, Germany. Correspondence: Professor Dr J Lötsch, Institute of Clinical Pharmacology, Goethe-University, Theodor Stern Kai 7, Frankfurt am Main 60590, Germany.
E-mail: j.loetsch@em.uni-frankfurt.de
Received 26 September 2015; revised 5 November 2015; accepted 13 November 2015; published online 3 May 2016

(Qiagen, Hilden, Germany). A multiplex amplification primer set for the exonic sequences of the opioid receptor genes (*OPRM1*, *OPRK1*, *OPRD1* and *SIGMAR1*, located on chromosomes 6, 8, 1 and 9, respectively) was designed online using a web tool (Ion Ampliseq Designer; Life Technologies, Darmstadt, Germany) provided by the manufacturer of the NGS device at http://www.ampliseq.com. Sequencing gaps affected only noncoding regions (for details, see Supplementary Table 2). A total of 10 ng DNA per sample were used for the target enrichment by a multiplex PCR and each DNA pool was amplified with the Ion AmpliSeq Library Kit 2.0 in conjunction with the Ion AmpliSeq 'custom Primer Pool' protocols according to the manufacturer's instructions (Life Technologies). After each pool had undergone 17 PCR cycles, amplicons were digested with FuPa Reagent partially removing the primer sequences followed by adaptors ligation. To enable multiplexing, sequencing adaptors with short lengths of index sequences (barcodes) were used (Ion Xpress Barcode Adapters Kit; Life Technologies). The adaptor-ligated amplicons were then purified using the Agentcourt AMPure XP beads (Beckman Coulter, Krefeld, Germany). After purification, fragment libraries were normalized to a final DNA concentration of 100 pM using the Ion Library Equalizer Kit (Life Technologies). Equalized barcoded libraries from 10 to 12 samples were pooled at a time. To clonally amplify the library DNA onto the Ion Sphere Particles (Life Technologies), a volume of 2 µl of the barcoded and pooled library was suspended to emulsion PCR using the Ion Personal Genome Machine (PGM) Template OT2 200 Kit on the Ion OneTouch 2 system (Life Technologies) according to the manufacturer's recommended protocol.

DNA NGS and variant identification. The template-positive Ion Sphere Particles were enriched on the Ion OneTouch ES (Life Technologies) and loaded on Ion 316 v2 Chips (500 Mb of expected sequence data output). Sequencing was performed with the Ion (PGM) system using Ion PGM 200 Sequencing Kit v2 (Life Technologies) with the 200-bp single-end run configuration following the manufacturer's instructions. Using the Torrent Suite software (version 4.4.2; Life Technologies), signal processing, base calling and the generation of unmapped and mapped BAM-files (hg19 reference genomic sequence) were performed. The variant calling (single-nucleotide polymorphisms, multi-nucleotide polymorphisms (MNPs), insertions (Ins) and deletions) across the hg19 reference genomic sequence was performed with the Torrent Variant Caller Plugin with following key parameters: minimum allele frequency = 0.1, minimum quality = 10, minimum coverage = 6 and minimum coverage on either strand = 0. The annotation of called variants was done with Ion Reporter software (version 4.2.2; Life Technologies) using VCF files from the Torrent Variant Caller as input. Data quality and coverage checks as well as variant identification were done using the single-nucleotide polymorphism and variation suite software (SVS version 8.3.3 for Linux 64-bit; Golden Helix, Bozeman, MT, USA). The correctness of the genotyping was verified using 10 amplifications of the coding parts of the genes that were completely conventionally sequenced by an independent commercial provider (LGC GmbH, Berlin, Germany).

Data analysis

Opioid receptor genotype differences between patients with high opioid dosage and controls were analyzed using the single-nucleotide polymorphism and variation suite software (SVS version 8.3.3 for Linux 64-bit; Golden Helix), the R software package (version 3.0.2 for Linux; http://CRAN.R-project.org/) and the Matlab numerical computing environment (version 8.3.0.532, MathWorks, Natick, MA, USA). The analyses are described briefly in the following; more detailed descriptions are provided in the Supplementary Materials.

Assessment of group differences in single opioid receptor variant or haplotype frequencies. A first analysis employed the classical approach to pharmacogenetic analyses consisting of χ^2 statistics with an α-level of 0.05 corrected for multiple testing according to the conservative criterion of Bonferroni. The dominant hereditary model (DD, Dd versus dd, where D denotes the minor allele and d the wild-type allele) was applied. In addition, haplotypes identified *in silico* via the minimizing historical recombination algorithm[11] were analogously assessed. To observe statistical power, only variants found at a frequency of $>10\%$[12] were included.

Assessment of combined genotypic group classifiers. Further analysis addressed combinations of genetic variants employing two classical methods comprising stepwise regression analysis and classification and regression tree (CART) analysis. In addition, advanced approaches at complex genetic group classifiers were used. As above approaches implicitly assumed homogeneity within the two patient cohorts, cluster analysis was used to reveal subgroups within each cohort. For this analysis, data were preprocessed to obtain the genotypes at each locus that could be expected by chance, which were obtained as the group size weighted means across all patients, separately for each variant. Subsequently, for each variant and patient the directed deviations were calculated as the difference between the actual observation and the above calculated expectation. These deviations were submitted to cluster analysis using the Ward algorithm and the Jaccard distance calculated as *1 – Jaccard coefficient*, the latter being identical to the percentage of nonzero values that differ. In a final step, these clusters served for the training of a *k*-nearest neighbor (*k*NN)[13] classifier with $k = 3$. Furthermore, the distance function was optimized by selection of the most informative variables using computed ABC analysis[14] that identified those genetic variants that promised to provide the best distinction between the two patient groups based on the difference in absolute group means of the number of rare-type alleles.

Classification performance analyses. For all single and complex classifiers cross-validated prediction performances were assessed. Test data sets (sample size $n = 20$) were drawn from the study cohort that always included (1) seven additional patients with high opioid dosage, and expanded by (2) further three patients randomly chosen from the already analyzed group with high opioid dosage to obtain a sufficiently large number for accuracy calculations and (3) 10 patients randomly chosen from the control group. Testing was 100 times repeated. The diagnostic accuracy, test sensitivity, specificity and positive predictive value were calculated using standard equations.

RESULTS

Group differences in single opioid receptor variant frequencies

Nucleotide information was completely available and sequencing runs met the standard requirements. The average throughput was 540 mega-bases, using 316 sequencing chips, which is within the upper third of the expected sequencing output according to the manufacturer's instructions, and the average chip loading was 71%, meeting the all expected assay quality parameters criteria.

NGS identified 152 variants in opioid receptor genes in the whole study population (Figure 1). A number of 100, 42, 3 and 7 variants were located in the covered sequences of the *OPRM1*, *OPRK1*, *OPRD1* and *SIGMAR1* genes. Group differences in the allelic frequencies were analyzed for 77 gene loci where variant alleles were found in at least 10% of the patients. This resulted in a corrected α-level of 0.000649351 as the upper limit for acceptance of a significant difference (Figure 2). A *P*-value below this limit resulted for the group frequency comparisons of the Chr6:154451812-single-nucleotide variation (SNV) in the µ-opioid receptor gene *OPRM1* ($P = 0.00049$, α-corrected $P = 0.038$) and for the Chr8:54147491-SNV in the κ-opioid receptor gene *OPRK1* ($P = 0.000278$, α-corrected $P = 0.021$), whereas a few other variants in the µ-opioid receptor gene, that is, Chr6:154444436-Ins (uncorrected $P = 0.0169$), Chr6:154452687-MNP (uncorrected $P = 0.036$) and Chr6:154567863-SNV (uncorrected $P = 0.0174$), merely displayed differences at an uncorrected significance level and were therefore rejected.

All opioid receptor genetic variants previously proposed as modulating opioid requirements (Table 1) failed to display significant group differences with respect to their allelic frequencies. For example, the variant *OPRM1* 118 G allele was found in 9 patients requiring high opioid doses and in 13 controls (uncorrected $P = 0.589$). However, none of the variants identified in this classical analysis provided an acceptable accuracy for the assignment of a patient to either group ($51.5 \pm 2\%$ and $51.4 \pm 5\%$) and the diagnostic sensitivities of these markers were also low (Table 3). Furthermore, a total of 24 different genetic variants carried only by pain patients who received very high opioid doses but not by any of the control patients provided an immediate biologically plausible cause for reduced or absent opioid receptor function (Table 2).

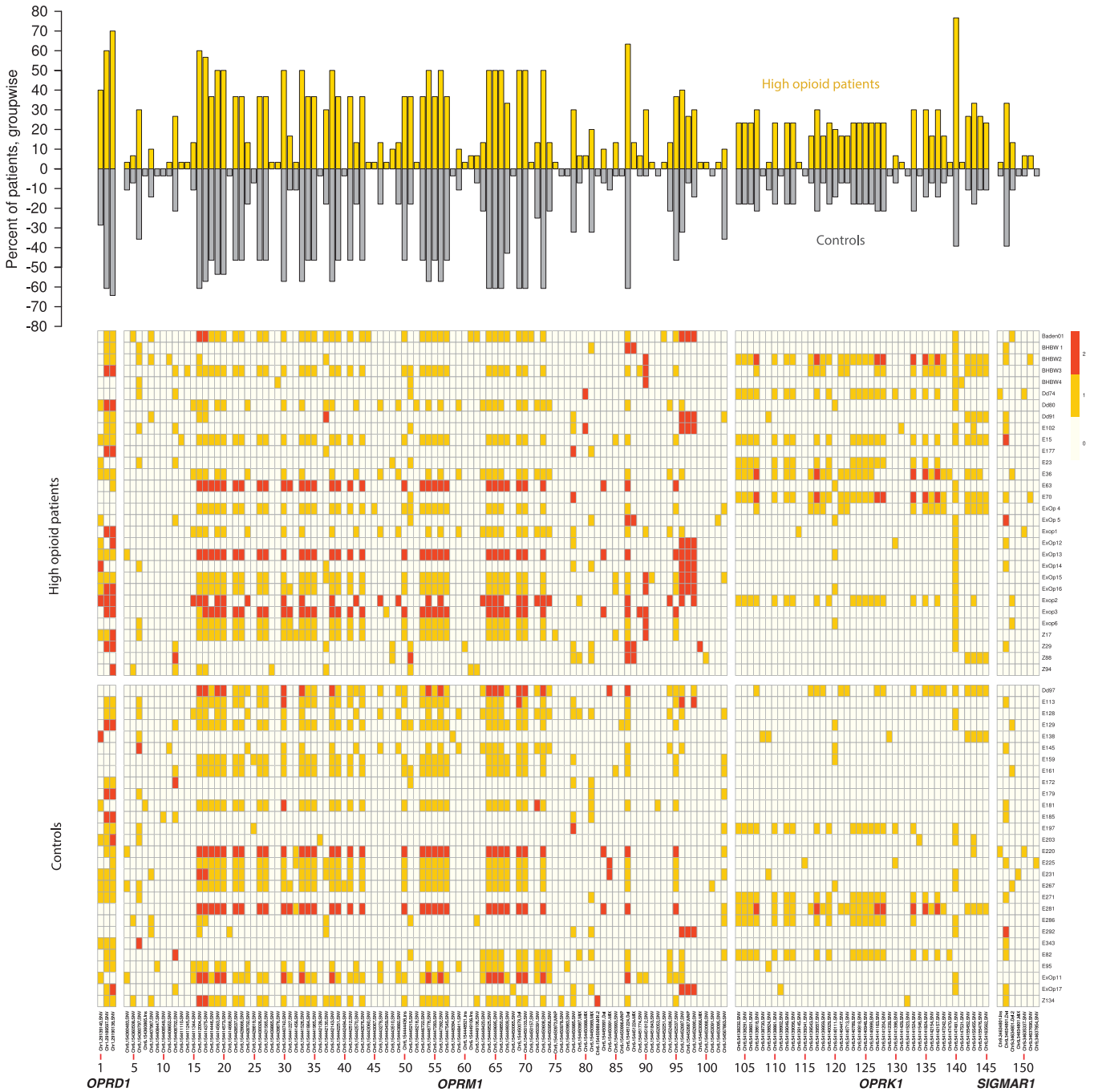


Figure 1. Overview of the opioid receptor genetic pattern of the present patients ($n = 30$ and 28 pain patients with high and average opioid dosage, respectively). The denominations of gene loci where variants have been detected is given at the bottom of the figure. (Top) Bar plot of the frequency of carriers of single-nucleotide polymorphisms (SNPs), with bar lengths indicating the percentage of patients carrying a variant allele (either heterozygously or homozygously); bars in the upper direction (yellow) show the high opioid patients ($n = 30$) and bars directed toward the bottom (dark gray) show the controls ($n = 28$). (Bottom) Matrix plot of the occurrence of variants (columns) per patient (lines; color coding is white: wild type, yellow: heterozygous, red: homozygous rare allele), separated by gaps for the two groups of patients and for the four opioid receptor genes (*OPRD1*, *OPRM1*, *OPRK1* and *SIGMAR1*) from the left to the right; the codes at the right are the patient's codes left in the figure for potential data validation purposes). For better visibility, the variants are numbered and the abscissa details are given in an enlarged version in the Supplementary Materials (DetailedAbscissa_Figure1.pdf).

Predictions using combined genetic markers obtained with classical approaches

An only moderate improvement of the opioid receptor genetics-based group classification performance was obtained using classical assessments of combined genotypic group classifiers. Stepwise regression analysis identified, based on the $P < 0.05$ inclusion and $P > 0.1$ rejection criteria, $n = 6$ opioid receptor variants, that is, Chr6:154443510-SNV, Chr6:154449106-Ins, Chr6:154450991-MIX, Chr6:154451812-SNV, Chr8:54147491-SNV and Chr8:54155452-SNV, as suitable components of a combined genotypic classifier for the

present group assignment. A linear regression model using these alleles performed with an average accuracy of 61.7 ± 6% on the 100 randomly drawn test data sets (Table 3). A decision tree classifier constructed with the CART algorithm used $n = 12$ opioid receptor variants, that is, Chr6:154360483-SNV, Chr6:154414666-SNV, Chr6:154439865-SNV, Chr6:154442128-SNV, Chr6:154443510-SNV, Chr6:154450065-SNV, Chr6:154450991-MIX, Chr6:154450996-MNP, Chr6:154451812-SNV, Chr6:154567863-SNV, Chr8:54147491-SNV and Chr8:54155452-SNV. The prediction of group assignment using this decision tree performed with an average accuracy of 74.4 ± 4% on the 100 randomly drawn test data sets (Table 3).

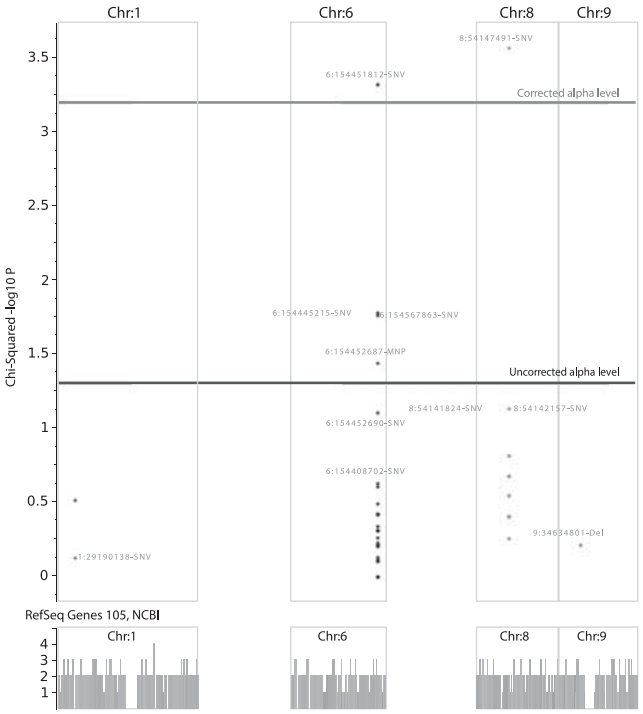


Figure 2. Manhattan plot showing the results of the genotype association test using the dominant hereditary model. Only chromosomes 1, 6, 8 and 9 are shown where the four opioid receptor genes are located. In addition, the α-levels before and after correction for multiple testing according to Bonferroni are indicated as horizontal lines.

Predictions using advanced approaches at complex genetic group classifiers

A Ward clustering using the Jaccard distance on all genetic markers led to clusters of $n = 5$ each in both the high opioid patients and the controls. The *k*NN classifier was trained with the pattern of each cluster per group that resulted in a number of typical patterns given by the number of clusters per patient group. Each of the 100 randomly drawn test data sets was compared in the high-dimensional data space with the obtained complete genotypes according to the Jaccard distance, and the case was assigned to that cluster to which the majority of its three neighbors belonged. The performance of this classifier on the test data resulted in an accuracy of 74.4 ± 7% matching that of the CART-derived classifier (Table 3). This result could be improved by eliminating those genetic markers that only introduced noise into the classifier. ABC analysis revealed that the genetic information contained at 34 loci in opioid receptor genes suffices for accurate group assignment to either patients with high opioid dosage or controls (Figure 3). With these markers, three clusters were identified for the control group, whereas four clusters appeared in the high opioid group. The performance of this *k*NN classifier on the test data, obtained as described in the previous paragraph, resulted in an accuracy of 80.6 ± 4% (Table 3).

DISCUSSION

The present analysis showed that patterns of opioid receptor genotypes indeed provide a basis for the high opioid dose occasionally observed in pain patients. However, this was obtained using a 'self-learning subsymbolic high-dimensional classifying biomarker'. That is, in machine learning a classifier is called symbolic if it can in principle answer the question of why a given data set has been assigned to a particular group; however, it is in principle not possible to get a reason for a particular classification. The presently used *k*NN classifier belongs to this type of algorithms. Its functioning is like an associative memory. For a given case the *k*NN classifier searches its data base of already learned correct classifications in order to find those cases that are most similar to the given data set.[15] If a data set has been classified successfully to its proper class, the data base of the classifier can be enhanced with this data set and classification ('machine learning'). In analogy to the

Table 1. Variants in opioid receptor genes for which functional consequences for opioid-based analgesic therapy had been reported previously, including example references for each variant

Gene	*cDNA*	*Nucleotide position*	*Effect*	*dbSNP*	*Clinical consequence*	*Reference*
OPRM1	c.575G>T	chr6:154411245	p.Cys192Phe	rs62638690	Heroin and cocaine addiction	26
	c.1206A>T	chr6:154414446	p.Gln402His	rs540825	Decreased effects in response to antidepressants	27
	c.118A>G	chr6:154360797	p.Asn40Asp	rs1799971	Decreased effects in response to opioids	28
	c.17C>T	chr6:154360696	p.Ala6Val	rs1799972	Heroin addiction; opioid dependence	29
	c.172C>T	chr6:154039373	p.Gln5His	rs6912029	Change-in-libido side effects; insomnia side effects	30
	c.1231C>T	chr6:154107531	p.Glu411AMB	rs677830	Decreased effects in response to opioids	27
	c.440C>G	chr6:154089975	p.Ser47Cys	rs17174794	Heroin and cocaine addiction	26
	c.1323A>G	chr6:154414563	p.(=)	rs675026	Increased risk in coronary heart disease	31
	c.1333C>T	chr6:154414573	p.(=)	rs562859	Decreased effects in response to antidepressants	32
OPRD1	c.921C>T	chr1:29189597	p.(=)	rs2234918	Increased effect sizes of pain	33
OPRK1	c.36G>T	chr8:54163562	p.(=)	rs1051660	Heroin addiction; alcohol dependence	34
	c.846C>T	chr8:54142154	p.(=)	rs16918875	Heroin addiction; alcohol dependence	35
	c.843A>G	chr8:54142157	p.(=)	rs702764	Heroin addiction; alcohol dependence	35
SIGMAR1	c.5A>C	chr9:34637690	p.Gln2Pro	rs1800866	Increased risk for developing Alzheimer's disease; decreased effects in response to antidepressants	36

Abbreviations: cDNA, complementary DNA; dbSNP, single-nucleotide polymorphism database; p.(=), synonymous variant.

Table 2. Genetic variants found only in patients receiving high opioid doses

Gene	*Variant*	*DNA change*	*Molecular consequence*	*Potential functional effect*[a]	*dbSNP ID*
OPRK1	54139145-SNV	c.*2712C>T	Noncoding	Reduced transcriptional efficiency	rs117602211
	54141429-SNV	c.*428G>A	Noncoding	Reduced transcriptional efficiency	rs182444059
	54147531-SNV	c.398T>C	p.Ile133Thr	Missense mutation	rs146859342
OPRM1	154411110-SNV	c.440C>G	p.Ser147Cys	Missense mutation	rs17174794
	154411245-SNV	c.575G>T	p.Cys192Phe	Missense mutation	rs62638690
	154439865-SNV	c.9C>T	—	—	rs11575858
	154439876-SNV	c.*20G>A	Noncoding	Reduced transcriptional efficiency	rs200778856
	154443060-SNV	—	—	—	—
	154443067-SNV	—	—	—	—
	154443459-SNV	—	—	—	—
	154443510-SNV	c.*3689A>G	Noncoding	Reduced transcriptional efficiency	rs188792757
	154446218-SNV	—	—	—	rs190450820
	154448521-Ins	—	—	—	—
	154449106-Ins	—	—	—	rs73022035
	154450157-SNV	—	—	—	—
	154450988-MIX	—	—	—	—
	154450973-MNP	—	—	—	—
	154451224-MIX	—	—	—	—
	154451843-SNV	c.*11987C>T	Noncoding	Reduced transcriptional efficiency	rs191957030
	154452251-SNV	c.*2395G>C	Noncoding	Reduced transcriptional efficiency	rs644261
	154453066-MIX	c.*13210A>G	Noncoding	Reduced transcriptional efficiency	rs184783311
	154453095-SNV	—	—	—	—
SIGMAR1	34637690-SNV	c.5A>C	p.Gln2Pro	Missense mutation	rs1800866

Abbreviations: dbSNP, single-nucleotide polymorphism database; mutation variants are as follows: Ins, insertion; MIX, a mixture of variation types; MNP, multi-nucleotide polymorphism; SNV, single-nucleotide variation. [a]Potential consequences according to this review.[21]

Table 3. Performances of selected classifiers to predict a patient with high opioid dosing

Classifier	*Group difference (P-uncorrected)**	*No. of genes*	*Sensitivity (%)*	*Specificity (%)*	*PPV (%)*	*Accuracy (%)*	*No. of 7 independent samples correct*
Chr6:154360797-SNV	$P=0.589$	1	13.9	87	51	50.6 ± 5	0
Chr6:154451812-SNV	0.00049	1	2.9	100	29	51.5 ± 2	0
Chr8:54147491-SNV	0.000278	1	10.5	92.2	52	51.4 ± 5	4
Regression	—	6	67.7	55.7	60	61.7 ± 6	4
CART	—	12	92.6	56.1	68	74.4 ± 4	3
Clustering+*k*NN classifier	—	152	75.3	73.5	74	74.4 ± 7	5
ABC analysis+clustering+*k*NN classifier	—	34	93.7	67.4	74	80.6 ± 4	4

Abbreviations: CART, classification and regression tree; *k*NN, *k*-nearest neighbor; PPV, positive predictive value; SNV, single-nucleotide variation. χ^2 test: the α-corrected significance level was 0.000649351.

subsymbolic/symbolic concept in machine learning we define a subsymbolic biomarker as a pattern of markers, here represented by the genetic changes, where (1) none of the markers needs to be directly related to any known biological function/process or a cellular component, although this is not excluded; and (2) the sequence of marker changes is not important, that is, the marker does not retrieve its classificatory value by its place in a sequence such as a nucleic or amino acid sequence in genes or proteins. In contrast, a symbolic biomarker would represent a genetic change that can be directly related to a biological function/process or a cellular component. However, if the pattern of the subsymbolic biomarkers is observed in an organism, it can be connected to biological function processes or a cellular component. This points at emergent properties in genetics.[10]

A subsymbolic high-dimensional classifying biomarker uses subsymbolic biomarkers and subsymbolic classifiers in order to decide between different biological conditions/classes/clusters. It is called self-learning, if successfully classified data are added to the knowledge base of the classifier. This can be called a 'self-learning subsymbolic high-dimensional classifying biomarker' (Figure 4). The present information from 30+28 patients can be successively enlarged by adding further patients. In the case that a new opioid receptor genotype pattern will be correctly identified by comparison with the present information from these 58 patients, the biomarker would be successful. In the opposite case, a wrongly classified genotype pattern can be implemented into the present data basis and new ABC marker sets. Jaccard distances can be computed to improve further diagnoses by a continuously developing system. Thus, the biomarker can 'learn'.

One of the keys to the success of the classification was the accounting for the heterogeneity of the patients. Beyond the main phenotype of high opioid dosage, the present patients were indeed a heterogeneous group. They had been submitted to our pharmacogenetic counseling mostly without further details of the clinical background except for a few cases such as those presented. Therefore, patients have different diseases underlying the pain. The high opioid doses may be accidental as most patients were sent from University tertiary care centers where the physicians were more inclined to raise the opioid doses, whereas in the periphery, the same patients might have been labeled as

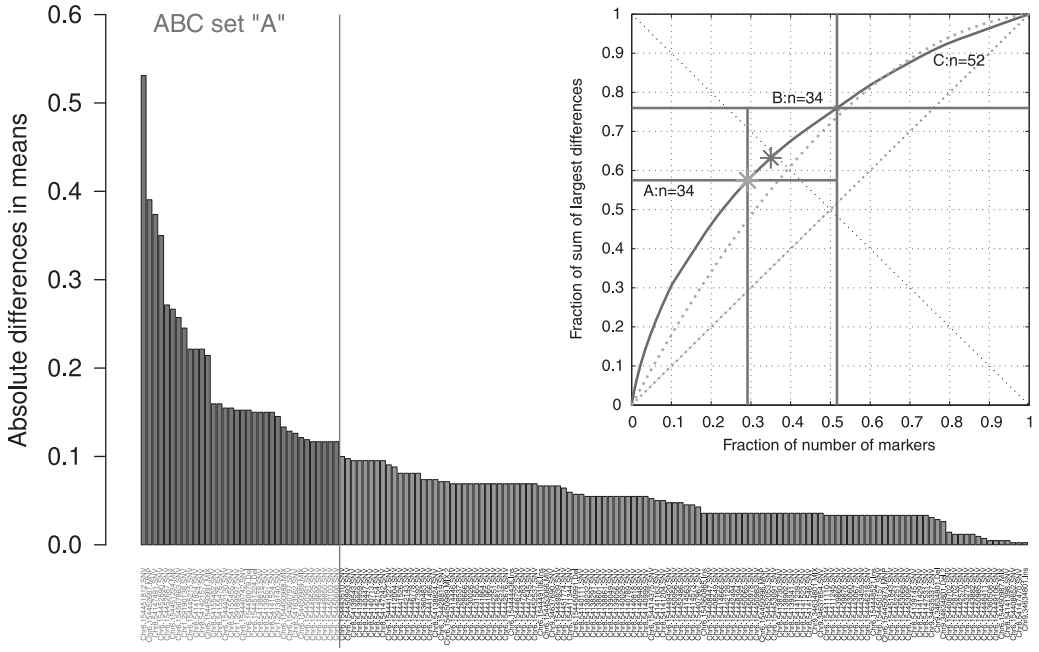


Figure 3. Identification of those genetic variants that promised to provide the comparatively best distinction between the two patient groups. The bar plot shows the absolute group size standardized differences of the occurrences of each variant between the two patients groups. The 34 bars left of the solid vertical line indicate the genetic variants found by a computed ABC analysis (right upper part of the figure) to provide a statistically valid set of markers to be included in a complex biomarker. At the upper right corner is the ABC plot[14] of the cumulative distribution function of the additional fraction of the absolute group differences in allelic numbers (solid curved line). In addition, ABC plots of the identity distribution, x_i = constant (dotted line, that is, each genetic variant occurs at the same number in each group), and of the uniform distribution in the data range (dotted line, that is, each variant allele had the same chance to occur) are shown. Compared with the latter distributions, the solid curved line clearly indicates a highly unequal distribution of the group differences in the numbers of variant alleles. Further marks in this plot consist of a light grey and a darker grey star denoting the so-called Pareto and *BreakEven* points, respectively. The Pareto point *A* (A_x, A_y) is the point at the smallest distance (left oblique black line) to the ideal point at xy where the effort would be zero to obtain the whole yield. The *BreakEven* point *B (Bx, By)* marks the point on the ABC curve where its slope, *dY/dE*, equals 1, that is, the so-called profit gain *dABC* equals 1. Beyond that point, more information can only be gained with inadequately high efforts. The ABC analysis comes from economical informatics and, in the present context, aims at identifying the most informative genetic variant for group classification by dividing the 152 variants into 3 distinct subsets. Set A should contain the 'critical few', that is, those elements that allow obtaining a maximum of yield with a minimal effort.[14] Set B comprises those elements where an increase in effort is proportional to the increase in yield. In contrast, set C contains the 'trivial many', that is, those elements that are not worth to be considered a biomarker. As a result, set A was used to establish a subsymbolic classifying biomarker.

opioid resistant already at doses below 400 mg OME per day and therefore not included into this analysis. An opioid resistance might indeed apply to one of the presented patients as discussed below. Thus, the heterogeneity of opioid-treated pain patients is a clinical reality and has to be considered in a successful biomarker. Here, a self-learning biomarker provides an ideal basis.

However, although the subsymbolic biomarker successfully classified the patients, the question remained of how genetic changes of no singular identifiable major molecular consequence can nevertheless underlie the phenotype. A possible explanation may be the above-mentioned emergent properties in genetics[10] to which the biomarker pointed. Specifically, we hypothesize that from the accumulation of several genetic variants that mostly provide only small quantitative modulations of gene transcription, a qualitative genetic change can emerge toward a substantially more or less efficiently transcribed gene that then gains the phenotypic consequences that were not produced by any single variant alone.

Indeed, molecular knowledge about members of the 'A' set of the ABC analysis (Figure 3) supports this hypothesis, and this will be illustrated by two examples. The Chr6:154451812-SNV, and various further variants that occurred only in patients receiving high analgesic opioid doses (not shown), is noncoding; yet, it is known that such variants can affect mRNA splicing, stability and structure, resulting in a reduced transcriptional efficiency.[16,17] Indeed, genome-wide association studies yielded that besides mutations in the coding regions of genes, even mutations in noncoding and intergenic regions can be associated with diseases.[18] These changes can affect the function of proteins, change the cellular response to therapeutic targets and can explain the different responses of individual patients to medications.[19] Furthermore, the Chr8:54147491-SNV variant (c.438G>T) changes the codon ATA, which translates to isoleucine, to codon ATC, which also translates to isoleucine. Although apparently this mutation does not lead to an alteration in the primary polypeptide sequence, that is, it is synonymous, this kind of mutation is now widely acknowledged to be able to cause changes in protein expression, conformation and function.[20] Recent genetic and biomedical studies have identified multiple mechanisms that provide an indication of the means by which synonymous mutations can affect physiological changes and consequently influence disease.[21]

A synonymous substitution might produce changes in the phenotype by affecting splicing that in consequence can lead to a shortened mRNA and subsequently a nonfunctional protein.[21,22] Moreover, synonymous mutations might alter the mRNA stability, leading to faster degradation and reduced protein expression.[23,24] Another possibility of how synonymous mutations may influence protein levels is affecting translation elongation. The relative synonymous codon usage (RSCU) value represents the local translation elongation rates and is directly linked with protein expression, as the speed of translation is often important for accurate protein folding. For example, a negative ΔRSCU value

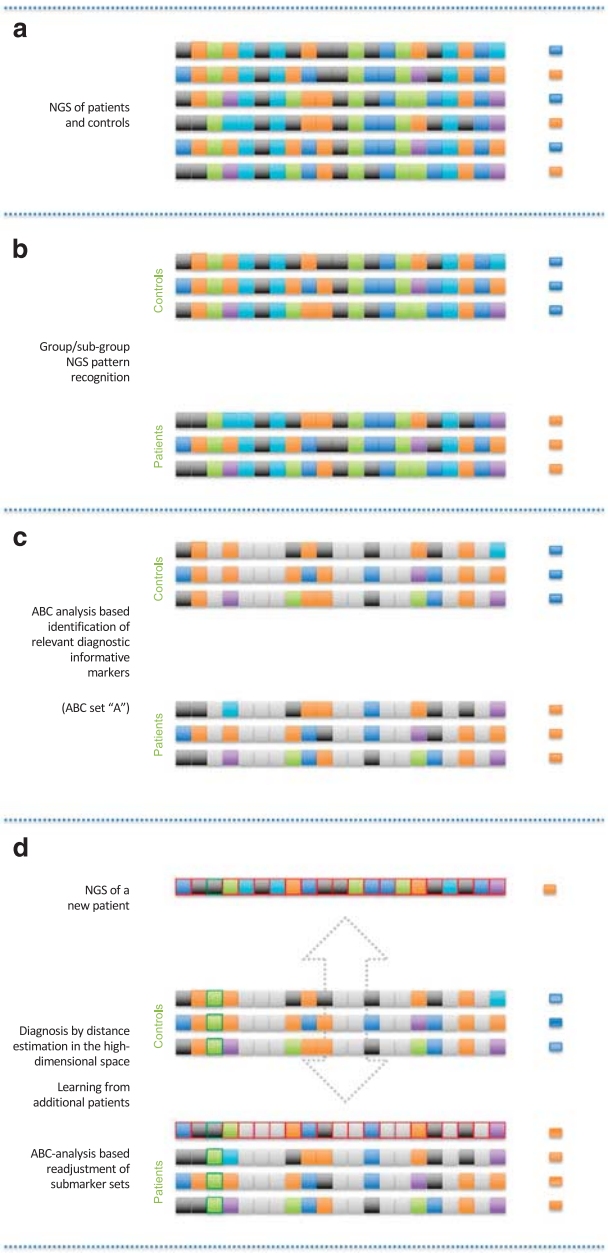


Figure 4. Schematic representation of main features of a prediction using the proposed 'high-dimensional subsymbolic biomarker' (from top to bottom). (**a**) The biomarker creation starts from the analysis of the relevant single markers, gene sequences, symbolized as arbitrarily colored squares (for example, let this be nucleotides and their heterozygous or homozygous presences). In the present analysis, let these squares denote the variants in the opioid receptor genes. (**b**) Based on the clinical background, the complex markers were grouped into either high-opioid doses demanding patients (orange small squares at the right) or controls (blue small squares at the right). Subgroups within the groups are possible and should be addressed by clustering (not shown). (**c**) From these composed makers, ABC analysis identifies the relevant submarkers that are most informative for the prediction (noninformative submarkers grayed out). (**d**) A new patient is analyzed (markers with red margins). Based on distance measuring in the high-dimensional space, this patient will be assigned to the most similar group, that is, to its nearest neighbors in the high-dimensional space based on the chosen distance measure (here, the Jaccard distance was used for genetic markers being either 0, 1 or 2). Concomitantly a learning process of the biomarker starts. In the present example, it proves useful, in a new ABC analysis, to include a further marker that, in a previous ABC analysis, had been found uninformative (third marker from the left, green margins). The process is repeated (dotted arrow) with each prediction where the disease background is known, that is, the biomarker 'learns' in the sense of Artificial Intelligence, where valid information about a patient's background is used to improve the marker. This will improve its predictive accuracy in cases where the background is not known.

(when a mutation introduces a rarer codon) might lead to a slower rate of translation elongation compared with the wild type.[21] The c.438G>T variant causes, as a consequence of codon modification, a change in the RSCU value (the RSCU value changes from 1.41 to 0.51 with $\Delta RSCU = -0.9$)[25] as queried from the Codon Usage Database at www.kazusa.or.jp/codon/. A negative $\Delta RSCU$ value means that the mutation introduces a rarer codon and this might be associated with a slower rate of translation elongation compared with the wild type.[21]

Taking this idea one step further, multiple mutations that affect codon usage thus might increase or decrease translation rates and therefore appreciably bias protein expression. In its consequence, the opioidergic system might have been altered to a degree that translated into a clinical phenotype. Specifically, the present subsymbolic classifying biomarker included 21 *OPRM1* variants, 9 *OPRK1* variants and 3 *OPRD1* variants. All coded receptors are involved in the endogenous nocifensive system. Hence, a pharmacological interpretation of the present results may involve the hypothesis of a general weakening of the endogenous opioid system, mainly but not exclusively μ-opioidergic, that was caused by the accumulation of genetic variants. This may have finally rendered the patient as nearly 'opioid resistant', directly because of a reduced activity of exogenous μ-opioid agonists and indirectly also owing to a disturbed opioidergic system. Alternatively, as the subsymbolic nature of the biomarker involves that its classification performance cannot be definitely assigned to one of its components, it could turn out that the main contributors to the successful classification are variants in the *OPRM1* gene. This is also well within the possibility of a self-learning classifier that, as explained above, by successively re-evaluating and modifying its composition, may still evolve toward a *OPRM1* variant-based classifier. In that case, the association with a reduced activity of mainly μ-opioid receptor agonists would become straightforward.

This analysis showed that opioid receptor genotyping, consistent with biological plausibility, has the potential to provide the desired predictively of particular (clinical) phenotypes as demonstrated with high opioid dose demands in pain patients. This may be obtained with a novel type of biomarker, called 'self-learning subsymbolic high-dimensional classifying biomarker'. A self-learning biomarker offers an ideal basis to account for the heterogeneity of patients as it uses flexible patterns for the prediction. This is a novel concept of biomarkers and it is not limited to the present clinical diagnostic tasks. Related to this concept, results of the analysis can also be taken as evidence for emergent properties in genetics.[10]

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGMENTS

This work has been funded by the European Union Seventh Framework Programme (FP7/2013) under Grant Agreement No. 602919 (JL, GLORIA). An additional contribution, in particular the funding of the next-generation genotyping equipment, has been received from the Landesoffensive zur Entwicklung wissenschaftlich-ökonomischer Exzellenz (LOEWE), Schwerpunkt: Anwendungsorientierte Arzneimittelforschung (JL). CS was supported by EU FP7-HEALTH-2013-INNOVATION-1 No. 602891-2.

REFERENCES

1 Lötsch J, Geisslinger G. Relevance of frequent mu-opioid receptor polymorphisms for opioid activity in healthy volunteers. *Pharmacogenomics J* 2006; **6**: 200–210.
2 Mardis ER. The impact of next-generation sequencing technology on genetics. *Trends Genet* 2008; **24**: 133–141.
3 Nielsen R, Paul JS, Albrechtsen A, Song YS. Genotype and SNP calling from next-generation sequencing data. *Nat Rev Genet* 2011; **12**: 443–451.
4 Pabinger S, Dander A, Fischer M, Snajder R, Sperk M, Efremova M *et al.* A survey of tools for variant analysis of next-generation genome sequencing data. *Brief Bioinform* 2014; **15**: 256–278.
5 Sarovich DS, Price EP. SPANDx: a genomics pipeline for comparative analysis of large haploid whole genome re-sequencing datasets. *BMC Res Notes* 2014; **7**: 618.
6 Stoddard JL, Niemela JE, Fleisher TA, Rosenzweig SD. Targeted NGS: a cost-effective approach to molecular diagnosis of PIDs. *Front Immunol* 2014; **5**: 531.
7 Galindo-González L, Pinzón-Latorre D, Bergen EA, Jensen DC, Deyholos MK. Ion Torrent sequencing as a tool for mutation discovery in the flax (Linum usitatissimum L.) genome. *Plant Methods* 2015; **11**: 19.
8 Glotov AS, Kazakov SV, Zhukova EA, Alexandrov AV, Glotov OS, Pakin VS *et al.* Targeted next-generation sequencing (NGS) of nine candidate genes with custom AmpliSeq in patients and a cardiomyopathy risk group. *Clin Chim Acta* 2015; **446**: 132–140.
9 Tenedini E, Artuso L, Bernardis I, Artusi V, Percesepe A, De Rosa L *et al.* Amplicon-based NGS: an effective approach for the molecular diagnosis of Epidermolysis Bullosa. *Br J Dermatol* 2015; **173**: 731–738.
10 Ultsch A (ed). Emergence in Self-Organizing Feature Maps In: *International Workshop on Self-Organizing Maps (WSOM '07)*. Neuroinformatics Group: Bielefeld, Germany, 2007.
11 Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B *et al.* The structure of haplotype blocks in the human genome. *Science* 2002; **296**: 2225–2229.
12 Daly MJ, Rioux JD, Schaffner SF, Hudson TJ, Lander ES. High-resolution haplotype structure in the human genome. *Nat Genet* 2001; **29**: 229–232.
13 Cover TM, Hart PE. Nearest neighbor pattern classification. *IEEE Trans Inform Theory* 1967; **13**: 21–27.
14 Ultsch A, Lotsch J. Computed ABC analysis for rational selection of most informative variables in multivariate data. *PLoS One* 2015; **10**: e0129767.
15 Altman NS. An introduction to kernel and nearest-neighbor nonparametric regression. *Am Stat* 1992; **46**: 175–185.
16 Fung KL, Gottesman MM. A synonymous polymorphism in a common MDR1 (ABCB1) haplotype shapes protein function. *Biochim Biophys Acta* 2009; **1794**: 860–871.
17 Fung KL, Pan J, Ohnuma S, Lund PE, Pixley JN, Kimchi-Sarfaty C *et al.* MDR1 synonymous polymorphisms alter transporter specificity and protein stability in a stable epithelial monolayer. *Cancer Res* 2014; **74**: 598–608.
18 Treutlein J, Cichon S, Ridinger M, Wodarz N, Soyka M, Zill P *et al.* Genome-wide association study of alcohol dependence. *Arch Gen Psychiatry* 2009; **66**: 773–784.
19 Glinskii AB, Ma J, Ma S, Grant D, Lim CU, Sell S *et al.* Identification of intergenic trans-regulatory RNAs containing a disease-linked SNP sequence and targeting cell cycle progression/differentiation pathways in multiple common human disorders. *Cell Cycle* 2009; **8**: 3925–3942.
20 Hunt R, Sauna ZE, Ambudkar SV, Gottesman MM, Kimchi-Sarfaty C. Silent (synonymous) SNPs: should we care about them? *Methods Mol Biol* 2009; **578**: 23–39.
21 Sauna ZE, Kimchi-Sarfaty C. Understanding the contribution of synonymous mutations to human disease. *Nat Rev Genet* 2011; **12**: 683–691.
22 Wang GS, Cooper TA. Splicing in disease: disruption of the splicing code and the decoding machinery. *Nat Rev Genet* 2007; **8**: 749–761.
23 Duan J, Wainwright MS, Comeron JM, Saitou N, Sanders AR, Gelernter J *et al.* Synonymous mutations in the human dopamine receptor D2 (DRD2) affect mRNA stability and synthesis of the receptor. *Hum Mol Genet* 2003; **12**: 205–216.
24 Nackley AG, Shabalina SA, Tchivileva IE, Satterfield K, Korchynskyi O, Makarov SS *et al.* Human catechol-O-methyltransferase haplotypes modulate protein expression by altering mRNA secondary structure. *Science* 2006; **314**: 1930–1933.
25 Sharp PM, Tuohy TM, Mosurski KR. Codon usage in yeast: cluster analysis clearly differentiates highly and lowly expressed genes. *Nucleic Acids Res* 1986; **14**: 5125–5143.
26 Clarke TK, Crist RC, Kampman KM, Dackis CA, Pettinati HM, O'Brien CP *et al.* Low frequency genetic variants in the mu-opioid receptor (OPRM1) affect risk for addiction to heroin and cocaine. *Neurosci Lett* 2013; **542**: 71–75.
27 Shabalina SA, Zaykin DV, Gris P, Ogurtsov AY, Gauthier J, Shibata K *et al.* Expansion of the human mu-opioid receptor gene architecture: novel functional variants. *Hum Mol Genet* 2009; **18**: 1037–1051.
28 Lötsch J, Skarke C, Grösch S, Darimont J, Schmidt H, Geisslinger G. The polymorphism A118G of the human mu-opioid receptor gene decreases the clinical activity of morphine-6-glucuronide but not that of morphine. *Pharmacogenetics* 2002; **12**: 3–9.
29 Bond C, LaForge KS, Tian M, Melia D, Zhang S, Borg L *et al.* Single-nucleotide polymorphism in the human mu opioid receptor gene alters beta-endorphin binding and activity: possible implications for opiate addiction. *Proc Natl Acad Sci USA* 1998; **95**: 9608–9613.
30 Wang S-C, Tsou H-H, Chen C-H, Chen Y-T, Ho I-K, Hsiao C-F *et al.* Genetic polymorphisms in the opioid receptor mu1 gene are associated with changes in libido and insomnia in methadone maintenance patients. *Eur Neuropsychopharmacol* 2012; **22**: 695–703.
31 Lettre G, Palmer CD, Young T, Ejebe KG, Allayee H, Benjamin EJ *et al.* Genome-wide association study of coronary heart disease and its risk factors in 8,090 African Americans: the NHLBI CARe Project. *PLoS Genet* 2011; **7**: e1001300.
32 Garriock HA, Tanowitz M, Kraft JB, Dang VC, Peters EJ, Jenkins GD *et al.* Association of mu-opioid receptor variants and response to citalopram treatment in major depressive disorder. *Am J Psychiatry* 2010; **167**: 565–573.
33 Doehring A, Küsener N, Fluhr K, Neddermeyer TJ, Schneider G, Lötsch J. Effect sizes in experimental pain produced by gender, genetic variants and sensitization procedures. *PLoS One* 2011; **6**: e17724.
34 Levran O, Londono D, O'Hara K, Nielsen DA, Peles E, Rotrosen J *et al.* Genetic susceptibility to heroin addiction: a candidate gene association study. *Genes Brain Behav* 2008; **7**: 720–729.
35 Wen S, Wang C, Berg A, Li Y, Chang MM, Fillingim RB *et al.* Modeling genetic imprinting effects of DNA sequences with multilocus polymorphism data. *Algorithms Mol Biol* 2009; **4**: 11.
36 Kishi T, Yoshimura R, Okochi T, Fukuo Y, Kitajima T, Okumura T *et al.* Association analysis of SIGMAR1 with major depressive disorder and SSRI response. *Neuropharmacology* 2010; **58**: 1168–1173.

Supplementary Information accompanies the paper on the The Pharmacogenomics Journal website (http://www.nature.com/tpj)

Erklärung zu den Autorenanteilen an der Publikation:

Next-generation sequencing of the human TRPV1 gene and the regulating co-players LTB4R and LTB4R2 based on a custom AmpliSeq™ panel (*printed*)

Name der Zeitschrift: PLOS ONE

Beteiligte Autoren: D Kringel, M Sisignano, S Zinn und J Lötsch

Was hat der Promovierende bzw. was haben die Koautoren beigetragen?

(1) zu Entwicklung und Planung

Promovierender: 50%
Autor JL: 50%

(2) zur Durchführung der einzelnen Untersuchungen und Experimente

Promovierender: 100% (DNA Extraktion, NGS-Panel Design, Sequenzierung, Varianten Annotation, Methoden Validierung)

(3) zur Erstellung der Datensammlung und Abbildungen

Promovierender: 70% (Datenerhebung, Run-Statistik, Erstellung der Abbildung Chip-Beladung und Abbildung Validierung)
Autor JL: 30% (Erstellung der Abbildung Genetische Vektoren)

(4) zur Analyse und Interpretation der Daten

Promovierender: 60% (Datenvorverarbeitung, Charakterisierung und Assoziation der Varianten, Dateninterpretation)
Autor JL: 40% (Datenvisualisierung, Datenanalyse, Dateninterpretation)

(5) zum Verfassen des Manuskripts

Promovierender: 45%
Autor MS: 5%
Autor SZ : 5%
Autor JL: 45%

Zustimmende Bestätigungen der oben genannten Angaben:

_____	_____
Datum/Ort	Unterschrift Promovend
_____	_____
Datum/Ort	Unterschrift Betreuer

PLOS ONE

RESEARCH ARTICLE

Next-generation sequencing of the human *TRPV1* gene and the regulating co-players *LTB4R* and *LTB4R2* based on a custom AmpliSeq™ panel

Dario Kringel[1], Marco Sisignano[1], Sebastian Zinn[1], Jörn Lötsch[1,2]*

1 Institute of Clinical Pharmacology, Goethe - University, Frankfurt am Main, Germany, **2** Fraunhofer Institute of Molecular Biology and Applied Ecology - Project Group Translational Medicine and Pharmacology (IME-TMP), Frankfurt am Main, Germany

* j.loetsch@em.uni-frankfurt.de

Citation: Kringel D, Sisignano M, Zinn S, Lötsch J (2017) Next-generation sequencing of the human *TRPV1* gene and the regulating co-players *LTB4R* and *LTB4R2* based on a custom AmpliSeq™ panel. PLoS ONE 12(6): e0180116. https://doi.org/10.1371/journal.pone.0180116

Editor: Sidney Arthur Simon, Duke University School of Medicine, UNITED STATES

Received: March 25, 2017

Accepted: June 11, 2017

Published: June 28, 2017

Data Availability Statement: Data can be accessed at the BioProject database. Specific accession numbers and URLs are included in Supporting Information files S1 Table, S2 Table, and S3 Table.

Funding: This work has been funded by the European Union Seventh Framework Programme (FP7/2007 - 2013) under grant agreement no. 602919 ("GLORIA", JL). Support of the laboratory equipment was gained from the Landesoffensive zur Entwicklung wissenschaftlich-ökonomischer Exzellenz (LOEWE), LOEWE-Zentrum für

Abstract

Background

Transient receptor potential cation channel subfamily V member 1 (TRPV1) are sensitive to heat, capsaicin, pungent chemicals and other noxious stimuli. They play important roles in the pain pathway where in concert with proinflammatory factors such as leukotrienes they mediate sensitization and hyperalgesia. TRPV1 is the target of several novel analgesics drugs under development and therefore, *TRPV1* genetic variants might represent promising candidates for pharmacogenetic modulators of drug effects.

Methods

A next-generation sequencing (NGS) panel was created for the human *TRPV1* gene and in addition, for the leukotriene receptors BLT1 and BLT2 recently described to modulate TRPV1 mediated sensitisation processes rendering the coding genes *LTB4R* and *LTB4R2* important co-players in pharmacogenetic approaches involving *TRPV1*. The NGS workflow was based on a custom AmpliSeq™ panel and designed for sequencing of human genes on an Ion PGM™ Sequencer. A cohort of 80 healthy subjects of Western European descent was screened to evaluate and validate the detection of exomic sequences of the coding genes with 25 base pair exon padding.

Results

The amplicons covered approximately 97% of the target sequence. A median of 2.81 x 10^6 reads per run was obtained. This identified approximately 140 chromosome loci where nucleotides deviated from the reference sequence GRCh37 hg19 comprising the three genes *TRPV1*, *LTB4R* and *LTB4R2*. Correspondence between NGS and Sanger derived nucleotide sequences was 100%.

Translationale Medizin und Pharmakologie (JL) and personnel support was received from the Else Kröner-Fresenius Foundation (EKFS), Research Training Group Translational Research Innovation – Pharma (TRIP, JL). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no further conflicts of interest exist.

Conclusions

Results suggested that the NGS approach based on AmpliSeq™ libraries and Ion Personal Genome Machine (PGM) sequencing is a highly efficient mutation detection method. It is suitable for large-scale sequencing of *TRPV1* and functionally related genes. The method adds a large amount of genetic information as a basis for complete analysis of TRPV1 ion channel genetics and its functional consequences.

Introduction

The transient receptor potential (TRP) family comprises several non-selective cation channels [1] enabling or inhibiting the transmembrane transport of several ions. Various members of this ion channel family are expressed at nociceptors and via their excitation by chemical, thermal or mechanical stimuli involved in the perception of pain [2]. This makes them primary candidates for the discovery of novel analgesic drugs [3]. A query of the Thomson Reuters "Drugs and Biologics Search Tool" (http://integrity.thomsonpharma.com) in June 2016 indicated that by far the most frequently regarded TRP member in analgesic drug development is TRP cation channel, subfamily V, member 1 (TRPV1 [4]) for which more than 200 agonists or antagonists are currently under development, which bases on the concept that endogenous agonists or sensitizers acting on TRPV1 provide a major contribution to pathophysiological pain conditions [5, 6]. The pharmacological modulation of this mechanism employs (i) the approach of direct antagonism of the TRPV1 ion channel, (ii) the exposure to agonists such as capsaicin that initially activates TRPV1 but upon prolonged exposure induces a deactivation via a calcineurin-dependent channel dephosphorylation and desensitization [7] and (iii) to prevent a sensitization and hyperactivation of the TRPV1 channel [8].

Given the importance of TRPV1 in pain and analgesic drug discovery and development, *TRPV1* genetics move into a focus of pharmacogenetic interest. A modulation of the effects of TRPV1 targeting analgesics is supported by observations that intronic *TRPV1* variants were associated with insensitivity to capsaicin [9] while the coding *TRPV1* variant rs8065080 was associated with altered responses to experimentally induced pain [10]. Moreover, gain-of-function mutations in TRPV1 have been associated with increased pain sensitivity [11], for which TRPV1 antagonists would enable a specific pharmacogenetics-based personalized cure. Hence, genetic variation of human *TRPV1* is in a focus of pain and analgesic research. With the broader availability of next generation sequencing (NGS) [12], a limitation to already investigated variants has fallen in favor of unrestricted access to the whole genetic information in agreement with the wider acceptance of whole genomic information as a valuable method in clinical research [13].

In this report, the evaluation of a new NGS method based on a custom AmpliSeq™ library and Ion Torrent sequencing for the fast detection of genetic variations in the human *TRPV1* gene is described. However, preclinical evidence indicates that leukotriene B4 mediates the inflammation via TRPV1 [14] and that the nociceptive function of TRPV1 is modulated by the activation of leukotriene receptors BLT1 and BLT2 [8] that are highly expressed in TRPV1 expressing dorsal root ganglion neurons. Both receptors form an antagonistic sensitizing system and have opposing roles in TRPV1 sensitisation. This renders them important co-players in pharmacogenetic approaches at analgesics aiming at modulation of the function of TRPV1. To provide a comprehensive basis for pharmacogenetic assessments of TRPV1 modulators,

the present NGS panel was extended with human *LTB4R* and *LTB4R2* genes that code for the leukotriene receptors of present interest.

Methods

DNA template preparation and amplification

The investigation followed the Declaration of Helsinki on Biomedical Research Involving Human Subjects and was approved by the Ethics Committee of the Medical Faculty of the Goethe-University, Frankfurt, Germany. All participating subjects had provided informed written consent. Genomic DNA was available from venous blood samples drawn from a random sample of 80 healthy volunteers of Western European descent according to self-assignment. DNA was extracted from 200 μl blood on a BioRobot EZ1 workstation applying the blood and body fluid spin protocol provided in the EZ1 DNA Blood 200 μl Kit (Qiagen, Hilden, Germany).

Exomic genotyping was performed for the *TRPV1* gene (NCBI ID 7442), located on chromosome 17 and encoding for the TRPV1 ion channel and for the *LTB4R* and *LTB4R2* genes (NCBI IDs 1241 and 56413), both located on chromosomes 14 and encoding for leukotriene B4 receptors BLT1 and BLT2. A multiplex PCR amplification strategy for the coding genes sequences was accomplished online (Ion Ampliseq™ Designer; http://www.ampliseq.com) to amplify the target region specified above (for primer sequences, see S1 Table) with 25 base pair exon padding. After comparison of several primer design options, the design providing the maximum target sequence coverage was chosen. The ordered amplicons covered 97.02% of the target sequence. A total of 10 ng DNA per sample were used for the target enrichment by a multiplex PCR and each DNA pool was amplified with the Ion Ampliseq™ Library Kit in conjunction with the Ion Ampliseq™ "custom Primer Pool"—protocols according to the manufacturer procedures (Life Technologies, Darmstadt, Germany).

After each pool had undergone 17 PCR cycles, the PCR primers were removed with FuPa Reagent and the amplicons were ligated to the sequencing adapters with short stretches of index sequences (barcodes) that enabled sample multiplexing for subsequent steps (Ion Xpress™ Barcode Adapters Kit; Life Technologies). After purification with AMPure XP beads (Beckman Coulter, Krefeld, Germany), the barcoded libraries were quantified with a Qubit® 2.0 Fluorimeter (Life Technologies, Darmstadt, Germany) and normalized for DNA concentration to a final concentration of 20 pmol/L using the Ion Library Equalizer™ Kit (Life Technologies, Darmstadt, Germany). Equalized barcoded libraries from 11 to 40 samples at a time were pooled. To clonally amplify the library DNA onto the Ion Sphere Particles (ISPs; Life Technologies, Darmstadt, Germany), the library pool was subjected to emulsion PCR by using an IT OneTouch template kit on an IT OneTouch system (Life Technologies, Darmstadt, Germany) following the manufacturer's protocol.

Sequencing

Enriched ISPs which carried many copies of the same DNA fragment were subjected to sequencing on an Ion 318 Chip to sequence pooled libraries with eleven to twelve samples. The 318 chip was chosen (instead of the low-capacity 314 or the middle-capacity 316 chip) to obtain a high sequencing depth of coverage which was averagely of 50x which means that, each base has been sequenced 50 times, when 40 samples were loaded. Sequencing was performed using the sequencing kit (Ion PGM 200 Sequencing Kit; Life Technologies, Darmstadt, Germany) as per the manufacturer's instructions with the 200-bp single-end run configuration.

Bioinformatics generation of sequence information

The raw data (unmapped BAM-files) from the sequencing runs were processed using Torrent Suite Software (Version 4.4.2, Life Technologies, Darmstadt, Germany) to generate read alignments which are filtered by the software into mapped BAM-files using the reference genomic sequence (hg19) of target genes. Variant calling was performed with the Torrent Variant Caller Plugin using as key parameters: minimum quality = 10, minimum coverage = 20, and minimum coverage on either strand = 3. The annotation of called variants was done using the Ion Reporter Software (Version 5.0; Life Technologies, Darmstadt, Germany) and the variant classification tool of the SNP and Variation Suite software (Version 8.4.4; Golden Helix, Bozeman, MT, USA) for the VCF (variant call format) files that contained the nucleotide reads and the GenomeBrowse® software (Version 2.0.4, Golden Helix, Bozeman, MT, USA) to map the sequences to the reference sequences GRCh37 g1k (dated February 2009).

On the basis of the observed allelic frequency, the expected number of homozygous and heterozygous carriers of the respective SNP (single nucleotide polymorphism) was calculated using the Hardy-Weinberg equation. Indicating that the study sample corresponded to a random sample of subjects, Fisher's exact test [15] was used as proposed previously [16]. Only variants within the Hardy-Weinberg equilibrium were retained. The SNP and Variation Suite software (Version 8.4.4; Golden Helix, Bozeman, MT, USA) was used for the analysis of sequence quality, coverage and for variant identification.

Method validation

Method validation was accomplished by means of Sanger sequencing [17, 18] in an independent external laboratory (Eurofins Genomics, Ebersberg, Germany). For the detected variant type, i.e., single nucleotide polymorphisms (SNV), nucleotide insertions (Ins) and nucleotide deletions (Del), the variant with the highest frequency of the rare allele was chosen for external sequencing: 17:3493769-SNV, 17:3496181_Ins, 17:3512619_Del. In addition, the variant 17:3480447-SNV, which is the functional rs8065080 SNP previously associated with altered pain sensitivity [10], was added accommodating the present context of analgesics' pharmacogenetic. Amplification of the respective DNA segments was done using PCR primer pairs (forward, reverse) of (i) 5´-CCATGTTGCGTCTCTCGATG-3´ and 5´-CAACCCGTTATTTCCTGTTCCCA-3´ (ii) 5´-CTCAGAGGTGAGCAGGCCTAGC-3´ and 5´-AAGGCCAGGATGCTTGACAGATG-3´, (iii) 5´-AAGGCACAAGACTCTGGAAGAAT-3´ and 5´-CGAGTTTGGGAAGCAGTCGTAT-3´ and (iv) 5´-ACCCAGTGCCTTCTCATTCAG-3´ and 5´-CACGTTCTCAAGACGCATCC-3´. Results of Sanger sequencing were aligned with the genomic sequence and analyzed using Chromas Lite® (Version 2.1.1, Technelysium Pty Ltd, South Brisbane, Australia) and the GenomeBrowse® (Version 2.0.4, Golden Helix, Bozeman, MT, USA) was used to compare the sequences obtained with NGS or Sanger techniques.

Results

The NGS assay of human *TRPV1*, *LTB4R2* and *LTB4R* genes was established on 80 genomic DNA samples obtained from a random selection of healthy subjects of Caucasian ethnicity. As proposed previously [19], only exons and their boundary sequences for which read-depths > 20 for each nucleotide could be obtained were considered as successfully analyzed. Applying this criterion, complete or nearly complete coverage of the relevant sequences was obtained (Table 1; for details on missing variants, see S2 Table). The sequencing of the whole cohort required two separate runs with each 40 patients' samples. Coverage statistics (Table 1) were comparable between both runs and were in the range of accepted quality criteria [20–22]. During the runs, a median of $2.81 \cdot 10^6$ reads per run was generated. The mean depth was near

Table 1. AmpliSeq™ amplicons and coverage details of the human *LTB4R2*, *LTB4R and TRPV1* NGS assay.

Gene	Chr*	Chr start	Chr end	Amplicons	Total bases	Covered bases	Coverage	Sum (total, covered, %)
LTBR42	Chr14	34634693	34635880	8	1187	1187	1.000	3520, 3427, 98.3%
		34636968	34637111	1	143	143	1.000	
		34636968	34637134	1	166	166	1.000	
		34637238	34637442	2	204	204	1.000	
		34637191	34637442	2	251	251	1.000	
		34637578	34637848	2	270	251	0.930	
		34637518	34637848	2	330	256	0.776	
		34634693	34635880	8	1187	1187	1.000	
LTBR4	Chr14	34637238	34637442	2	204	204	1.000	5683, 5117, 97%
		34637191	34637442	2	251	251	1.000	
		34637578	34637848	2	270	251	0.930	
		34637518	34637848	2	330	256	0.776	
		34634693	34635880	8	1187	1187	1.000	
		34637238	34637442	2	204	204	1.000	
TRPV1	Chr17	24636968	24637111	1	143	143	1.000	23787, 21859, 98.8%
		24636968	24637134	1	166	166	1.000	
		24637238	24637442	2	204	204	1.000	
		24637191	24637442	2	251	251	1.000	
		24636968	24637134	1	166	166	1.000	
		24637238	24637442	2	204	204	1.000	
		24637191	24637442	2	251	251	1.000	
		24636968	24637134	1	166	166	1.000	
		24637238	24637442	2	204	204	1.000	
		24637191	24637442	2	251	251	1.000	
		24636968	24637134	1	166	166	1.000	
		24637238	24637442	2	204	204	1.000	
		24637191	24637442	2	251	251	1.000	
		24637191	24637442	2	251	251	1.000	
		24636968	24637134	2	270	251	0.930	
		24637238	24637442	2	330	256	0.776	
		24637191	24637442	8	1187	1187	1.000	
		24636968	24637134	2	251	251	1.000	
		24637238	24637442	2	270	251	0.930	
		24637191	24637442	2	330	256	0.776	
		24636968	24637134	8	1187	1187	1.000	

*: Chr: Chromosome.

https://doi.org/10.1371/journal.pone.0180116.t001

from 200 reads, the mean read length evaluated 198 bases and average chip loading was 66% (Fig 1). To ensure a high density of ISPs on a chip and hence, a high sequencing output, the chip loading value should be ≥ 60%. The observed NGS results agreed with the results obtained with conventional sequencing of random samples (Fig 2). In all validation samples, the correspondence between NGS and Sanger derived nucleotide sequences was 100%, all of the tested nucleotide variants could be verified.

Following elimination of nucleotides agreeing with the standard human genome sequence GRCh37 g1k (dated February 2009), the result of the NGS consisted of a vector of nucleotide

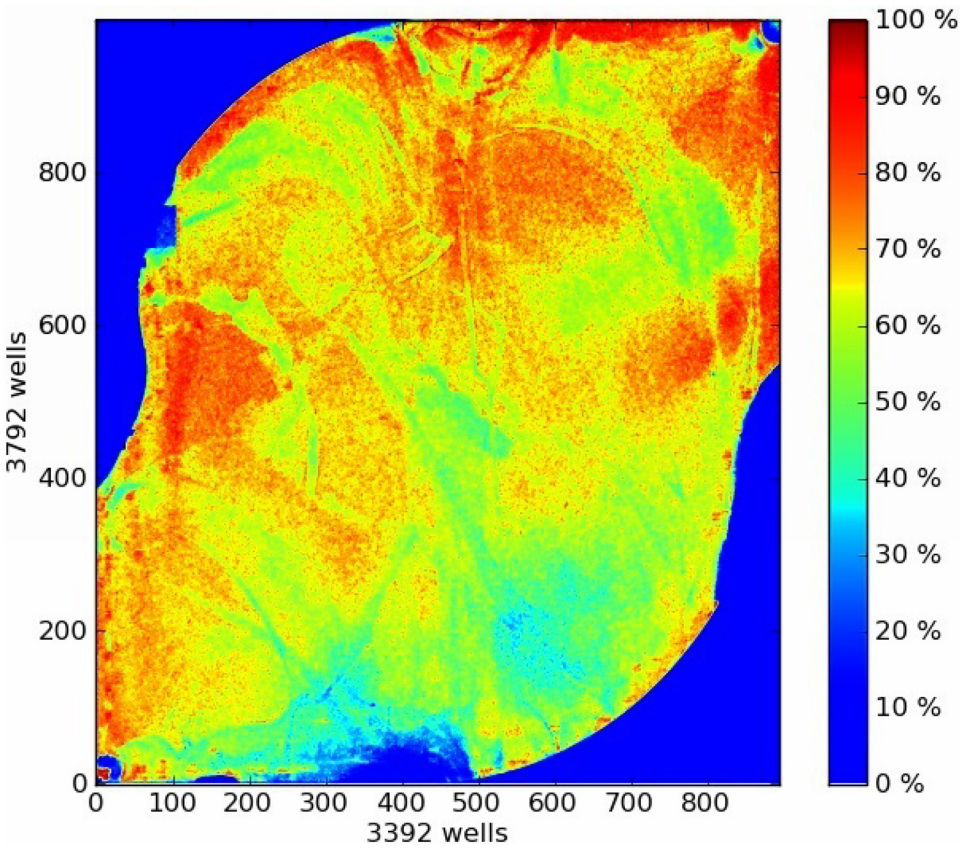


Fig 1. Pseudo-color image of the Ion 318™ v2 chip plate showing percent loading across the physical surface. This sequencing run had a 70% loading, which ensures a high ISP density. Every 318 chip contains more than 6 million wells and the color scale on the right side conduces as a loading indicator. Deep red coloration stays for a 100% loading, which means that every well in this area contains an ISP (templated and non-templated) whereas deep blue coloration implies that the wells in this area are empty.

https://doi.org/10.1371/journal.pone.0180116.g001

information about the *LTB4R2*, *LTB4R* and *TRPV1* genes for each individual DNA sample (Fig 3). This vector had a length equaling the set union of the number of chromosomal positions in which a non-reference nucleotide had been found in any probe of the actual cohort of randomly chosen healthy subjects. Specifically, a total of 156 genetic variants was found, of which 11, 28 and 117 were located in the *LTB4R2*, *LTB4R* and *TRPV1* genes, respectively (Fig 3). Of the observed variants, 38 were located in coding parts of the genes (Table 2), 56 were located in introns, 33 in the 3'-UTR, 16 in the 5'-UTR, 5 variants were assigned to both UTR's and 8 were located downstream. The nucleotidic and, if present, the resulting amino acid exchanges, of the coding variants are listed in Table 2. The allelic frequencies corresponded to those expected based on the Hardy-Weinberg equilibrium (Fisher's exact tests: p always > 0.05) and, for variants with reported clinical functional association, the observed allelic frequency was comparable to reported frequencies (Table 3). Most of the observed variants were single nucleotide polymorphisms (n = 135; 9, 25 and 101 in the *LTB4R2*, *LTB4R* and *TRPV1* genes, respectively) whereas classified as mixed polymorphisms (n = 8), nucleotide insertions (n = 6), nucleotide deletions (n = 5) or multinucleotide polymorphisms (n = 2) were more rarely found in the present cohort.

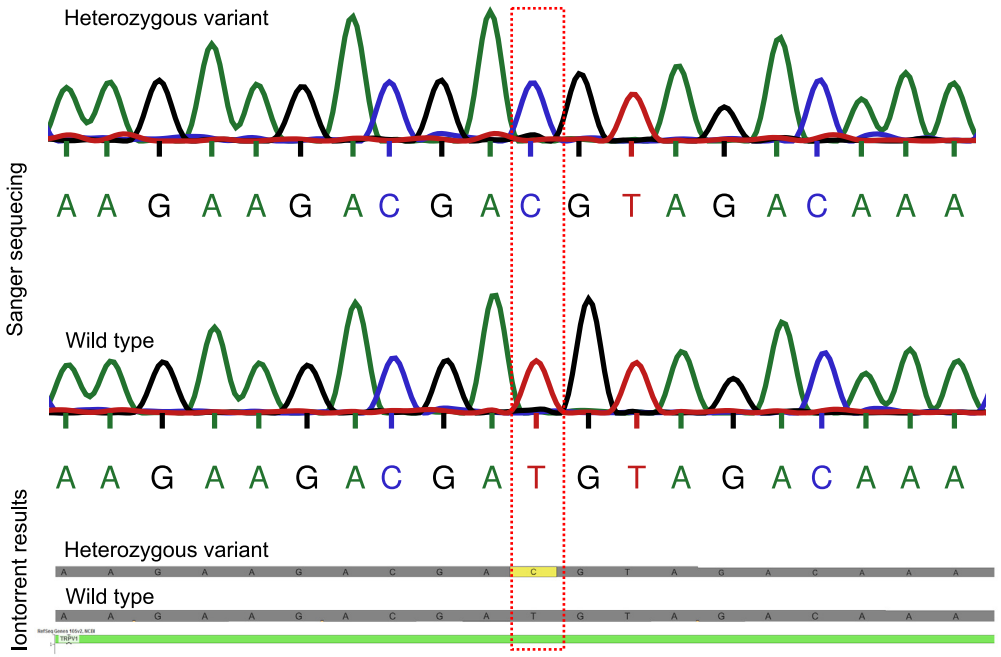


Fig 2. Alignment of the ion torrent sequence of the *TRPV1* gene illustrated by Golden Helix Genome Browse® readout versus the same sequence according to a Sanger electrophereogram. Highlighted is the coding *TRPV1* variant rs8065080 as a heterozygous mutation and a non-mutated wild type.

https://doi.org/10.1371/journal.pone.0180116.g002

Discussion

An NGS assay for the exons and regulatory parts of the human genes coding for the TRPV1 ion channel and those coding for its recently associated co-players comprising the leukotriene receptors BLT1 and BLT2 (*LTB4R*, *LTB4R2)*. The NGS assay produced valid nucleotide sequences corresponding to those obtained with the classical Sanger sequencing technique. The NGS assay is suitable for small to large-scale experimental setups aiming at accessing the information about any nucleotide in a study cohort, with a selection of those that differ from the reference nucleotide.

TRPV1 ion channels mediate pain induced by noxious heat (> 43˚C) [23]. A most striking phenotype of *Trpv1* –/– mice is a severe deficit of inflammation-induced thermal hyperalgesia [24]. In addition to heat, TRPV1 expression is largely associated with small diameter primary afferent nerve fibers, which are sensitive to various chemical excitants including protons (low pH), capsaicin, lipoxygenase, resiniferatoxin, ethanol, N-arachidonoyl-dopamine and the endogenous cannabinoid anandamide [3, 25]. Based on evidence that TRPV1 channels are necessary for the development of inflammatory hyperalgesia to thermal stimuli [24] their role in pain has been acknowledged for more than two decades [24, 26]. Currently, they are used as target of capsaicin containing analgesics. However, TRPV1 remains a primary candidate for the discovery of novel analgesic drugs [3] and approximately 200 modulators of this target are currently under development (http://integrity.thomsonpharma.com). This establishes a strong future pharmacogenetic context of *TRPV1* considering the increasing acknowledgment that the treatment of pain will benefit from individualized approaches including those based on the patient's genotype [8].

Research on the genetic variation of variants in human *TRPV1* or leukotriene receptor genes is an active topic that has already provided several clinically relevant functional associations. A

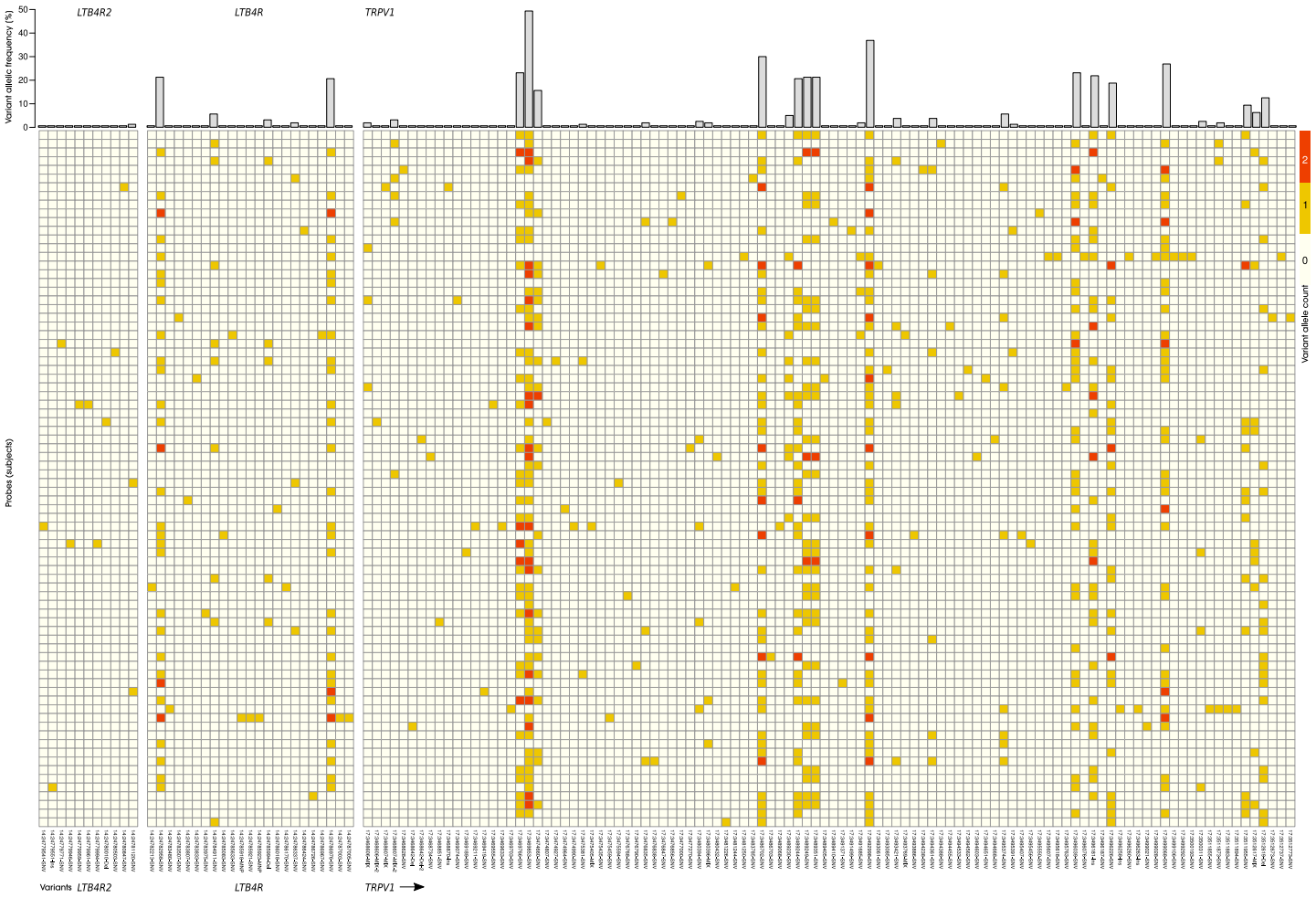


Fig 3. *LTB4R2*, *LTB4R* and *TRPV1* genetic pattern of 80 healthy volunteers of Caucasian ethnicity. The mosaic plot shows the occurrence of variants (lines) per DNA sample (columns) as vectors of a length corresponding to the number of gene loci in which a non-reference nucleotide was found in any sample of the whole cohort. The vectors are composed of information about the number of non-reference alleles found at the respective locus in the respective sample, color codes as white, 0 non-reference alleles = wild type genotype, yellow, heterozygous, and red, 2 non-reference alleles). The bar plot on the top shows the number of variant alleles found in the cohort.

https://doi.org/10.1371/journal.pone.0180116.g003

query of the 156 genetic variants in various publicly available data sources (Online Mendelian Inheritance in Man" (OMIM®) database at http://www.ncbi.nlm.nih.gov/omim, NCBI gene index database at http://www.ncbi.nlm.nih.gov/gene; GeneCards at http://www.genecards.org [27] and the "1000 Genomes Browser" at https://www.ncbi.nlm.nih.gov/variation/tools/1000genomes; all accessed in May 2017) yielded 13 clinical associations (Table 3). The clinical associations included a variety of pathologies such as pain, asthma or osteoarthritis. Specifically, variants in both, *TRPV1* and *LTB4R* have been associated with a higher susceptibility to bronchial asthma [28–32]. Moreover, *TRPV1* variants have been associated with a higher risk of type 2 diabetes [33] or of functional dyspepsia [34]. Finally, of potential importance for a pharmacogenetic modulation of the effects of future analgesics, *TRPV1* variants have been associated with altered pain phenotypes in clinical or human experimental settings [10, 35, 36]. This fits to the particular role of TRPV1 as a major target for novel analgesic drugs under development.

Winter and colleagues recently created an overview of site-directed mutagenesis studies on *Trpv1* receptor in rodents [37]. Their study summarized information about several mutated

Table 2. A list of variants found in the coding parts of the *LTB4R2, LTB4R and TRPV1* genes in a random sample of 80 healthy volunteers of Caucasian ethnicity.

Gene	Variant	Chr*	Position	Classification	Exon	Coding	Protein
LTB4R2	14:24779946-SNV	14	24779946	Nonsyn SNV	2	c.76T>C	p.Phe26Leu
	14:24779959-SNV	14	24779959	Nonsyn SNV	2	c.89C>T	p.Ala30Val
	14:24779961-SNV	14	24779961	Nonsyn SNV	2	c.91G>A	p.Ala31Thr
	14:24779994-SNV	14	24779994	Nonsyn SNV	2	c.124G>A	p.Val42Met
	14:24780010-Del	14	24780010	Frameshift Del	2	c.140_164del	p.Ala51fs
	14:24780503-SNV	14	24780503	Synonymous	2	c.633C>T	p.=
	14:24780847-SNV	14	24780847	Nonsyn SNV	2	c.977A>G	p.Glu326Gly
LTB4R	14:24784911-SNV	14	24784911	Synonymous	2	c.54T>C	p.=
	14:24785083-SNV	14	24785083	Nonsyn SNV	2	c.226C>T	p.His76Tyr
	14:24785633-SNV	14	24785633	Nonsyn SNV	2	c.776T>C	p.Val259Ala
	14:24785784-SNV	14	24785784	Synonymous	2	c.927C>T	p.=
TRPV1	17:3474927-SNV	17	3474927	Synonymous	14	c.2238C>T	p.=
	17:3475435-SNV	17	3475435	Nonsyn SNV	13	c.2212G>T	p.Asp738Tyr
	17:3475459-SNV	17	3475459	Nonsyn SNV	13	c.2188G>A	p.Gly730Arg
	17:3475490-SNV	17	3475490	Synonymous	13	c.2157G>A	p.=
	17:3476990-SNV	17	3476990	Synonymous	12	c.2040C>T	p.=
	17:3477000-SNV	17	3477000	Nonsyn SNV	12	c.2030A>G	p.Asn677Ser
	17:3480432-SNV	17	3480432	Nonsyn SNV	11	c.1768G>A	p.Gly590Arg
	17:3480447-SNV	17	3480447	Nonsyn SNV	11	c.1753A>G	p.Ile585Val
	17:3480910-SNV	17	3480910	Synonymous	10	c.1695T>C	p.=
	17:3483785-SNV	17	3483785	Nonsyn SNV	9	c.1513A>G	p.Thr505Ala
	17:3486702-SNV	17	3486702	Nonsyn SNV	8	c.1406C>T	p.Thr469Ile
	17:3486703-SNV	17	3486703	Nonsyn SNV	8	c.1405A>T	p.Thr469Ser
	17:3489068-SNV	17	3489068	Synonymous	7	c.1377T>C	p.=
	17:3491499-SNV	17	3491499	Nonsyn SNV	6	c.1207A>G	p.Ser403Gly
	17:3493200-SNV	17	3493200	Nonsyn SNV	5	c.945G>C	p.Met315Ile
	17:3494361-SNV	17	3494361	Synonymous	3	c.501C>T	p.=
	17:3494388-SNV	17	3494388	Synonymous	3	c.474T>C	p.=
	17:3494533-SNV	17	3494533	Synonymous	2	c.399G>A	p.=
	17:3494562-SNV	17	3494562	Stopgain	2	c.370C>T	p.Gln124*
	17:3494603-SNV	17	3494603	Nonsyn SNV	2	c.329T>C	p.Leu110Pro
	17:3495374-SNV	17	3495374	Nonsyn SNV	1	c.271C>T	p.Pro91Ser
	17:3495391-SNV	17	3495391	Nonsyn SNV	1	c.254A>G	p.Gln85Arg
	17:3495407-SNV	17	3495407	Nonsyn SNV	1	c.238C>T	p.Pro80Ser
	17:3495456-SNV	17	3495456	Synonymous	1	c.189C>T	p.=
	17:3495550-SNV	17	3495550	Nonsyn SNV	1	c.95G>T	p.Arg32Met
	17:3495607-SNV	17	3495607	Nonsyn SNV	1	c.38C>T	p.Ala13Val
	17:3495618-SNV	17	3495618	Nonsyn SNV	1	c.27G>C	p.Leu9Phe

*: Chr: Chromosome.

https://doi.org/10.1371/journal.pone.0180116.t002

sites along the *Trpv1*, which influenced the effect or binding of different compounds like agonists, antagonists, and channel blockers and alter the responsiveness to heat and influence the regulation of the receptor function. Of peculiar interest is the c-terminus part of the receptor, because it contained several mutations implicated in binding of capsaicin. To reference this information to our study, we took out an alignment blast with http://www.uniprot.org/blast/, which is a search tool to find regions of local similarity between sequences and can be used to

Table 3. A list of human variants of the *LTB4R and TRPV1* genes, found in the present random sample of 80 healthy volunteers of Caucasian ethnicity, for which functional associations in clinical or human experimental settings have been reported.

Gene	Variant	dbSNP# accession number	Allelic frequency [%] (CI*) Present cohort*	HAPMAP CEU	Known clinical association	Reference
LTB4R	14:24786060-SNV	rs1046587	46.2 (38.7–54)	47.4	Asthma susceptibility	[28]
	14:24786293-SNV	rs4981503	28.1 (21.7–35.4)	-	Asthma susceptibility	[29]
TRPV1	17:3469853-SNV	rs4790522	49.4 (41.7–57)	56.2	Bronchial asthma susceptibility	[30, 31, 66]
					Susceptibility to cough	[67]
					Altered pain sensitivity	[35]
	17:3480447-SNV	rs8065080	33.7 (26.9–41.4)	35.8	Altered cold pain sensitivity	[10]
					Painful knee osteoarthritis	[36]
					Altered salt taste perception	[68]
					Higher risk of type 2 diabetes	[33]
	17:3486702-SNV	rs224534	30 (23.4–37.5)	33.5	Sickle cell pain	[69]
	17:3493200-SNV	rs222747	25.6 (19.5–32.9)	18.3	Functional dyspepsia	[34]
					Ménière's disease	[70]
	17:3495391-SNV	rs55916885	1.2 (0.3–4.4)	-	Cerebellar hypoplasia	[71]

*: CI denotes 95% binomial confidence intervals of the allelic frequencies are given in parentheses after the observed frequency.

#: Database of Single Nucleotide Polymorphisms (dbSNP). Bethesda (MD): National Center for Biotechnology Information, National Library of Medicine. Available from: http://www.ncbi.nlm.nih.gov/SNP/ [72]

https://doi.org/10.1371/journal.pone.0180116.t003

infer functional and evolutionary relationships between sequences revealed that *TRPV1* is highly conserved. With the present NGS assay, several functional SNPs could be identified in the coding area of *TRPV1*; one variant (17:3477000-SNV) was located in exactly the c-terminus area mentioned above. On this basis, the impact of this variant on nociception can be prospectively studied.

A pharmacogenetic modulation of the effects of TRPV1-targeting analgesics is supported by evidence of associations of rare and of common variants in the human *TRPV1* gene with pain-related clinical phenotypes. Based upon the direction of change of each phenotype and cumulative changes in each SNP, three functional categories of *TRPV1* variants were proposed: gain of function (hTRPV1 Q85R, P91S, and T469I), loss of function (I585V), and mixed (M315I) [38]. These *in vitro* results support clinical observations of *TRPV1* genotypic effects. A Korean subject who was insensitive to capsaicin and displayed mRNA and protein expression levels of TRPV1 reduced by 50% from average subjects was found to carry seven intronic *TRPV1* single nucleotide polymorphisms (SNPs) [9]. Similarly, women carrying a coding *TRPV1* variant were found to be less sensitive to cold [10]. The association possibly involves interactions among TRP channels [39] based on evidence that TRPA1 channels are often co-expressed with heat ($> 43^{\circ}C$ [4]) gated TRPV1 [40, 41]) and the channels act in concert. TRPV1 can oligomerize with other TRP family subunits including TRPV3 and TRPA1 [42–44] and the heteromerization can affect the calcium signaling pathways of TRPA1 homomers [44]. While heat hyperalgesia was initially attributed solely to TRPV1, currently TRPA1 and TRPV1 are regarded to be regulated downstream of PLC-coupled bradykinin (BK_2) receptors [45] contributing together to hypersensitivity to heat [46]. Hence, this evidence supports a possible pharmacogenetic importance. Further evidence about functional associations of *TRPV1* gene variants has been raised in Spanish Caucasian migraine patients in whom the presence of the *TRPV1* rs222741 variant conferred a disease risk [47].

The addition of leukotriene B4 (LTB4) receptors to the *TRPV1* gene panel anticipates a possible pharmacogenetic role in TRPV1 targeting analgesics resulting from recent evidence about a co-expression of the receptors at nociceptive neurons and functional their interplay [8]. LTB4 is a potent proinflammatory agent and its signaling pathway involves two distinct G protein coupled receptors of which BLT1 is a high-affinity and BLT2 a low-affinity LTB4 receptor [48]. The interaction of LTB4 at these receptors is a contributing factor in the pathogenesis of inflammatory diseases [49]. Studies involving the targeted deletion of murine BLT1 and the effect of antagonizing LTB4 receptors in inflammatory models have highlighted the therapeutic potential of BLT receptors with regard to inflammatory diseases [49]. LTB4 has also been shown to activate the TRPV1 channel [50, 51] which leads to excitation of nociceptors and evokes pain-related behaviors [25]. While variants in the two LTB4 receptors potentially affect TRPV1 modulation based analgesic therapies, evidence about functional polymorphisms in these genes is sparse. Studies have suggested a role of polymorphisms overreaching leukotriene pathway genes in determining leukotriene production and susceptibility to allergic disorders, such as inflammatory cell chemotaxis and asthma [52]. Both receptor genes were shown to be polymorphic, in addition, *LTB4R* and *LTB4R2* show splice variations at multiple regions, however, the functional significance has yet to be determined [53].

The present NGS method is suitable for large-scale sequencing of an extended set of human genes involving the main target, *TRPV1*, and recently identified co-players, *LTB4R* and *LTB4R2*. By covering almost, the complete relevant coding and regulatory parts of these genes, the method includes all variants studied so far for functional associations and adds a large amount of genetic information as a basis for complete analysis of human TRPV1 ion channel genetics and its functional consequences. The assay aimed at the complete coding and regulatory information of the selected genes, which regards the increasing acknowledgment of the insufficiency of addressing a limited selection of published functional genetic variants in providing a satisfactory genetic diagnosis of the clinical phenotype. Research interest in the complete genomic information dates back to the seventies of last century when the selective incorporation of chain-terminating dideoxynucleotides by DNA polymerase during *in vitro* DNA replication had been introduced [17, 18]. Techniques significantly improved during the last decades with the development of contemporary machines in the late 1990s re-leased to the market around the year 2005. The term "next generation" DNA sequencing refers to high-throughput technologies capable of parallel analyzes of large numbers of different DNA sequences in a single reaction [54]. NGS has been attributed the potential to accelerate biomedical research [12, 55, 56].

Currently, two commercial NGS platforms are widely used for diagnostic purposes: the MiSeq/HiSeq/NextSeq (Illumina, Hayward, CA, USA) and the Ion Torrent PGM (Life Technologies, Carlsbad, CA, USA). Both platforms combine conceptually similar workflows, starting with the creation of the genetic sample, which commences library preparation involving fragmentation of genomic DNA, purifying to uniform and desired fragment size and ligation to sequencing adapters specific to the platform. Differences apply to the reaction biochemistry and the way how the sequencing information is read [54]. In the present ion semiconductor sequencing method, libraries are immobilized to beads and amplified in microdroplets of aqueous solution and oil using emulsion PCR. Individual nucleotide bases are incorporated via DNA polymerase, which in the case of success triggers the release of a proton. The semiconductor chip that acts as a pH meter [57] providing the final readout. Alternative techniques use the detection of light instead, i.e., from optical fluorescence signals in the case of successful nucleotide incorporation the DNA nucleotide sequence is assembled. The different techniques differ with respect to the obtained throughput and accuracy, but multiple studies have shown that both NGS platforms provide reliable sequencing results in routine clinical diagnostics

[58–61] and a recent study came up with a 100% concordance between NGS and an alternative diagnostic approach in mutant allele detection [62].

The high throughput and comprehensive information about DNA sequences are presently reflected in the assay costs. The sequencing of the *TRPV1*, *LTB4R* and *LTB4R2* receptor genes of 80 patients required € 1,500 for the AmpliSeq™ custom panel, € 5,880 for library preparation, € 980 for template preparation and € 1,400 for sequencing. In addition, approximately € 600 were spent for consumables and laboratory supplies. With 40 barcoded samples loaded on two chips, respectively, analysis costs for a single patient's gene sequence were approximately € 130. NGS costs are expected to quickly fall in near future [63]. However, despite this rapid technological progress, the analysis of the generated large data sets remains challenging [64]. As the sequencing process is only the beginning of the procedure, the analysis of the resulting "big data" requires substantial computational power, bioinformatics expertise and "up to date" databases of genomic variations. NGS technologies seem to shift the workload essentially away from the laboratory sample preparation toward various data analysis processes.

We report a NGS assay based on AmpliSeq™ libraries and Ion Personal Genome Machine (PGM) suitable for large-scale sequencing of *TRPV1* and functionally related genes. While the aim of assay development had the pharmacogenetics of TRPV1-targeting novel analgesics in mind, the roles of TRPV1 and the two LTB4 receptors are not restricted to this setting. By contrast, the expression of TRPV1 is also observed in non-neuronal sites such as the epithelium of bladder and lungs and in hair cells of the cochlea. At these sites, TRPV1 serves as a potential drug target for treating various diseases such as cystitis, asthma and hearing loss [65].

Supporting information

S1 Table. A list of PCR primer used for the NGS assay.
(DOCX)

S2 Table. A list of missed parts from the gene panel.
(DOCX)

S3 Table. The accession numbers of the original data at the BioProject database.
(DOCX)

Author Contributions

Conceptualization: DK JL MS SZ.

Data curation: DK JL.

Formal analysis: JL DK.

Funding acquisition: JL.

Investigation: DK JL.

Methodology: DK JL MS SZ.

Project administration: JL.

Resources: JL.

Supervision: JL.

Validation: DK JL MS SZ.

Visualization: JL DK.

Writing – original draft: JL DK.

Writing – review & editing: DK JL MS SZ.

References

1. Montell C. Drosophila TRP channels. Pflügers Archiv: European journal of physiology. 2005; 451:19–28. https://doi.org/10.1007/s00424-005-1426-2 PMID: 15952038
2. Clapham DE. TRP channels as cellular sensors. Nature. 2003; 426(6966):517–24. https://doi.org/10.1038/nature02196 PMID: 14654832
3. Patapoutian A, Tate S, Woolf CJ. Transient receptor potential channels: targeting pain at the source. Nat Rev Drug Discov. 2009; 8(1):55–68. https://doi.org/10.1038/nrd2757 PMID: 19116627
4. Caterina MJ, Schumacher MA, Tominaga M, Rosen TA, Levine JD, Julius D. The capsaicin receptor: a heat-activated ion channel in the pain pathway. Nature. 1997; 389(6653):816–24. https://doi.org/10.1038/39807 PMID: 9349813
5. Szallasi A, Cortright DN, Blum CA, Eid SR. The vanilloid receptor TRPV1: 10 years from channel cloning to antagonist proof-of-concept. Nat Rev Drug Discov. 2007; 6(5):357–72. https://doi.org/10.1038/nrd2280 PMID: 17464295
6. Julius D. TRP channels and pain. Annual review of cell and developmental biology. 2013; 29:355–84. https://doi.org/10.1146/annurev-cellbio-101011-155833 PMID: 24099085
7. Docherty RJ, Yeats JC, Bevan S, Boddeke HW. Inhibition of calcineurin inhibits the desensitization of capsaicin-evoked currents in cultured dorsal root ganglion neurones from adult rats. Pflugers Arch. 1996; 431(6):828–37. PMID: 8927498
8. Zinn S, Sisignano M, Kern K, Pierre S, Tunaru S, Jordan H, et al. The leukotriene B4 receptors BLT1 and BLT2 form an antagonistic sensitizing system in peripheral sensory neurons. J Biol Chem. 2017; 292(15):6123–34. https://doi.org/10.1074/jbc.M116.769125 PMID: 28242764
9. Park JJ, Lee J, Kim MA, Back SK, Hong SK, Na HS. Induction of total insensitivity to capsaicin and hypersensitivity to garlic extract in human by decreased expression of TRPV1. Neurosci Lett. 2007; 411 (2):87–91. https://doi.org/10.1016/j.neulet.2006.10.046 PMID: 17110039
10. Kim H, Neubert JK, San Miguel A, Xu K, Krishnaraju RK, Iadarola MJ, et al. Genetic influence on variability in human acute experimental pain sensitivity associated with gender, ethnicity and psychological temperament. Pain. 2004; 109(3):488–96. https://doi.org/10.1016/j.pain.2004.02.027 PMID: 15157710
11. Boukalova S, Touska F, Marsakova L, Hynkova A, Sura L, Chvojka S, et al. Gain-of-function mutations in the transient receptor potential channels TRPV1 and TRPA1: how painful? Physiological research / Academia Scientiarum Bohemoslovaca. 2014; 63 Suppl 1:S205–13.
12. Metzker ML. Sequencing technologies—the next generation. Nat Rev Genet. 2010; 11(1):31–46. https://doi.org/10.1038/nrg2626 PMID: 19997069
13. Frank M, Prenzler A, Eils R, Graf von der Schulenburg JM. Genome sequencing: a systematic review of health economic evidence. Health economics review. 2013; 3(1):29. https://doi.org/10.1186/2191-1991-3-29 PMID: 24330507
14. Vigna SR, Shahid RA, Nathan JD, McVey DC, Liddle RA. Leukotriene B4 mediates inflammation via TRPV1 in duct obstruction-induced pancreatitis in rats. Pancreas. 2011; 40(5):708–14. https://doi.org/10.1097/MPA.0b013e318214c8df PMID: 21602738
15. Fisher RA. On the Interpretation of Chi Square from Contingency Tables, and the Calculation of P. Journal of the Royal Statistical Society. 1922; 85(1):87–94.
16. Emigh TH. A comparison of tests for Hardy-Weinberg equilibrium. Biometrics. 1980; 36(4):627–42. PMID: 25856832
17. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. ProcNatlAcadSciUSA. 1977; 74(12):5463–7.
18. Sanger F, Coulson AR. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. J Mol Biol. 1975; 94(3):441–8. PMID: 1100841
19. Tarabeux J, Zeitouni B, Moncoutier V, Tenreiro H, Abidallah K, Lair S, et al. Streamlined ion torrent PGM-based diagnostics: BRCA1 and BRCA2 genes as a model. 2014;(1476–5438 (Electronic)).
20. Millat G, Chanavat V, Rousson R. Evaluation of a new NGS method based on a custom AmpliSeq library and Ion Torrent PGM sequencing for the fast detection of genetic variations in cardiomyopathies. Clin Chim Acta. 2014; 433:266–71. https://doi.org/10.1016/j.cca.2014.03.032 PMID: 24721642
21. Concolino P, Costella A, Minucci A, Scaglione GL, Santonocito C, Salutari V, et al. A preliminary Quality Control (QC) for next generation sequencing (NGS) library evaluation turns out to be a very useful tool

for a rapid detection of BRCA1/2 deleterious mutations. Clin Chim Acta. 2014; 437:72–7. https://doi.org/10.1016/j.cca.2014.06.026 PMID: 25007954

22. Glotov AS, Kazakov SV, Zhukova EA, Alexandrov AV, Glotov OS, Pakin VS, et al. Targeted next-generation sequencing (NGS) of nine candidate genes with custom AmpliSeq in patients and a cardiomyopathy risk group. Clin Chim Acta. 2015; 446:132–40. https://doi.org/10.1016/j.cca.2015.04.014 PMID: 25892673

23. Tominaga M, Caterina MJ, Malmberg AB, Rosen TA, Gilbert H, Skinner K, et al. The cloned capsaicin receptor integrates multiple pain-producing stimuli. Neuron. 1998; 21(3):531–43. PMID: 9768840

24. Davis JB, Gray J, Gunthorpe MJ, Hatcher JP, Davey PT, Overend P, et al. Vanilloid receptor-1 is essential for inflammatory thermal hyperalgesia. Nature. 2000; 405(6783):183–7. https://doi.org/10.1038/35012076 PMID: 10821274

25. Szallasi A, Bíró T, Szabó T, Modarres S, Petersen M, Klusch A, et al. A non-pungent triprenyl phenol of fungal origin, scutigeral, stimulates rat dorsal root ganglion neurons via interaction at vanilloid receptors. Br J Pharmacol. 1999; 126(6):1351–8. https://doi.org/10.1038/sj.bjp.0702440 PMID: 10217528

26. Szallasi A, Blumberg PM. Vanilloid (Capsaicin) receptors and mechanisms. Pharmacol Rev. 1999; 51 (2):159–212. PMID: 10353985

27. Rebhan M, Chalifa-Caspi V, Prilusky J, Lancet D. GeneCards: integrating information about genes, proteins and diseases. Trends Genet. 1997; 13(4):163. PMID: 9097728

28. Tulah AS, Parker SG, Moffatt MF, Wardlaw AJ, Connolly MJ, Sayers I. The role of ALOX5AP, LTA4H and LTB4R polymorphisms in determining baseline lung function and COPD susceptibility in UK smokers. BMC Med Genet. 2011; 12:173. https://doi.org/10.1186/1471-2350-12-173 PMID: 22206291

29. Tulah AS, Beghe B, Barton SJ, Holloway JW, Sayers I. Leukotriene B4 receptor locus gene characterisation and association studies in asthma. BMC Med Genet. 2012; 13:110. https://doi.org/10.1186/1471-2350-13-110 PMID: 23167751

30. Chen CL, Li H, Xing XH, Guan HS, Zhang JH, Zhao JW. Effect of TRPV1 gene mutation on bronchial asthma in children before and after treatment. Allergy and asthma proceedings: the official journal of regional and state allergy societies. 2015; 36(2):e29–36.

31. Ren YF, Li H, Xing XH, Guan HS, Zhang BA, Chen CL, et al. Preliminary study on pathogenesis of bronchial asthma in children. Pediatr Res. 2015; 77(4):506–10. https://doi.org/10.1038/pr.2015.11 PMID: 25585038

32. Wang MY, Cummock MD, Yu Y, Trivedi RA. An analysis of the differences in the acute hospitalization charges following minimally invasive versus open posterior lumbar interbody fusion. J Neurosurg Spine. 2010; 12(6):694–9. https://doi.org/10.3171/2009.12.SPINE09621 PMID: 20515357

33. Park S, Zhang X, Lee NR, Jin HS. TRPV1 Gene Polymorphisms Are Associated with Type 2 Diabetes by Their Interaction with Fat Consumption in the Korean Genome Epidemiology Study. J Nutrigenet Nutrigenomics. 2016; 9(1):47–61. https://doi.org/10.1159/000446499 PMID: 27287034

34. Triantafyllou K, Kourikou A, Gazouli M, Karamanolis GP, Dimitriadis GD. Functional dyspepsia susceptibility is related to CD14, GNB3, MIF, and TRPV1 gene polymorphisms in the Greek population. Neurogastroenterol Motil. 2017; 29(1).

35. Kim H, Mittal DP, Iadarola MJ, Dionne RA. Genetic predictors for acute experimental cold and heat pain sensitivity in humans. J Med Genet. 2006; 43(8):e40. https://doi.org/10.1136/jmg.2005.036079 PMID: 16882734

36. Valdes AM, De Wilde G, Doherty SA, Lories RJ, Vaughn FL, Laslett LL, et al. The Ile585Val TRPV1 variant is involved in risk of painful knee osteoarthritis. Ann Rheum Dis. 2011; 70(9):1556–61. https://doi.org/10.1136/ard.2010.148122 PMID: 21616913

37. Winter Z, Buhala A, Otvos F, Josvay K, Vizler C, Dombi G, et al. Functionally important amino acid residues in the transient receptor potential vanilloid 1 (TRPV1) ion channel—an overview of the current mutational data. Mol Pain. 2013; 9:30. https://doi.org/10.1186/1744-8069-9-30 PMID: 23800232

38. Wang S, Joseph J, Diatchenko L, Ro JY, Chung MK. Agonist-dependence of functional properties for common nonsynonymous variants of human transient receptor potential vanilloid 1. Pain. 2016; 157 (7):1515–24. https://doi.org/10.1097/j.pain.0000000000000556 PMID: 26967694

39. Clapham DE, Runnels LW, Strubing C. The TRP ion channel family. Nat Rev Neurosci. 2001; 2(6):387–96. https://doi.org/10.1038/35077544 PMID: 11389472

40. Malin S, Molliver D, Christianson JA, Schwartz ES, Cornuet P, Albers KM, et al. TRPV1 and TRPA1 function and modulation are target tissue dependent. J Neurosci. 2011; 31(29):10516–28. https://doi.org/10.1523/JNEUROSCI.2992-10.2011 PMID: 21775597

41. Story GM, Peier AM, Reeve AJ, Eid SR, Mosbacher J, Hricik TR, et al. ANKTM1, a TRP-like channel expressed in nociceptive neurons, is activated by cold temperatures. Cell. 2003; 112(6):819–29. PMID: 12654248

42. Cheng W, Yang F, Takanishi CL, Zheng J. Thermosensitive TRPV channel subunits coassemble into heteromeric channels with intermediate conductance and gating properties. J Gen Physiol. 2007; 129 (3):191–207. https://doi.org/10.1085/jgp.200709731 PMID: 17325193

43. Cheng W, Yang F, Liu S, Colton CK, Wang C, Cui Y, et al. Heteromeric heat-sensitive transient receptor potential channels exhibit distinct temperature and chemical response. J Biol Chem. 2012; 287 (10):7279–88. https://doi.org/10.1074/jbc.M111.305045 PMID: 22184123

44. Ho KW, Ward NJ, Calkins DJ. TRPV1: a stress response protein in the central nervous system. American journal of neurodegenerative disease. 2012; 1(1):1–14. PMID: 22737633

45. Bautista DM, Jordt SE, Nikai T, Tsuruda PR, Read AJ, Poblete J, et al. TRPA1 mediates the inflammatory actions of environmental irritants and proalgesic agents. Cell. 2006; 124(6):1269–82. https://doi.org/10.1016/j.cell.2006.02.023 PMID: 16564016

46. Guimaraes MZP, Jordt SE. TRPA1: A Sensory Channel of Many Talents. In: Liedtke WB, Heller S, editors. TRP Ion Channel Function in Sensory Transduction and Cellular Signaling Cascades. Frontiers in Neuroscience. Boca Raton (FL) 2007.

47. Carreno O, Corominas R, Fernandez-Morales J, Camina M, Sobrido MJ, Fernandez-Fernandez JM, et al. SNP variants within the vanilloid TRPV1 and TRPV3 receptor genes are associated with migraine in the Spanish population. Am J Med Genet B Neuropsychiatr Genet. 2012; 159b(1):94–103. https://doi.org/10.1002/ajmg.b.32007 PMID: 22162417

48. Yokomizo T, Kato K, Hagiya H, Izumi T, Shimizu T. Hydroxyeicosanoids bind to and activate the low affinity leukotriene B4 receptor, BLT2. J Biol Chem. 2001; 276(15):12454–9. https://doi.org/10.1074/jbc.M011361200 PMID: 11278893

49. Tager AM, Luster AD. BLT1 and BLT2: the leukotriene B(4) receptors. Prostaglandins Leukot Essent Fatty Acids. 2003; 69(2–3):123–34. PMID: 12895595

50. Craib SJ, Ellington HC, Pertwee RG, Ross RA. A possible role of lipoxygenase in the activation of vanilloid receptors by anandamide in the guinea-pig bronchus. Br J Pharmacol. 2001; 134(1):30–7. https://doi.org/10.1038/sj.bjp.0704223 PMID: 11522594

51. Hwang SW, Oh U. Hot channels in airways: pharmacology of the vanilloid receptor. Current opinion in pharmacology. 2002; 2(3):235–42. PMID: 12020463

52. Duroudier NP, Tulah AS, Sayers I. Leukotriene pathway genetics and pharmacogenetics in allergy. Allergy. 2009; 64(6):823–39. https://doi.org/10.1111/j.1398-9995.2009.02015.x PMID: 19416143

53. Tulah AS, Beghé B, Barton SJ, Holloway JW, Sayers I. Leukotriene B4 receptor locus gene characterisation and association studies in asthma. BMC Med Genet. 2012; 13:110. https://doi.org/10.1186/1471-2350-13-110 PMID: 23167751

54. Rizzo JM, Buck MJ. Key principles and clinical applications of "next-generation" DNA sequencing. Cancer Prev Res (Phila). 2012; 5(7):887–900.

55. Mardis ER. Next-generation DNA sequencing methods. Annu Rev Genomics Hum Genet. 2008; 9:387–402. https://doi.org/10.1146/annurev.genom.9.081307.164359 PMID: 18576944

56. Shendure J, Ji H. Next-generation DNA sequencing. Nature biotechnology. 2008; 26(10):1135–45. https://doi.org/10.1038/nbt1486 PMID: 18846087

57. Rothberg JM, Hinz W, Rearick TM, Schultz J, Mileski W, Davey M, et al. An integrated semiconductor device enabling non-optical genome sequencing. Nature. 2011; 475(7356):348–52. https://doi.org/10.1038/nature10242 PMID: 21776081

58. de Leng WW, Gadellaa-van Hooijdonk CG, Barendregt-Smouter FA, Koudijs MJ, Nijman I, Hinrichs JW, et al. Targeted Next Generation Sequencing as a Reliable Diagnostic Assay for the Detection of Somatic Mutations in Tumours Using Minimal DNA Amounts from Formalin Fixed Paraffin Embedded Material. PLoS One. 2016; 11(2):e0149405. https://doi.org/10.1371/journal.pone.0149405 PMID: 26919633

59. Sie D, Snijders PJ, Meijer GA, Doeleman MW, van Moorsel MI, van Essen HF, et al. Performance of amplicon-based next generation DNA sequencing for diagnostic gene mutation profiling in oncopathology. Cell Oncol (Dordr). 2014; 37(5):353–61.

60. Lin MT, Mosier SL, Thiess M, Beierl KF, Debeljak M, Tseng LH, et al. Clinical validation of KRAS, BRAF, and EGFR mutation detection using next-generation sequencing. Am J Clin Pathol. 2014; 141 (6):856–66. https://doi.org/10.1309/AJCPMWGWGO34EGOD PMID: 24838331

61. McCourt CM, McArt DG, Mills K, Catherwood MA, Maxwell P, Waugh DJ, et al. Validation of next generation sequencing technologies in comparison to current diagnostic gold standards for BRAF, EGFR and KRAS mutational analysis. PLoS One. 2013; 8(7):e69604. https://doi.org/10.1371/journal.pone.0069604 PMID: 23922754

62. Portier BP, Kanagal-Shamanna R, Luthra R, Singh R, Routbort MJ, Handal B, et al. Quantitative assessment of mutant allele burden in solid tumors by semiconductor-based next-generation

sequencing. Am J Clin Pathol. 2014; 141(4):559–72. https://doi.org/10.1309/AJCP1JUGQMW7ZNTL PMID: 24619758

63. Lohmann K, Klein C. Next generation sequencing and the future of genetic diagnosis. Neurotherapeutics: the journal of the American Society for Experimental NeuroTherapeutics. 2014; 11(4):699–707.

64. Norton N, Li D, Hershberger RE. Next-generation sequencing to identify genetic causes of cardiomyopathies. Current opinion in cardiology. 2012; 27(3):214–20. https://doi.org/10.1097/HCO.0b013e328352207e PMID: 22421630

65. Brito R, Sheth S, Mukherjea D, Rybak LP, Ramkumar V. TRPV1: A Potential Drug Target for Treating Various Diseases. Cells. 2014; 3(2):517–45. https://doi.org/10.3390/cells3020517 PMID: 24861977

66. Wang Q, Bai X, Xu D, Xu D, Li H, Fang J, et al. [TRPV1 UTR-3 polymorphism and susceptibility of childhood asthma of the Han Nationality in Beijing]. Wei Sheng Yan Jiu. 2009; 38(5):516–21. PMID: 19877503

67. Smit LA, Kogevinas M, Anto JM, Bouzigon E, Gonzalez JR, Le Moual N, et al. Transient receptor potential genes, smoking, occupational exposures and cough in adults. Respir Res. 2012; 13:26. https://doi.org/10.1186/1465-9921-13-26 PMID: 22443337

68. Dias AG, Rousseau D, Duizer L, Cockburn M, Chiu W, Nielsen D, et al. Genetic variation in putative salt taste receptors and salt taste perception in humans. Chem Senses. 2013; 38(2):137–45. https://doi.org/10.1093/chemse/bjs090 PMID: 23118204

69. Jhun EH, Yao Y, He Y, Mack AK, Wilkie DJ, Molokie RE, et al. Prevalence of pain-related single nucleotide polymorphisms in patients of African origin with sickle cell disease. Pharmacogenomics. 2015; 16 (16):1795–806. https://doi.org/10.2217/pgs.15.126 PMID: 26555434

70. Vrabec JT, Liu L, Li B, Leal SM. Sequence variants in host cell factor C1 are associated with Meniere's disease. Otol Neurotol. 2008; 29(4):561–6. https://doi.org/10.1097/MAO.0b013e318168d23b PMID: 18520591

71. Gulsuner S, Tekinay AB, Doerschner K, Boyaci H, Bilguvar K, Unal H, et al. Homozygosity mapping and targeted genomic sequencing reveal the gene responsible for cerebellar hypoplasia and quadrupedal locomotion in a consanguineous kindred. Genome Res. 2011; 21(12):1995–2003. https://doi.org/10.1101/gr.126110.111 PMID: 21885617

72. Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, et al. dbSNP: the NCBI database of genetic variation. Nucleic Acids Res. 2001; 29(1):308–11. PMID: 11125122

Erklärung zu den Autorenanteilen an der Publikation:

Machine-learned analysis of the association of next-generation sequencing–based human TRPV1 and TRPA1 genotypes with the sensitivity to heat stimuli and topically applied capsaicin (*printed*)

Name der Zeitschrift: PAIN

Beteiligte Autoren: D Kringel, G Geisslinger, E Resch, B G Oertel, M C Thrun, S Heinemann und J Lötsch

Was hat der Promovierende bzw. was haben die Koautoren beigetragen?

(1) zu Entwicklung und Planung

Promovierender: 50%
Autor JL: 50%

(2) zur Durchführung der einzelnen Untersuchungen und Experimente

Promovierender: 80% (DNA Extraktion, NGS-Panel Design, Sequenzierung, Varianten Annotation, Methoden Validierung)
Autor SH: 20% (Human Experimente)

(3) zur Erstellung der Datensammlung und Abbildungen

Promovierender: 30% (Datenerhebung NGS, Run-Statistik)
Autor SH: 10% (Datenerhebung Schmerzschwellen)
Autor JL: 60% (Erstellung der Abbildungen)

(4) zur Analyse und Interpretation der Daten

Promovierender: 30 % (Datenvorverarbeitung, Charakterisierung und Assoziation der Varianten, Dateninterpretation)
Autor JL: 70% (Datenvisualisierung, Programmierung der KI, Anwendung der Klassifikatoren, Datenanalyse, Dateninterpretation)

(5) zum Verfassen des Manuskripts

Promovierender: 35%
Autor GG: 5%
Autor ER: 5%
AutorBGÖ: 5%
Autor MCT: 10%
Autor SH: 5%
Autor JL: 35%

Zustimmende Bestätigungen der oben genannten Angaben:

_____	_____
Datum/Ort	Unterschrift Promovend
_____	_____
Datum/Ort	Unterschrift Betreuer

Research Paper

PAIN®

OPEN

Machine-learned analysis of the association of next-generation sequencing–based human *TRPV1* and *TRPA1* genotypes with the sensitivity to heat stimuli and topically applied capsaicin

Dario Kringel[a], Gerd Geisslinger[a], Eduard Resch[b], Bruno G. Oertel[b], Michael C. Thrun[b], Sarah Heinemann[a], Jörn Lötsch[a,b,*]

Abstract

Heat pain and its modulation by capsaicin varies among subjects in experimental and clinical settings. A plausible cause is a genetic component, of which TRPV1 ion channels, by their response to both heat and capsaicin, are primary candidates. However, TRPA1 channels can heterodimerize with TRPV1 channels and carry genetic variants reported to modulate heat pain sensitivity. To address the role of these candidate genes in capsaicin-induced hypersensitization to heat, pain thresholds acquired before and after topical application of capsaicin and *TRPA1*/*TRPV1* exomic sequences derived by next-generation sequencing were assessed in n = 75 healthy volunteers and the genetic information comprised 278 loci. Gaussian mixture modeling indicated 2 phenotype groups with high or low capsaicin-induced hypersensitization to heat. Unsupervised machine learning implemented as swarm-based clustering hinted at differences in the genetic pattern between these phenotype groups. Several methods of supervised machine learning implemented as random forests, adaptive boosting, k-nearest neighbors, naive Bayes, support vector machines, and for comparison, binary logistic regression predicted the phenotype group association consistently better when based on the observed genotypes than when using a random permutation of the exomic sequences. Of note, *TRPA1* variants were more important for correct phenotype group association than *TRPV1* variants. This indicates a role of the *TRPA1* and *TRPV1* next-generation sequencing–based genetic pattern in the modulation of the individual response to heat-related pain phenotypes. When considering earlier evidence that topical capsaicin can induce neuropathy-like quantitative sensory testing patterns in healthy subjects, implications for future analgesic treatments with transient receptor potential inhibitors arise.

Keywords: Data science, Machine learning, Next-generation sequencing, Genetics of pain, Human, Experimental pain models

1. Introduction

The perception of pain after noxious stimulation involves a complex pathophysiology[13] processed in a large network of nociceptive molecular pathways.[30] This complexity extends to the perception of apparently uniform stimuli such as heat shown to follow a multimodal distribution.[86] This hints at interindividual differences in involved sensors of which the largest group belongs to the transient receptor potential (TRP) channels.[9] In particular, TRPV1 is known as a thermosensitive channel involved in nociception,[61] and in addition to heat, it is also gated by pungent chemicals such as vanilloids including capsaicin.[15]

Synergistic effects of chemical and thermal gating are used to study mechanisms of thermal hyperalgesia in humans.[58] Although hyperalgesia varies among patients with pain,[42] in experimental settings topical capsaicin application induces hyperalgesia only in a fraction of subjects.[42] This may point at a genetic background where *TRPV1* as a primary candidate gene is playing a role in both, heat sensation and capsaicin hypersensitization. However, associations of genetic variants with the heat sensitivity or hypersensitization by capsaicin were only rarely reported. However, the only hint at an association of *TRPV1* genetics with heat pain sensitivity in humans points at the rs8065080 single-nucleotide polymorphism (SNP),[36] which was not replicated.[35] Moreover, an unexpected role of a *TRPA1* genetic variant rs11988795 in heat pain was reproduced.[35,68] The coexpression of TRPV1 and TRPA1[72] and the flexibility of the TRP family channels raise the possibility that these channels might interact to influence the properties of one another.[20] In this regard, recently reported heteromerization among the TRP channels is suggestive of the mechanism for interactions.[28]

Based on this evidence and considering the unresolved role of *TRPV1* variants for the modulation of human pain sensitivity, despite the molecular plausibility of an involvement, the present

Sponsorships or competing interests that may be relevant to content are disclosed at the end of this article.

[a] Institute of Clinical Pharmacology, Goethe-University, Frankfurt am Main, Germany, [b] Fraunhofer Institute for Molecular Biology and Applied Ecology IME, Project Group Translational Medicine and Pharmacology TMP, Frankfurt am Main, Germany

**Corresponding author. Address: Goethe-University, Theodor Stern Kai 7, 60590 Frankfurt am Main, Germany. Tel.: +49-69-6301-4589; fax: +49-69-6301-4354. E-mail address: j.loetsch@em.uni-frankfurt.de (J. Lötsch).*

Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's Web site (www.painjournalonline.com).

PAIN 159 (2018) 1366–1381

http://dx.doi.org/10.1097/j.pain.0000000000001222

analysis addressed the association between *TRPV1* and *TRPA1* genotypes with a human phenotype of capsaicin-induced hyperalgesia to heat stimuli. With the broader availability of next-generation sequencing (NGS), the limitation to known functional variants has fallen in favor of unrestricted access of TRP channel genetics. Therefore, it is not necessary to assess the phenotypic role of TRP channel genotypes for selected single variants. This accommodates increasing molecular evidence that noncoding variants can affect mRNA splicing, stability, and structure, resulting in a reduced transcriptional efficiency[22,23,77] rendering them potentially functionally relevant. Hence, a recently developed genetic panel[38] was used to address the role *TRPV1* and *TRPA1* genetic variants in the sensitivity to nociceptive heat and in the reaction to hypersensitization with topical capsaicin recently assessed in a cohort of healthy subjects.[46]

2. Methods

2.1. Data sets, subjects, and study design

The phenotype data sets and DNA samples were available from a previous study,[46] enrolling n = 100 healthy volunteers (46 men) of Caucasian ethnicity by self-assignment, aged 19 to 42 years (mean ± SD 25 ± 3.5 years). In this data set, phenotypic measurements from n = 82 subjects were nonmissing and included in the present analysis. The study followed the Declaration of Helsinki and was approved by the Ethics Committee of the Goethe-University Medical Faculty, Frankfurt am Main, Germany. Informed written consent in the study procedures including the genotyping had been obtained from each participating subject.

Inclusion criteria were age between 18 and 50 years and no relevant current medical history. The subjects' actual health had been ascertained by medical history and physical examination including vital signs. Exclusion criteria were drug intake during the previous week, except for oral contraceptives and vitamin or hormone-substituting drugs (eg, L-thyroxin), a current clinical condition involving pain, and current diseases according to questioning and medical examination. Alcohol was prohibited for 24 hours before the actual experiments. Before the experimental tests, all subjects completed training sessions with pain tests applied to an area different from the planned skin areas.

2.2. Assessment of heat pain sensitivity

In the capsaicin experimental pain model, chemical methods of nociceptor stimulation were used to produce stable and long-lasting hyperalgesia with a low potential for skin injury, in the original publication supplemented by heat stimulation.[58] Topical application of 150 mg capsaicin cream (0.2%, manufactured by the local Hospital Pharmacy) onto a 3 × 3 cm^2 skin area was used. Subsequently, the area was covered with a plaster for 30 minutes.

Quantitative sensory testing (QST) was performed at baseline and after application of capsaicin. A clinically established QST test battery proposed by the German Research Network on Neuropathic Pain[63,64] was used. For the present report, pain thresholds to noxious heat were selected. They were assessed using a 3 × 3 cm thermode (TSA 2001—II; Ramat Yishai, Israel) on a 9 cm^2 skin area at the inside of the forearm without any superficial veins or birth marks. Heat pain thresholds (HPTs) were measured by increasing the temperature of the thermode by 1°C/s, starting at 32°C, until the subject indicated pain, which triggered the reversal of the temperature ramp back to the baseline. According to the published instructions for the QST test battery,[59,63,64] the HPT was defined as the mean of 3 measurement repetitions. During testing, the room temperature was kept at 20 to 25°C.

Data were preprocessed according to the QST test battery instructions,[59,63,64] which included uniform direction along increasing stimulus intensity as $HPT_T = HPT - 32°C$, where the subscript T denotes the data transformation. The values of HPT_T were mapped onto the distribution of the reference group that consists of 180 healthy subjects, in whom a data set of 1080 QST parameter values has been obtained. This serves as the reference for all QST-based diagnoses.[52] Therefore, according to the QST standard procedure, the individual QST parameter values were z-transformed as $z_{QST,individual} = \frac{QST_{individual} - QST_{reference}}{SD_{reference}}$, with QST reference values with regard to the sex, age, and tested body site of the actual subject taken from.[52] The signs of the z-scores, $zHPT_T$, were adjusted to denote that a z-score >0 indicates high sensitivity and z-score <0 indicates low sensitivity, according to the standardized instructions. The effect of capsaicin was quantified as the difference between the measurement after capsaicin application and the measurement without the presence of capsaicin, ie, $CapsEff = zHPT_{T,capsaicin} - zHPT_{T,baseline}$.

2.3. Transient receptor potential channel genotyping using next-generation sequencing

Next-generation sequencing of *TRPA1* and *TRPV1* genes was based on a custom AmpliSeq library and performed using a validated assay on an Ion Torrent personal genome machine as described in detail previously.[38] In brief, genomic DNA was extracted from 200 μL venous blood on a BioRobot EZ1 workstation applying the blood and body fluid spin protocol provided in the EZ1 DNA Blood 200 μL Kit (Qiagen, Hilden, Germany). A multiplex amplification primer set for the exomic sequences of the TRP channel genes was designed online using a web tool (Ion Ampliseq Designer; Life Technologies, Darmstadt, Germany) provided by the manufacturer of the NGS device at http://www.ampliseq.com.

The present amplification design obtained coverage of 96% of target sequence. After sequencing, signal processing was performed using Torrent Suite software (version 5.2.2; Life Technologies), base calling and the generation of unmapped and mapped binary alignment map files (hg19 reference genomic sequence) were performed. Variant calling across the hg19 reference genomic sequence was performed with the Torrent Variant Caller Plugin (minimum quality = 10, minimum coverage = 20, and minimum coverage on either strand = 3) and variant annotation was performed using Ion Reporter Software (version 5.2.2; Life Technologies). Variant call format files containing the nucleotide reads were processed toward the individual genotypes using GenomeBrowse software (Version 2.0.4; Golden Helix, Bozeman, MT) and SNP and Variation Suite software (Version 8.7.1; Golden Helix).

2.4. Data analysis

To accommodate a large number of genetic variants expected to result from the NGS-based genotyping, the main genotype–phenotype association analysis was implemented using a novel approach based on machine-learned techniques (for an overview on machine learning in pain research, see 49). The main idea was to train an artificial intelligence, implemented as different types of machine learning, to learn the association of the genetic information with the pain-related phenotype, and to subsequently use the trained intelligence to predict a phenotype in new data

from genetic information. If this performed better than guessing the phenotype or than using genetic information unrelated to the phenotype, a genotype–phenotype association can be concluded as supported by the data. Machine learning was a priori preferred to the sole use of traditional approaches such as logistic regression analysis because of the expected high dimensionality and collinearity of the rich genetic information; indeed, the nevertheless included regression analysis was outperformed by several machine-learned methods (see Results section). The concept of training an artificial intelligence with genetic information to enable it to correctly associate an individual with a pain phenotype class required measures against overfitting,[54] which are usually implemented as splitting the data set into a training subset that is provided to the artificial intelligence during the learning phase and a test subset which is not seen by the artificial intelligence during learning but provided when the learned algorithm is used for classification; usually, this procedure is repeated several times in a resampling design.[54]

Data were analyzed using the R software package (version 3.4.1 for Linux; http://CRAN.R-project.org/)[60] on an Intel Xeon computer running on Ubuntu Linux 16.04.3 64-bit. Supervised and unsupervised machine learning was used for genotype vs phenotype association. Machine learning addresses the so-called data space $D = \{(x_i, y_i)|x_i \in X, y_i \in Y,\ i = 1, \ldots, n\}$ including an input space X comprising vectors $x_i = < x_{i,1}, \ldots, x_{i,d} >$ with $d > 0$ different parameters (here, the genetic information) acquired from $n > 0$ cases belonging to the output classes y_i (eg, a pain-related phenotype). In unsupervised learning, the class information is disregarded and only the so-called feature space comprising an unlabeled data set of $D = \{(x_i)|x_i \in X,\ i = 1, \ldots, n\}$, composed of values $x_i \in X \subset \mathbb{R}^d$ comprising the d features, respectively, genetic markers is searched with the goal to find "interesting" structures, which can be associated subsequently with the phenotypes. By contrast, in supervised machine learning, an algorithm is trained on data for which the class labels of the cases are known that is able to assign future cases for which this class label information is unknown to the right class (prediction and generalization[18]).

The analysis was performed in 4 main steps comprising (1) creation of a phenotype group structure, (2) preprocessing of the *TRPV1* and *TRPA1* NGS genetic information, (3) identification of a genetic marker pattern and its relation to the phenotype classes, and (4) finding a mapping of the genetic parameters to the phenotype classes.

2.5. Identification of capsaicin sensitivity phenotype classes

The first step of the data analysis aimed at establishing the output data space, ie, a phenotype class structure. Therefore, the distribution of the changes after capsaicin application, CapsEff, was investigated by analyzing the probability density function (PDF) as described previously.[43,86] In brief, the Pareto density estimation (PDE), ie, a kernel density estimator particularly suitable for the discovery of groups in the data,[81] was used. A multimodal distribution of the pain responses was assessed by fitting a Gaussian mixture model (GMM) to the PDEs as $P(x) = \sum_{i=0}^{M} w_i N(x|m_i, s_i) = \sum_{i=1}^{M} w_i \cdot \frac{1}{\sqrt{2\pi}s_i} \cdot e^{-\frac{(x-m_i)^2}{2s_i^2}}$, where $N(x|m_i, s_i)$ denotes Gaussian probability densities (components) with mean values m_i and SDs s_i. The w_i denotes the mixture weights indicating the relative contribution of each Gaussian component to the overall distribution, which add up to a value of 1. M denotes the number of components in the mixture. Gaussian mixture model fitting was performed with our R package "AdaptGauss" (https://cran.r-project.org/package=AdaptGauss).[86] To determine the optimum number of components, model optimization was performed for M = 1 to 5 components. The final model was selected based on likelihood ratio tests.[73] In addition, the Kolmogorov–Smirnov test[70] was applied to assess whether the observed distribution differed significantly from the expectation from the model, and the quality of the model to fit the distribution was assessed visually using a quantile–quantile (QQ) plot. Subject association to the identified subgroups was obtained using the Bayes' theorem[2] that provided the probability that an individual observation belongs to mode i calculated as the posterior probability. Thus, the output space Y was obtained, comprising $y_i \in C = \{1, \ldots, c\}$, where c denotes possible unambiguous classes c where every y_i has a unique class label and the number of classes was equal to the number of Gaussian modes, M.

2.6. Preprocessing of the genetic information

The determination of single-nucleotide variants from the NGS data refers to the Software plugin "The Torrent Variant Caller" (TVC) provided by Life Technologies. A variant is defined as a nucleotide disagreeing with the nucleotide in the reference sequence. The TVC plugin calls SNPs, multinucleotide polymorphisms, insertions, and deletions in a sample across a reference (hg19). In the second step of the analysis, the genetic information (mainly SNPs) was curated by (1) eliminating non-informative variants and (2) creating of negative and positive genetic control data sets with respect to a possible association of the genotype with the phenotypes. Variants were eliminated for which the distribution of homozygous and heterozygous carriers differed from expectation according to the Hardy–Weinberg equilibrium.[26] This was judged by means of Fishers exact tests[21] using the R package "HardyWeinberg" (https://cran.r-project.org/package=HardyWeinberg).[25] To avoid the inclusion of non-informative variants such as those carried by only very few subjects into the classifier, informative gene loci were detected based on the Shannon information[69] computed as $\text{Info} = -P_{0,i} \cdot \ln(P_{0,i}) - P_{1,i} \cdot \ln(P_{1,i})$, where $P_{0,i}$ and $P_{1,i}$ are the observed probabilities of the nonobservation (0) or observation (1), respectively, of a variant allele in the ith gene locus. The precise limit of the Shannon information up to which a gene locus could be regarded sufficiently informative, was calculated by means of a computed ABC analysis.[83] This is a categorization technique for the selection of a most important subset among a larger set of positive numerical items. It divides the set into 3 disjoint subsets "A," "B," and "C"[93] referred to in economic sciences where the method originates as "the important few" (set "A") vs "the trivial many" (set "C"),[31] whereas set "B" comprises items between the 2 extremes including elements where an increase in effort is proportional to the increase in yield. However, although earlier applications of ABC analyses parted the item set according to the so-called 80/20 rule, which sets the limit between sets "A" and "B" at 80% of the yield achieved with 20% efforts, this limit is based on mathematical calculations in computed ABC analysis[83] implemented in our R package "ABCanalysis" (http://cran.r-project.org/package=ABCanalysis).[83] As subset "A" can be regarded as containing the most profitable features,[31,55] it was chosen for classifier establishment. The limit to set "B" was found at Shannon information = 0.339. Furthermore, as implemented previously,[39] further variants unlikely to provide a suitable basis for phenotype class assignment were excluded. In the present analysis, this was approached through the effect sizes of the allelic distribution between the phenotype classes used classic χ^2 statistics.[57] The

values of χ^2 obtained for each gene locus were submitted to a computed ABC analysis described above. Here, only the clearly unsuitable variants were omitted, ie, ABC set "C" regarded as comprising "the trivial many."[31]

Genetic control data sets were created by rearranging the original genotype information. Specifically, a negative control feature set was obtained by random permutation of the genetic data. The expectation was that the association with the phenotypes was not better than guessing and should be consistently outperformed by the mapping of the true genotypes to the phenotypes using different machine-learned methods. In addition, a positive control feature set was obtained by sorting the original genotype information at each locus in descending order data of the number of variant alleles along the sorted phenotype classes (**Fig. 1**). The expectation was that the association with the phenotypes could be almost perfectly obtained by all machine-learning methods.

2.7. Identification of a genetic marker pattern and its relation to the phenotype classes

In the third step of the analysis, the genetic information was explored for data structures. Their existence would support that the *TRPA1* and *TRPV1* NGS genotypes were not homogenously distributed among the subjects but hinted at subgroups of subject based on the genetic information. This would be a first step to further explore the data for a possible relation of the genotype-based subgroups with the phenotype classes. Hence, the preprocessed genetic information was analyzed for patterns using unsupervised machine learning, which was implemented as a swarm of intelligent agents called "DataBots."[80] The data space $D = \{x_i, i = 1, \ldots, n\} \subset \mathbb{R}^d$, comprising d genetic markers acquired in n subjects was explored for distance-based structures using the cityblock (Manhattan) distance[12] as used elsewhere[94] for genetic data scaled [0,1,2]. To explore this feature space, topographic mapping was used, which provides data projection methods to create low-dimensional images from high-dimensional data. Specifically, topographic mapping was implemented as swarm intelligence, ie, an algorithm guided by the flocking behavior of numerous independent but cooperating the so-called DataBots, which are self-organizing artificial "life forms" identified with single data objects (subjects). These "DataBots" can move on a 2-dimensional grid, and their movements are either random or follow the attractive or repulsive forces proportionally to the (dis-)similarities of neighboring

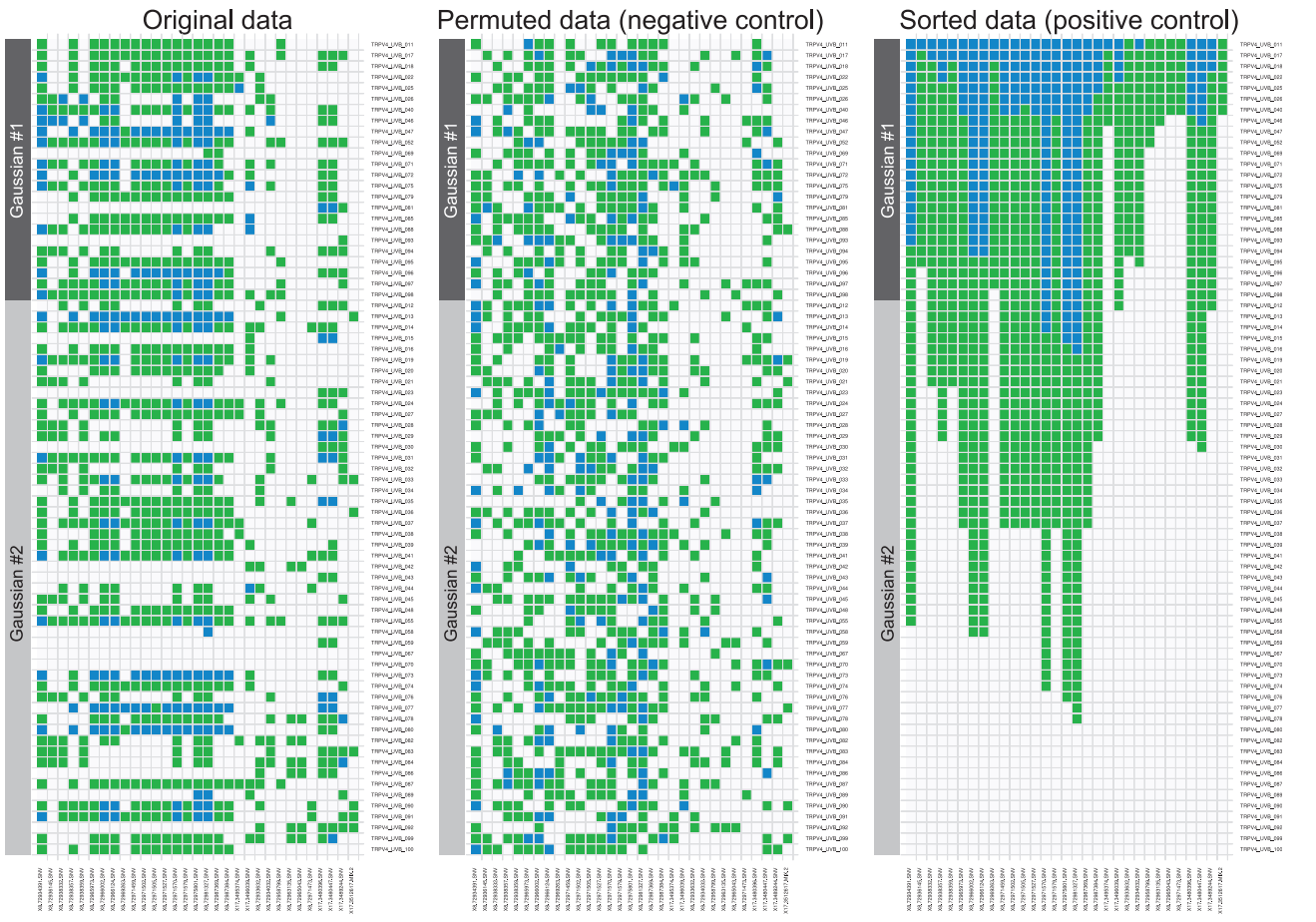


Figure 1. Patterns of the *TRPA1* (chromosome 8: X8) and *TRPV1* (chromosome 17: X17) genotypes observed in n = 75 healthy volunteers of Caucasian ethnicity for whom phenotype data of the heat hypersensitization after capsaicin application were available. The heat plot shows the occurrence of variants (columns) per subject (lines). The genetic information is color coded as the number of nonreference alleles found at the respective locus in the respective sample as white, 0 nonreference alleles = wild type genotype; green, heterozygous; and blue, 2 nonreference alleles. Thus, the individual genotypes are given by the vectors (rows) associated with each subject (subjects count at the right of each panel). The bar plot at the left shows the phenotype group association, with gray indicating Gaussian #1 and black indicating Gaussian #2 in Figure 2. The original genotype information (left) was permuted to obtain a negative control data set for the association of genotypes with phenotypes, and sorted in descending order of alleles at each gene locus to obtain a positive control data set for the genotype–phenotype association. The figure has been created with the R software package (version 3.4.1 for Linux; http://CRAN.R-project.org/)[60] using the library "gplots" (Warnes et al., https://cran.r-project.org/package=gplots).

"DataBots." Specifically, a parameter-free focusing projection method of a polar swarm, *Pswarm*, was used that exploits concepts of self-organization and swarm intelligence.[75] During construction of this type of projection, which is called the learning phase and requires an annealing scheme, structure analysis shifts from global optimization to local distance preservation (focusing). Intelligent agents of *Pswarm* operate on a toroid grid where positions are coded into polar coordinates allowing for a precise definition of their movement, neighborhood function, and annealing scheme. The size of the grid and, in contrast to other focusing projection methods,[17,89] the annealing scheme is data based and therefore, the method does not require any parameters. During learning, each DataBot searched for the strongest "scent,"[27] ie, for other agents that carried data with most similar features as it carried itself, by moving across the grid or staying in its current position, with a decreasing search radius.

After successful swarm learning, DataBots carrying items with similar features, ie, DataBots associated with similar data points, are placed in groups on the projection grid. The identification of emergent structures was enhanced on top of the learned structure. To this end, the distances between data points were calculated with the so-called U matrix[51,85] shown previously to provide emergent structures corresponding to clusters[51] and outperforming classic clustering methods.[84] Every value (height) in the U matrix depicts the average high-dimensional distance of a prototype to all immediate neighboring prototypes regarding a grid position. The corresponding visualization technique is a topographical map with hypsometric colors[76] facilitating the recognition of data structures. The calculations were performed using the R library "DatabionicSwarm" (Thrun M, https://cran.r-project.org/package=DatabionicSwarm).[76] Subsequently, clusters in the projected data were verified using the Ward method.[92] Finally, a possible association of the genotype-based clusters with the phenotype classes was assessed using the Fisher exact statistics.[21] In case of a positive association, this established that the genetic data were related to the pain phenotype, which was addressed in the next step of the data analysis.

2.8. Mapping of the genetic parameters to the phenotype classes

After establishment of a data structure in the genotype that reflected the phenotype structure, the association of the genotype with the phenotype was further analyzed. Therefore, in the fourth step of the data analysis, the question was pursued whether the phenotype can be predicted from the genotype. This was achieved by means of supervised machine learning, which addresses the data space $D = \{(x_i, y_i) | x_i \in X, y_i \in Y, i = 1, \ldots, n\}$ and tries to find a mapping of the input space X, comprising vectors $x_i = < x_{i,1}, \ldots, x_{i,d} >$ with $d > 0$ different parameters (here, the genetic information) acquired from $n > 0$ cases, to the output space Y, comprising y_i classes, eg, a pain-related phenotypes obtained through GMM and subsequent calculation of the Bayesian decision limits used for class separation.

In the present analysis, the mapping of the input space to the output space was performed using different methods of supervised machine learning, ie, (1) random forests,[6] (2) adaptive boosting,[66] (3) k-nearest neighbors (kNNs),[11] (4) naive Bayesian[2] classifiers, (5) support vector machines,[10] and (6) logistic regression,[91] which provided an internal validation of the results without the intention to compare the performances between machine-learning methods. The machine-learning methods were applied on the original data set and on the negative and positive control data set created as described above. The expectation was to observe a prediction of the phenotypes that were consistently better across several methods when using the original genotypes than when using the permuted genotypes, which should provide a classification performance not superior to guessing. In all 3 data sets, the classifiers were trained at training data subsets comprising 2/3 of the data, and subsequently their performance was estimated on the test data subset consisting of the remaining 1/3 of the data. This was repeated in 1000 cross-validation runs using Monte-Carlo[24] resampling and random splits of the original training data set into new training and test data subsets, using the R library "sampling" (https://cran.r-project.org/package=sampling).[78]

Random forests create sets of different, uncorrelated, and often very simple decision trees[6] with conditions on features as vertices and classes as leaves. The splits of the features are random and the classifier relates on the majority vote for class membership provided by a large number of decision trees. In the present analysis, 1000 decision trees were built containing sqrt(d) features, respectively, to nucleotide positions as the standard setting implemented in the R library "randomForest" (https://cran.r-project.org/package=randomForest).[41] The number of trees was heuristically based on visual analysis of the relationship between the number of decision trees and the classification accuracy, which indicated that beyond 100 trees, the classification balanced accuracy remained stable and a larger number merely consumed available computation time (Supplementary Fig. 1, available online at http://links.lww.com/PAIN/A561).

Boosting[66] approaches classification through a set of weak learners from which a single strong learner is created.[33] As weak classifiers served small classification and regression trees,[7] which provide a simple form of classification rules using the Gini impurity to find optimal (local) dichotomic decisions. In the present analysis, adaptive boosting as a successful algorithm for binary classification[67] was used, in which during the learning phase, subsequent weak learners are tweaked in favor of those data instances that had been misclassified by previous classifiers. Initially, each of n data point is associated with the same weight $w_i = 1/n$. A learner was trained to assign the correct class to each data point. Iteratively, the weights of misclassified data points were increased such that the subsequent learner gave more focus on the misclassified items. The final model combined all models using a weighted sum of the outputs that reflect the accuracy of all the constituent models. The number of iterations was heuristically based on the classification accuracy, which indicated no improvement beyond 500 runs, from which 1000 iterations were considered to provide robust results. These calculations were performed using the R package "ada" (http://cran.r-project.org/package=ada,[14] with the partitioning and classification package "rpart" https://cran.r-project.org/package=rpart).

The kNN classification[11] is a nonparametric method that belongs to the most frequently used algorithms in data science, although it is one of the basic methods in machine learning. During kNN model building, the entire labeled training data set is stored while a test case is placed in the feature space in the vicinity of the test cases at the smallest high-dimensional distance. The test case receives the class label according to the majority vote of the class labels of the k-training cases in its vicinity. In the present implementation, the size of k was established in resampling experiments with k set at 3 or 5. Even numbers for k intuitively make a majority vote on which the class assignment is based difficult when one of the nearest neighbors belongs to class 1 and the other to class 2. We tested 3 and 5 because these are often used and the default in various implementations of kNN. A silhouette plot would show the quality

of a clustering and to compare alternatives, eg, with different numbers of clusters. However, here, we used kNN as a classifier for a predefined number of classes (c = 2), not to obtain clusters or to reassess the number of classes in the data that had been obtained by means of GMM. At k = 3 and using the Manhattan distance[12] as used elsewhere[94] for genetic data, the best classification accuracy of the classifier was observed in 100 runs on randomly resampled data. Other distances such as the Euclidian, Jaccard, or Bray–Curtis distances, or more sophisticated implementations of nearest neighbor–based class assignment such as weighting or the use of kernel of different shapes were tried but did not provide any improvements regarding the basic version. These calculations were performed using the R package "KernelKnn" (Mouselimis L, https://cran.r-project.org/package=KernelKnn).

Bayesian classifiers were used that provide the probability that a data point being assigned to a specific class calculated by application of the Bayes' theorem.[2] In naive Bayesian classifiers, the oversimplified assumption is included that all features are conditionally independent of each other, which is a widely used technique to assign class labels to the samples from the available set of features, describing a special case of the more general Bayesian network model. The calculations were performed using the R package "e1071" (Meyer D, https://cran.r-project.org/package=e1071).

Support vector machines are supervised learning methods that classify data mainly based on geometrical and statistical approaches used for finding an optimum decision surface (hyperplane) that can separate the data points of 1 class from those belonging to another class in the high-dimensional feature space.[10] Using a kernel function, the hyperplane is frequently selected in a way to obtain a tradeoff between minimizing the misclassification rate and maximizing the distance of the plane to the nearest properly classified data point. In the present analysis, a Gaussian kernel with a radial basis was used. The analyses were performed using the R library "kernlab" (https://cran.r-project.org/package=kernlab).[32]

Finally, logistic regression[91] was used to map the genotype information to the 2 phenotype classes. This accommodated the inclusion of a more classic data analysis method well known from statistics. Logistic regression estimates the probability of falling into a certain level of the categorical response given a set of predictors. The calculations were performed using the "glm" command and the "family = binomial" switch as implemented in the R "stats" package[60] provided with the basic installation of the software core package (http://www.R-project.org/). The performances of all classifiers were assessed on the test data subsets created during cross-validation and are reported as the median of the resampling runs. Finally, a classic χ^2 test–based genotype vs association was performed.

3. Results

3.1. Capsaicin sensitivity phenotype classes

Phenotype data (HPTs acquired before and after topical application of capsaicin) were complete from 82 subjects. For technical reasons, data from 18 subjects were incomplete and therefore, these subjects were excluded from all analyses. After capsaicin application, a right shift in the pain thresholds to heat stimuli, calculated as stimulus intensity $HPT_T = HPT - 32°C$ (**Fig. 2**), was observed. The shift was pronounced enough to place the cohort in the range of HPT values typical for neuropathic patients according to the reference values of the QST test battery.[52] That is, while at baseline, only 6 pathological values were observed; after capsaicin application, 78 of the 82 subjects displayed pathological HPT values.

Visual inspection of the probability density distribution (PDF) of the capsaicin effects, CapsEff, suggested a multimodal distribution (**Fig. 2**). This was statistically supported by a significant likelihood ratio test ($P = 1.87 \times 10^{-6}$) comparing the goodness of the fits of the PDF, estimated using the PDE, between a single Gaussian mode and a GMM using M = 2 modes. No more significant improvement of the fit was obtained when a further Gaussian was added, based on likelihood ratio tests ($P = 0.9403$ for M = 3 vs M = 2). A satisfactory fit by a GMM with M = 2 was also supported by the nonsignificant result of the Kolmogorov–Smirnov test ($P = 0.952$) and the visual inspection of the QQ plot (**Fig. 2**). The parameter values of the final GMM are provided in **Table 1**. Thus, the output space was structured into 2 classes containing n = 24 and 58 subjects with low or high hypersensitization response to heat after topical application of capsaicin, respectively.

3.2. Association of TRPV1 and TRPA1 genotypes with capsaicin sensitivity phenotypes

Next-generation sequencing data were obtained from 75 subjects distributed across phenotype classes in a proportion of n = 24 and n = 51. The genetic information initially comprised 278 loci wherein at least 1 subject an allele differing from the hg19 reference genomic sequence was observed. The *TRPA1* gene at chromosome 8 displayed 134 loci with variant alleles and the *TRPV1* gene at chromosome 17 displayed 144 loci with variant alleles. All variant alleles were observed at frequencies corresponding to the expectations from the Hardy–Weinberg equilibrium (Fisher exact tests: P always > 0.05). After feature selection based on the Shannon information criterion and the ABC analysis of the Chi2 statistics for phenotype group differences (**Fig. 3**), d = 31 genetic features remained in the data set comprising 25 variants in the *TRPA1* gene and 6 loci in the *TRPV1* gene (**Fig. 1**) with different putative molecular functional consequences (**Table 2**). This corresponded to the size of the genetic features used in a previous study with comparable data analysis.[39] The frequencies of the minor alleles, ie, those disagreeing with the hg19 reference genomic sequence, in the analyzed data set ranged between 5% and 61%, with a median of 28%.

3.3. Genetic marker pattern and its relation to the phenotype classes

Unsupervised machine learning, aiming at data structure detection, was applied to analyze the 75 × 31-sized matrix comprising d = 31 genetic variants acquired in n = 75 subjects. Training of a swarm of intelligent data bots provided a structure-preserving projection of the high-dimensional data space $D = \{x_i, i = 1, \ldots, n\} \subset \mathbb{R}^d$ onto a 2-dimensional toroid projection grid (**Fig. 4**). After addition of the U matrix, a cluster structure emerged from the separation of the data bots carrying the genetic information into 2 distinct groups as visually indicated by a "mountain range" on the topographic map analogy (**Fig. 4** top). This was verified by Ward clustering that indicated 2 clusters differing with respect to the pattern of genetic variants (**Fig. 4**). Finally, the cluster membership was found to be unequally distributed among the phenotypes (the Fisher exact test: $P = 0.01199$), ie, the swarm-based cluster #1 comprising subjects carrying few variant alleles was underrepresented in phenotype cluster (Gaussian) #1 comprising subjects with low heat

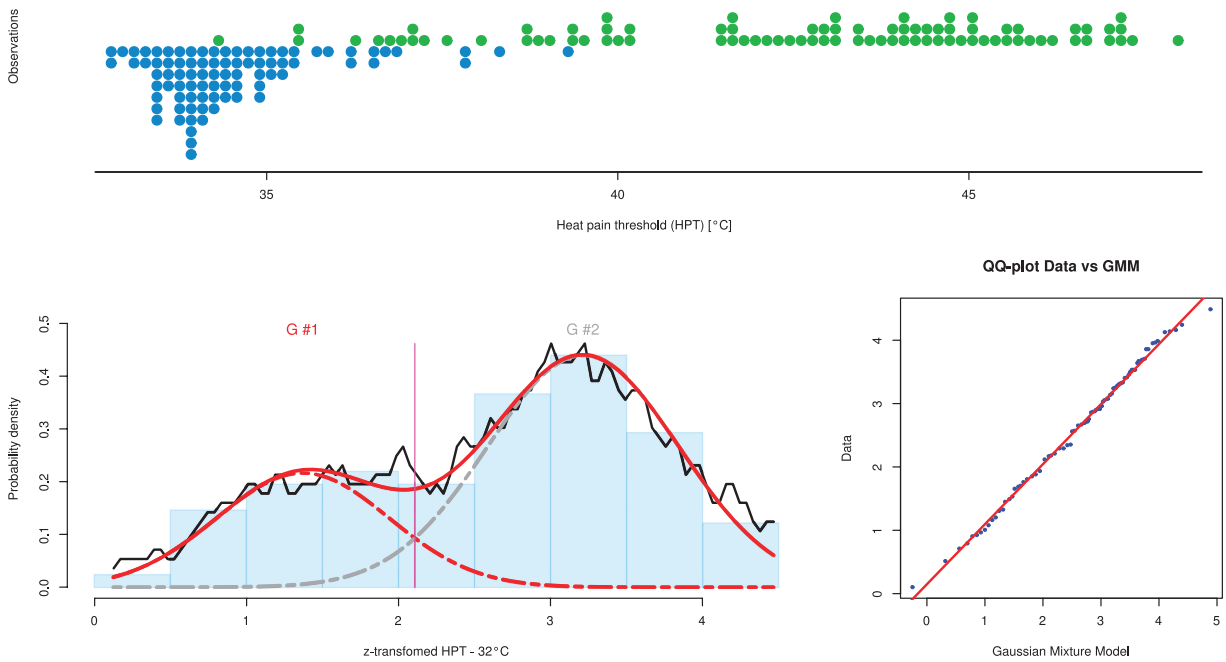


Figure 2. Original heat pain thresholds (HPTs) and distribution of the effects of capsaicin. Top: One-dimensional scatter plot of the observed individual heat pain sensitivity (dots; raw data). At the upper half (green dots), the values acquired at baseline are shown, whereas at the lower half, the values acquired after topical application of capsaicin are shown (blue dots). Bottom: The distribution of the capsaicin effects, obtained from the *z*-transformed HPTs according to the QST standard procedure[52] as CapsEf = $z\text{HPT}_{\text{T,capsaicin}} - z\text{HPT}_{\text{T,baseline}}$ and shown as probability density function (PDF) estimated by means of the Pareto density estimation (PDE[81]; black line) overlaid on a histogram could be fitted using a Gaussian mixture model (GMM) given as $P(x) = \sum_{i=0}^{M} w_i N(x|m_i,)$, with M = 2 modes. The fit is shown as a red line and the M = 2 mixes are indicated as differently colored dashed lines (G #1–#2). The Bayesian boundary between the Gaussians is indicated as a perpendicular magenta line. At the right side, a quantile–quantile (QQ) plot is shown comparing the observed distribution of cold pain data (ordinate) with the distribution expected from the GMM (abscissa). The blue dots symbolize the quantiles of observed data vs predicted data and the red line indicates identity, ie, the agreement between the data distribution expected from the model with the observed data distribution. The close vicinity of the dots to this line indicates satisfactory fits of the data by the respective GMM. The figure has been created using the R software package (version 3.4.1 for Linux; http://CRAN.R-project.org/)[60]; in particular, the dot plot was drawn using the R library "beeswarm" (Eklund A, https://cran.r-project.org/package=beeswarm) and the GMM plots were obtained using our package "AdaptGauss" (https://cran.r-project.org/package=AdaptGauss).[86] QST, quantitative sensory testing.

sensitivity and hypersensitization response (**Fig. 4**). This supported further exploration of the genetic information for relevance for the phenotypic classification.

3.4. Mapping of the genetic parameters to the phenotype classes

After establishment of a relation between the *TRPA1* and *TRPV1* NGS-based genetic patterns with the phenotype classes, the genotype–phenotype association was further analyzed. The classic χ^2-based genotype vs phenotype association analysis was negative, ie, only the 2 *TRPA1* variants X8.72934391.SNV and X8.72969263.SNV differed in allelic distribution between phenotype groups, but only at the uncorrected α level (**Fig. 3**) while when corrected according to Bonferroni,[3] the α level of 0.0016 resulting for the d = 31 genetic variants was exceeded for all gene loci.

Subsequently, supervised machine leaning was applied in cross-validation experiments using 1000 Monte-Carlo random resamplings of 2/3 vs (new training) 1/3 (new test) of the data provided the consistent observation that when using the true *TRPA1* and *TRPV1* NGS genotypes, the class assignment was better than that obtained with the permuted and therefore meaningless genotype information (**Fig. 5**). With the best median classification accuracy with the true genotypes of 62.5% (**Table 3**; n = 14, 5, 3, and 3 true positives, false positives, false negatives, and true negatives, respectively, as the average confusion matrix across the 1000 model runs), obtained with random forest and the best median classification accuracy obtained with the permuted genotype data of 50.7%, the improvement was almost by 1/8. However, the classification improvement associated with the true genotype data over the permuted data was small as compared to that obtained with the sorted, ie, positive control data (**Table 3**).

Finally, random forests allowed convenient access to the features' relative importance, which was numerically provided as mean decrease in classification accuracy when the respective feature (gene locus) was omitted from forest building (**Fig. 3**). The feature ranking pointed at *TRPA1* variants as most important,

Table 1

Parameter values of Gaussian mixture models (GMMs) applied as $P(\mathbf{x}) = \sum_{i=0}^{M} w_i N(\mathbf{x}|m_i, s_i)$ where m_i, s_i, and w_i are the parameters mean, SD, and relative weight of each of the Gaussians, i, respectively, obtained in the fit of the probability density distributions the pain thresholds to heat stimuli, calculated from the *z* values of the heat pain thresholds $z\text{HPT}_T = z(\text{HPT} - 32°\text{C})$ as CapsEf = $z\text{HPT}_{\text{T,capsaicin}} - z\text{HPT}_{\text{T,baseline}}$.

GMM parameter	i = 1 (Gaussian 1)	i = 2 (Gaussian 2)
m_i [$z\text{HPT}_{\text{T,capsaicin}} - z\text{HPT}_{\text{T,baseline}}$]	1.371	3.215
s_i	0.567	0.629
w_i	0.307	0.693
Bayesian decision limit [$z\text{HPT}_{\text{T,capsaicin}} - z\text{HPT}_{\text{T,baseline}}$]	2.107	

A mixture of M = 2 Gaussians (Fig. 2) was found to provide the best fits, as indicated by likelihood ratio tests.
HPT, heat pain threshold.

Table 2

Genetic variants that after feature selection were included in the genotype–phenotype associations, and their potential biological consequences as queried from several publicly available databases (NCBI gene index database at http://www.ncbi.nlm.nih.gov/gene; GeneCards at http://www.genecards.org, Short Genetic Variations database [dbSNP] at https://www.ncbi.nlm.nih.gov/snp and the "1000 Genomes Browser" at https://www.ncbi.nlm.nih.gov/variation/tools/1000genomes; all accessed in August 2017).

Gene	Variant	DNA change	Molecular consequence	dbSNP ID	Region score	TSS score	Unmatched score
TRPA1	X8.72933632.SNV	C>T	3 prime UTR variant	rs6996723	0.34	0.63	0.78
	X8.72934032.SNV	A>G	3 prime UTR variant	rs7827617	—	—	—
	X8.72934391.SNV	G>T	3 prime UTR variant	rs9298197	0.34	0.7	0.77
	X8.72936145.SNV	T>C	Missense variant	rs959976	0.29	0.07	0.36
	X8.72938332.SNV	G>A	Intron variant	rs2305017	0.25	0.19	0.22
	X8.72938357.SNV	A>C	Intron variant	rs2305018	0.24	0.17	0.18
	X8.72938359.SNV	T>C	Intron variant	rs2305019	0.23	0.2	0.17
	X8.72958799.SNV	G>A	Synonymous variant	rs61757563	0.23	0.1	0.37
	X8.72963135.SNV	A>G	Intron variant	rs1025927	0.23	0.28	0.07
	X8.72965043.SNV	T>C	Intron variant	rs13271151	0.51	0.68	0.79
	X8.72965973.SNV	G>A	Intron variant	rs3735942	0.21	0.2	0.49
	X8.72966002.SNV	G>A	Synonymous variant	rs3735943	0.32	0.25	0.44
	X8.72966124.SNV	A>G	Intron variant	rs3735944	0.39	0.38	0.54
	X8.72969263.SNV	A>C	Intron variant	rs3779752	0.49	0.31	0.31
	X8.72971459.SNV	T>C	Intron variant	rs12541196	0.23	0.17	0.1
	X8.72971473.SNV	C>T	Intron variant	rs71525150	0.15	0.19	0.11
	X8.72971502.SNV	T>C	Intron variant	rs12541199	0.26	0.2	0.1
	X8.72971505.SNV	T>G	Intron variant	rs12541200	0.14	0.17	0.15
	X8.72971527.SNV	C>T	Intron variant	rs12548486	—	—	—
	X8.72971570.SNV	C>T	Intron variant	rs9298198	0.12	0.12	0.18
	X8.72971579.SNV	A>C	Intron variant	rs11423229	0.27	0.39	0.18
	X8.72975801.SNV	T>A	Missense variant	rs7819749	0.27	0.37	0.49
	X8.72981327.SNV	A>G	Synonymous variant	rs1811457	0.15	0.32	0.46
	X8.72987369.SNV	C>T	Intron variant	rs2278654	0.16	0.46	0.87
	X8.72987384.SNV	A>T	Intron variant	rs2278653	0.17	0.52	0.82
TRPV1	X17.3480396.SNV	C>T	Intron variant	rs8078936	0.21	0.23	0.22
	X17.3480447.SNV	T>C	Missense variant	rs8065080	0.18	0.29	0.43
	X17.3489244.SNV	G>A	Intron variant	rs161394	0.18	0.11	0.21
	X17.3495374.SNV	G>A	Missense variant	rs222749	0.42	0.42	0.76
	X17.3496039.SNV	C>T	5 prime UTR variant	rs729271	0.44	0.26	0.18
	X17.3512617.MIX.2	—	Deletion/Insertion	rs775128810	—	—	—

The putative functional consequences according to 65 are amino acid or protein changes for missense and deletion/insertion variants, and reduced transcriptional efficiency for UTR and synonymous exonic variants. At the right of the tables, the values of 3 scores are provided by the genome-wide annotation of variants tool (GWAVA; at http://www.sanger.ac.uk/sanger/StatGen_Gwava)[62] that generates 3 different so-called GWAVA scores, ie, the "region score," the "TSS score," and the "unmatched" score, all in the range [0,…,1]. A high GWAVA score means more active functionality with respect to a low GWAVA score.
MIX, A mixture of variation types; SNV, single-nucleotide variation; TSS, transcription start site.

whereas the first *TRPV1* variants figured only at rank 7 among the classification-relevant gene loci. This observation accompanied the results of the classic χ^2-based genotype vs phenotype association analysis, in which only the 2 *TRPA1* variants X8.72934391.SNV and X8.72969263.SNV differed in allelic distribution between phenotype groups, however, only at the uncorrected α level (**Fig. 3**). These variants could also be used for phenotype class association; however, when eliminating them from the data set, a phenotype association was still consistently better than chance (**Table 3** and **Fig. 3B**), which supports that a complex genotype rather than a single variant modulated the phenotype.

4. Discussion

In the present analysis, several different methods of data analysis pointed toward a contribution of human TRP channel genotypes to the individual susceptibility to capsaicin-induced hypersensitization to heat stimuli. This was firstly hinted at by a high-dimensional pattern that emerged in the genotypes and could be statistically significantly associated with the 2 generated phenotype classes. Subsequently and most importantly, an importance of TRP genotypes for the heat pain–related phenotypes could be supported by the consistently better prediction of phenotypes from the genetic information than by chance, which was similarly observed across all machine-learned methods applied that always outperformed the phenotype class prediction when using randomly permuted genetic markers. Thus, the results can be summarized as an association of a complex TRP channel–related NGS genotype with the phenotype of the individual sensitivity to heat pain–related phenotypes.

The 31 genetic variants in the *TRPA1* and *TPRV1* genes that after feature selection were included in the association analyses, comprised 4 missense, 3 synonymous, and 1 deletion/insertion variation (**Table 2**), whereas the majority was located in introns or untranslated regions of the genes. The 2 polymorphism that differed in allelic distribution between phenotype classes at the uncorrected α level, ie, rs9298197 and rs3779752, and in addition, the rs2278654 variant that got the highest random-forest–based rank among all genetic loci, are located in noncoding areas of the *TRPA1* gene. Although they cannot affect the protein structure directly, recent studies in cancer tissue have highlighted the importance of noncoding variants and indeed, the majority of variants, both somatic and germline, had

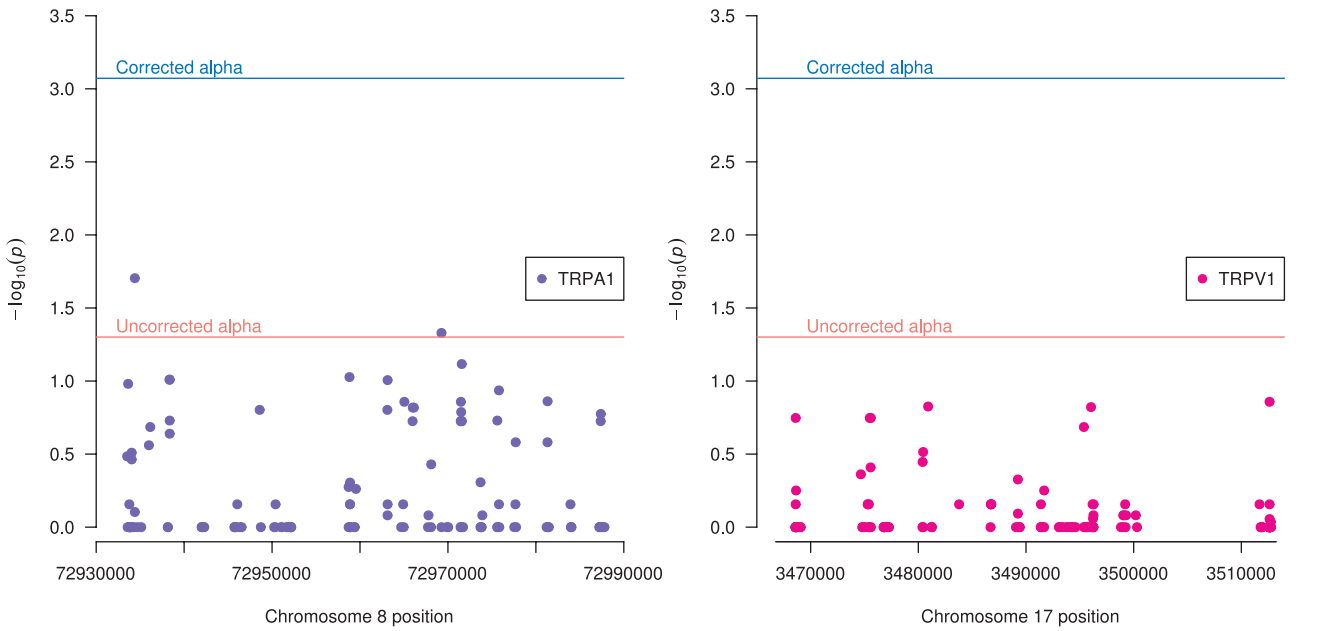


Figure 3. Dot plot of the results of the χ^2-based genotype association tests for d = 278 loci at the *TRPA1* (left panel) and *TRPV1* (right panel) genes. The α levels before (red) and after (blue) correction for multiple testing according to Bonferroni[3] are indicated as horizontal lines. A distribution differing between phenotypes above the uncorrected α level was observed for the variants X8.72934391.SNV and X8.72969263.SNV (Table 2). The figure has been created using the R software package (version 3.4.1 for Linux; http://CRAN.R-project.org/)[60] and the package "qqman" (https://cran.r-project.org/package=qqman).[78]

been observed in noncoding portions of the genome.[34] This observation implies that variants that affect the risk of complex diseases often exert their effect by altering the regulation of genes rather than by directly affecting the gene and protein function.[62] They act by affecting gene expression, eg, by disrupting a transcription factor–binding site[74] or by affecting mRNA splicing, stability, and structure, which may result in a reduced transcriptional efficiency.[22]

Along this line, to further assess the biological plausibility of a functional consequence of the present machine-learned derived selection of gene loci, the 31 selected variants were queried in the genome-wide annotation of variants tool (GWAVA, https://www.sanger.ac.uk/sanger/StatGen_Gwava). This web-based tool produces a prediction of the functional impact of noncoding genetic variants that are based on machine learning from a wide range of annotations of noncoding elements for which the functional consequences are known. For this task, it uses a tailored random-forest algorithm that builds 3 different classifiers, the so-called GWAVA scores named "region score," "TSS score," and "unmatched" score and all scaled in the range [0,…,1], by using all available annotations to discriminate between disease variants and variants from 3 control data sets.[62] Specifically, the "unmatched" classifier bases on a random selection of single-nucleotide variations (SNVs) from across the genome to get a reasonable sample of the background, the "TSS score" includes genome-wide variants matched for distance to the nearest transcription start site and the "region score" is composed of all variants in the 1 kb surrounding each of the disease variants. The machine-learned algorithm is trained with a set of variants with known function and learns to predict the function of further variants from their location within the gene. A high GWAVA score means more active functionality with respect to a low GWAVA score. The quality of the prediction was addressed in the original publication[62] where the authors showed that the classifier for each training set could usefully discriminate between disease and control variants. The area under the receiver operating characteristic curves were 0.97, 0.88, and 0.71, respectively, where a value of 0.5 denotes a bad classifier and 1 denotes an excellent classifier.

A GWAVA analysis for all the 134 gene loci in the *TRPA1* yielded 58 hits; 76 of the present variants that had not been reported previously were not found. Interestingly, the GWAVA tool found all but 2 of the 25 *TRPA1* variants included in the final analyses, which provides a first support for the potential importance, ie, for the successful of the applied machine-learned methods in selecting relevant gene loci for phenotype association (**Table 2**). Moreover, the 3 *TRPA1* variants highlighted by the random-forest classifier as most important (**Fig. 6**), ie, x8.72934392.SNV (rs9298197), x8.72969263.SNV (rs3779752), and x8.72987369.SNV (rs2278654) figured at the first or second positions of at least 1 of the GWAVA prediction scores (**Table 2**). This supports (1) the present data analysis approach and (2) the functional role of variants, although located in noncoding regions of the genes. Further variants included in the selection of d = 31 gene loci in the *TRPA1* and *TRPV1* genes could be supported by previous evidence of a functional role. This includes the *TRPA1* variants rs11988795, rs3735942, and rs3735943, which have been reported as associated with different sensitivity to pain,[35] or the *TRPA1* variant rs12548486, which has been associated with menthol preference among smokers.[79] In addition, the Ile585Val encoded by rs8065080 in the *TRPV1* gene has been reported to be associated with genetic risk of painful knee osteoarthritis,[87] and carriers of the *TRPA1* variant rs8065080 had a 1.6 time longer cold withdrawal time than noncarriers.[36,45] A further positive hit was the missense variant Lys186Asn (rs7819749) in *TRPV1*, which has been linked with glioblastoma multiforme.[1]

The pattern of variant alleles differed between phenotype groups in the direction that carriers of fewer variant alleles were underrepresented in the phenotype group with less pronounced changes of HPTs after topical application of capsaicin. Both directions of changes would seem biological plausible, and in particular, gain-of-function mutations in ion channels may lead to increased agonist sensitivity or altered gating properties, and may render the channel constitutively active.[5] For example, an

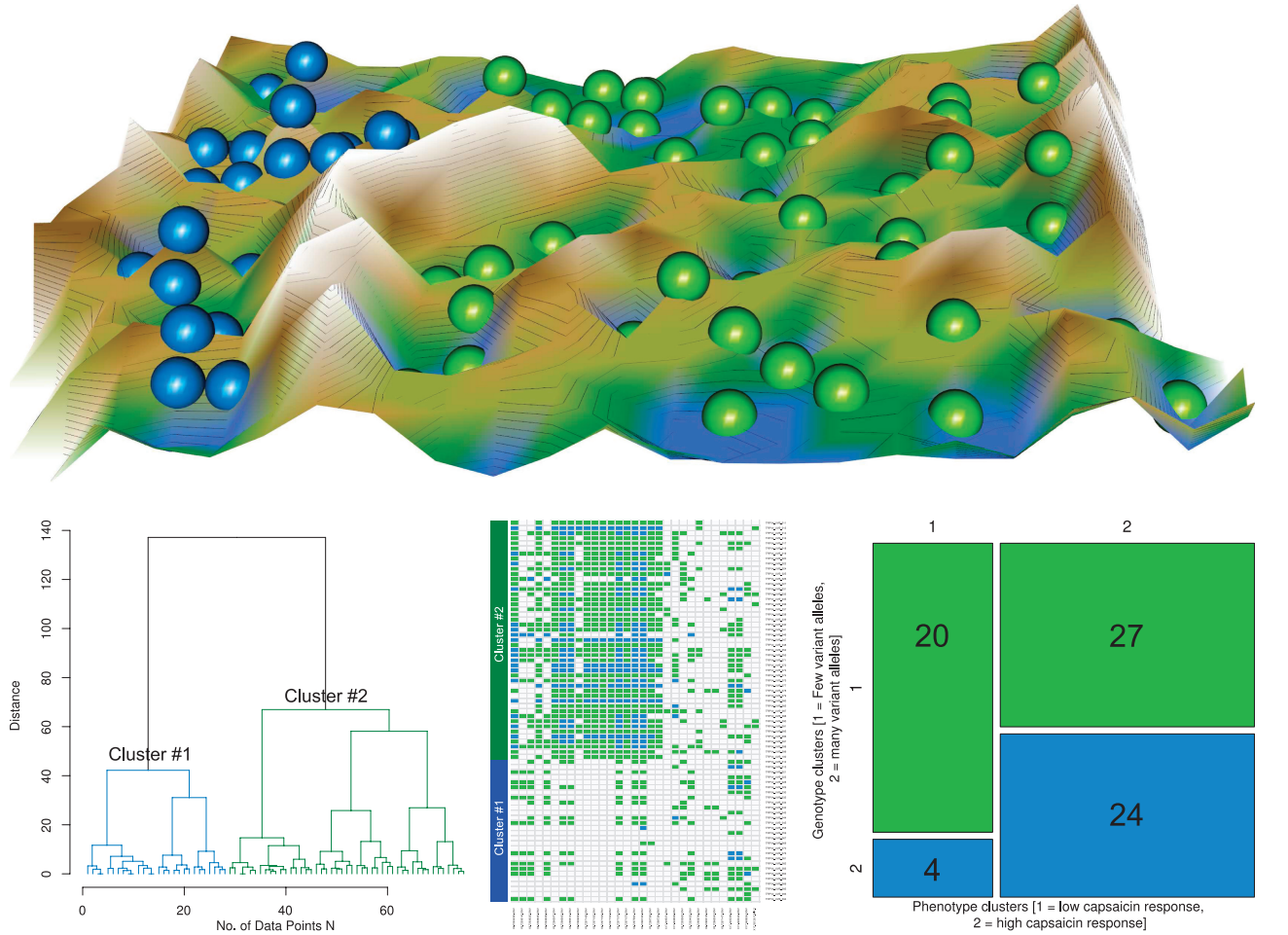


Figure 4. Data structure found in the *TRPA1/TRPV1* NGS genotypes and its relation with the phenotypes. Top: Visualization of high-dimensional data consisting of d = 31 gene loci analyzed in n = 75 subjects. The data were projected onto a 2-dimensional grid using a parameter-free projection polar swarm, *Pswarm*.[75] During the learning phase, the DataBots were allowed for adaptively adjusting their location on the grid close to DataBots carrying data with similar features, with successively decreasing search radius. When the algorithm ends, the DataBots become projected points. To enhance the emergence of data structures on this projection, a generalized U matrix displaying the distance in the high-dimensional space was added as a third dimension to this visualization.[75] The U matrix was colored in hypsometric colors[76] making the visualization appear as a geographical map with brown (up to snow-covered) heights and green valleys with blue lakes. Watersheds indicate borderlines between different groups of subjects according to the pattern of repeated cold pain measurements. The data points are colored according to the emerging 2-cluster structure. Bottom left: Ward clustering of the projected data clearly indicated 2 clusters using the Manhattan distance. Bottom center: Heat plot of the pattern of genetic variants (columns) per subject (lines), grouped for the data structure of the genetic information. The 75 × 31 matrix is a visualization of high-dimensional data consisting of d = 31 gene loci analyzed in n = 75 subjects. The allele occurrences are shown color coded as the number of nonreference alleles found at the respective locus in the respective sample as white, 0 nonreference alleles = wild type genotype; green, heterozygous; and blue, 2 non-reference alleles. Bottom right: Subjects belonging to the different genotype clusters were unevenly distributed across the phenotype clusters, ie, assignment to the 2 Gaussian modes in the distribution of capsaicin effects (Fig. 2), at a statistical significance level of $P < 0.05$ (the Fisher exact test). The mosaic plots represent the contingency table of the genotype vs phenotype class structure (membership sizes given as numbers in the fields of the mosaic). The figure has been created using the R software package (version 3.4.1 for Linux; http://CRAN.R-project.org/),[60] in particular the libraries "DatabionicSwarm" (M. Thrun, https://cran.r-project.org/package=DatabionicSwarm)[76] and "gplots" (Warnes G et al., https://cran.r-project.org/package=gplots). NGS, next-generation sequencing.

autosomal-dominant hereditary form of high-pain sensitivity, the so-called familial episodic pain syndrome, FEPS1 (accession number 615040 in the Online Mendelian Inheritance in Man (OMIM) database; http://www.ncbi.nlm.nih.gov/omim), which is characterized by episodes of debilitating upper-body pain, triggered by fasting and physical stress, is caused by a gain-of-function SNP (rs398123010) in the *TRPA1* gene.[37] In carriers, QST showed normal baseline sensory thresholds but enhanced secondary hyperalgesia to punctate stimuli after treatment with mustard oil.[37] Accordingly, this mutation increases the chemical sensitivity of TRPA1, but leaves the voltage sensitivity unchanged. Other gain-of-function mutations, rs753375978 and rs575489206, located in the analogous region of the *TRPV1* gene, severely affect all aspects of channel activation and lead to spontaneous activity.[5]

The more important role of TRPA1 as compared to that of TRPV1 in the sensitivity to heat or the hypersensitization response to capsaicin bears implications for the development of novel analgesic treatments[56] involving TRP channel inhibitors. Specifically, a query of the Thomson Reuters "Drugs and Biologics Search Tool" (http://integrity.thomsonpharma.com) in August 2017 indicated (**Table 4**) that by far, the most frequently regarded TRP channel family member in analgesic drug development is TRPV1, for which are 29 agonists or antagonists are currently under active development. TRPA1 agonists or antagonists figured with only 7 entries. If, based on the present results, the functional impact of *TRPA1* variants exceeds *TRPV1* variants, TRPA1 may play a greater role in pain including neuropathic pain when considering that topical capsaicin can induce a neuropathy-like QST results pattern in a small subgroup of healthy subjects.[46]

Table 3

Performance of classifiers obtained using different machine-learned methods (random forests, adaptive boosting, k-nearest neighbors [kNNs], naive Bayes, and support vector machines [SVMs]) on (1) the original data, (2) a reduced data set from which the 2 variants that differently distributed between phenotype groups at a noncorrected significance level (Fig. 3) and which alone provided a separation between phenotypes better than guessing (X8.72934391.SNV and X8.72969263.SNV) were left out, (3) a data set constructed to provide a negative control by permuting the original genotypes, and (4) a data set constructed to provide a positive control by sorting the genotype information in descending order of alleles at each gene locus (Fig. 1).

Parameter	Random forests			Boosting			kNN			Naive Bayes			SVM			Regression		
[%]	Original data	Permuted data (negative control)	Sorted data (positive control)	Original data	Permuted data (negative control)	Sorted data (positive control)	Original data	Permuted data (negative control)	Sorted data (positive control)	Original data	Permuted data (negative control)	Sorted data (positive control)	Original data	Permuted data (negative control)	Sorted data (positive control)	Original data	Permuted data (negative control)	Sorted data (positive control)
Sensitivity, recall	88.2	94.1	100	76.5	70.6	100	82.4	76.5	100	100	62.5	100	37.5	25	100	64.7	58.8	100
Specificity	37.5	12.5	100	37.5	25	100	37.5	25	100	11.8	29.4	100	70.6	76.5	100	50	37.5	87.5
Positive predictive value, precision	75	68	100	73.7	68	100	72.7	68.2	100	34.8	33.3	100	40	33.3	100	72.2	66.7	94.4
Negative predictive value	55.6	25	100	44.4	31.3	100	50	33.3	100	100	68.4	100	72.2	68.2	100	40	31.3	100
Balanced accuracy	62.5	50	97.1	59.9	50	97.1	59.6	50.4	100	55.9	50.7	100	57	50.4	97.1	57	48.9	93.8
Area under the ROC curve	62.5	50	97.1	59.9	50	97.1	59.6	50.4	100	55.9	50.7	100	57	50.4	97.1	57	48.9	93.8

Results represent the medians of the test performance measures from 1000 model runs using Monte-Carlo resampling with splits into 2/3 of the data (new training data) and 1/3 (new test data).
ROC, receiver operating characteristic.

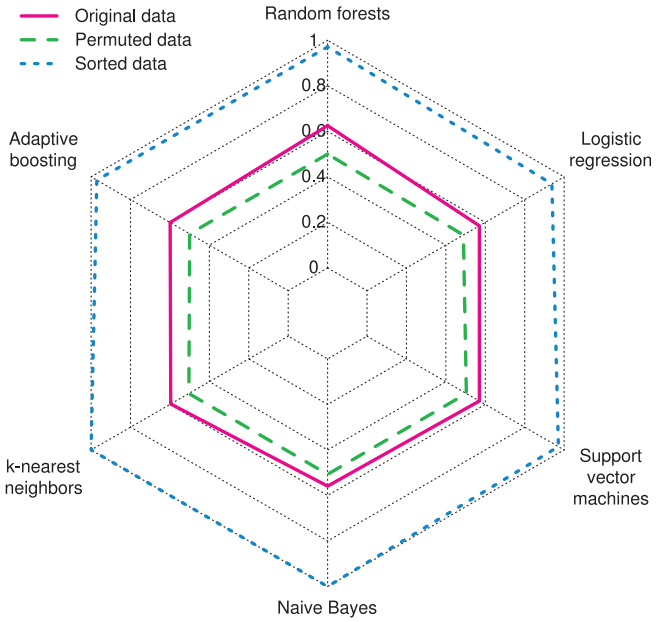


Figure 5. Radar plot of the balanced accuracy of different classifiers (random forests, adaptive boosting, k-nearest neighbors, naive Bayes, support vector machines, and logistic regression) to detect of a membership to the group with high response to capsaicin-induced hypersensitization against heat pain stimuli (Gaussian #2 in Fig. 2). The classification performance has been assessed in 1000 model runs using Monte-Carlo resampling runs with splits into 2/3 of the data (new training data) and 1/3 (new test data). The performance measures are comparatively shown for the results obtained on the original *TRPV1*/*TRPA1* NGS genotype and capsaicin sensitivity phenotype classes data set, on data constructed to provide as a negative control by permuting the genotypes, and on data constructed to provide a positive control by sorting the genotype information in descending order of alleles at each gene locus (Table 3). The plot shows the balanced accuracies in a spider web form. Each category, ie, machine-learning method, has a separate axis, scaled from 0% to 100% balanced accuracy. The axes are arranged in a circle in 360° evenly, and the values of each series are connected with lines indicating the results obtained with either of the 3 data sets, each with a different color. The figure has been created using the R software package (version 3.4.1 for Linux; http://CRAN.R-project.org/)[60] with the "radarchart" function provided in the library "fmsb" (Nakazawa M, https://cran.r-project.org/package=fmsb). NGS, next-generation sequencing.

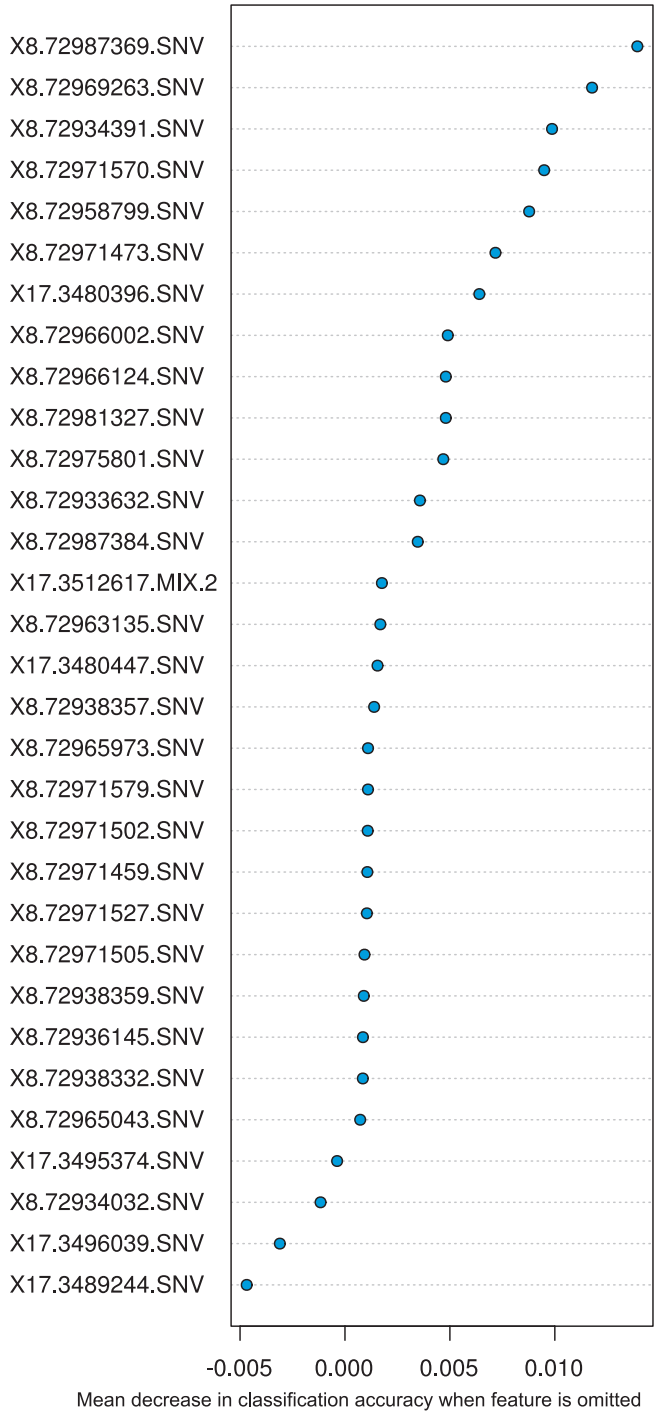


Figure 6. Importance of single-gene loci among the *TRPA1* (chromosome 8: X8) and *TRPV1* (chromosome 17: X17) genotypes for the random-forests–based classification into the 2 capsaicin hypersensitization phenotype groups (Fig. 2). The stripchart shows the importance of each gene locus, measured as the mean decrease in the classification accuracy when the respective feature is omitted from random-forests building. The figure has been created using the R software package (version 3.4.1 for Linux; http://CRAN.R-project.org/).[60] SNV, single-nucleotide variation.

The improvement of phenotype prediction provided by the *TRPA1* and *TRPV1* genotypes over a nonsense genotype was consistent yet modest when comparing the almost perfect phenotype association with an idealized arbitrary genotype. This points at further factors modulating the individual sensitivity to heat pain or the response to capsaicin, which is highly plausible and a monogenetic regulation of heat pain sensitivity or its enhancement by capsaicin was not expected considering the current knowledge about the complex genetic architecture of pain[19,44] and the role of competitive genotype effects not controlled for.[45] Indeed, although the present assessments had an explicit focus on *TRPV1* and *TRPA1*, further genetic variants are known to play a role in thermal pain sensitivity.[29] For example, the third heat transducer, TRPM3 (*TRPM3*), was not addressed in this study but may also contribute to heat pain sensitivity as shown in mice.[90] Furthermore, variants implicated in the present phenotype have been found in the genes coding for GTP cyclohydrolase 1 (*GCH1*),[8] for the melanocortin 1 receptor (*MC1R*)[16] or for the vasopressin receptor 1A (*AVPR1A*).[53] Furthermore, nongenetic factors play a role[40] up to the estimate that only 26% of the variance in heat pain responses can be explained by genetic factors.[29]

The present data-driven analyses were based on machine learning, which in its unsupervised form was applied to detect structures in the genetic data that hinted at a group separation, and in its supervised form was applied to assess the question whether the genotype provides information suitable for correct pain phenotype assignment. The methods were selected heuristically; possible alternatives such as self-organizing maps as used previously,[48,50] multidimensional scaling[4] t-SNE[88] or principal component analysis did not offer obvious advantages

Table 4

Novel drugs intended as analgesics targeting TRPV1 or TRPA1 ion channels, which are currently under active clinical development.

Drug	Action	Company
Zucapsaicin	TRPV1 agonist	Winston Laboratories
Resiniferatoxin		Icos
Capsaicin		Perrigo
Cannabidivarin		GW Pharmaceuticals
Etodolac		MEDRx
CGS-125		Vizuri Health Sciences
Hyaluronan		Vizuri Health Sciences
Diclofenac sodium		Boehringer Ingelheim
Propofol		Aspen Pharmacare
Axomadol		Grunenthal
Tivanisiran	TRPV1 expression inhibitor	Sylentis
Davasaicin	TRPV1 ligand	Dong-A
DWP-05195	TRPV1 antagonist	Daewoong
TR-1		Daewoong
V-116517		Purdue Pharma
JYL-1421		AmorePacific
NGD-8695		Ligand
NGD-8243		Ligand
NGD-9611		Ligand
Mavatrep		Johnson & Johnson
JTS-653		Japan Tobacco (JT)
DD-04107		BCN Peptides
AMG-51		Amgen
AMG-628		Amgen
ABT-102		Abbott
SAR-115740		Sanofi
AZD-1386		AstraZeneca
GRC-6211		Lilly
Catharanthine	TRPA1 agonist	University of Toronto
KDS-4103		Kadmus Pharmaceuticals
Cannabidivarin		GW Pharmaceuticals
ODM-108	TRPA1 modulator	Orion (F1)
HC-030031	TRPA1 antagonist	Hydra Biosciences
CB-625		Merck & Co
GRC-17536		Glenmark Pharmaceuticals

The information was queried on August 23, 2017, from the Thomson Reuters Integrity database at https://integrity.thomson-pharma.com.

over a swarm-based data projection. By contrast, it could not be excluded that methods mail fail such as on data that contain a cluster structure not separable using hyperplanes where multidimensional scaling may fail, or on data displaying high intrinsic data dimensionality were t-SNE is not recommended (eg, Figure 5.2 in Ref. 76), or on data not linearly separable where PCA has also been shown to fail in some settings where the swarm-based clustering was correct (eg, Figure 5.3 in Ref. 76). Similarly, the choice of supervised methods was heuristic; however, chosen to cover a variety of machine-learned classifiers previously used in pain research[49] such as prototype based (eg, kNN) or collective decision based (eg, boosting and random forests), with the addition of classic methods such as logistic regression included for its vicinity to classic statistical approaches, or naive Bayes. Indeed, the agreement among results obtained using different kinds of machine-learning methods and the biological plausibility of the results did not indicate an immediate need to include further methods.

The present analyses used machine learning for knowledge discovery, ie, an association of the *TRPV1* and *TRPA1* genetics with the heat-related pain phenotype was sought rather than a clinical tool for diagnostics. The moderate classification performance strongly suggests to base such a diagnostic tool on further factors including demographic, psychological and clinical parameters and factors derived from "omics," ie, proteomics, lipidomics, or genome-wide based features. Moreover, the present methods produced subsymbolic[71] classifiers where a better performance of machine-learned algorithm is sought by waiving the possibility to understand the details, ie, it is impossible to obtain complete biomedical explanations for the functioning of the algorithm. For example, random forests use hundreds or thousands of simple decision trees that escape interpretation; the classification is obtained through the complete set of trees, ie, the "forest."[6] The subsequently applied ranking of the importance of single variants only partly provided a biological explanation. With other classifiers, this was even less possible or completely impossible. However, the purpose of the present analysis was to study whether or not the genetic information contained in the sequences of *TRPV1* and *TRPA1* may contribute to the prediction of the phenotype, which establishes a genotype–phenotype association as the main purpose of this analysis.

5. Conclusions

In a cross-validated scenario, several analytical paths supported a role of *TRPA1* and, to a lower degree, *TRPV1* NGS-based genotypes for a potentially clinically relevant pain phenotype. The analysis shows that the complexity of the genotype is a relevant factor and machine-learned methods provide biologically plausible results, outperforming classic statistical genotype vs phenotype association analyses. The results were biologically plausible and fit with evidence of function *TRPA1* or *TRPV1* variants. Moreover, the relative importance of the variants observed with the machine learning agrees with an independent computer-based prediction of the biological roles of noncoding gene variants obtained in a GWAVA analysis. From this, a role of *TRPA1* or *TRPV1* NGS genotyping in personalized approaches at analgesic therapy with the respective novel analgesics may be expected. However, the improvement of phenotype prediction over chance was consistent but small when compared with a virtual extreme phenotype where most variant alleles were moved into a single phenotype group, which hints at further factors such as the genetics of other ion channels, generally pain-relevant genes[82] or nongenetic factors.

Conflict of interest statement

The authors have no conflict of interest to declare.

This work has been funded by the European Union Seventh Framework Programme (FP7/2013) under grant agreement no. 602919 (J.L., GLORIA) and by the Landesoffensive zur Entwicklung wissenschaftlich-ökonomischer Exzellenz (LOEWE),

LOEWE-Zentrum für Translationale Medizin und Pharmakologie (G.G. and J.L.).

Appendix A. Supplemental digital content

Supplemental digital content associated with this article can be found online at http://links.lww.com/PAIN/A561.

Article history:
Received 14 September 2017
Received in revised form 16 February 2018
Accepted 15 March 2018
Available online 27 March 2018

References

[1] Backes C, Harz C, Fischer U, Schmitt J, Ludwig N, Petersen BS, Mueller SC, Kim YJ, Wolf NM, Katus HA, Meder B, Furtwängler R, Franke A, Bohle R, Henn W, Graf N, Keller A, Meese E. New insights into the genetics of glioblastoma multiforme by familial exome sequencing. Oncotarget 2015;6:5918–31.
[2] Bayes M, Price M. An essay towards solving a problem in the doctrine of chances. By the late rev. Mr. Bayes, F. R. S. Communicated by Mr. Price, in a Letter to John Canton, A. M. F. R. S. Philosophical Trans 1763;53: 370–418.
[3] Bonferroni CE. Teoria statistica delle classi e calcolo delle probabilita. Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commerciali di Firenze 1936;8:3–62.
[4] Borg I, Groenen P. Modern multidimensional scaling: theory and applications. New York: Springer, 2005.
[5] Boukalova S, Touska F, Marsakova L, Hynkova A, Sura L, Chvojka S, Dittert I, Vlachova V. Gain-of-function mutations in the transient receptor potential channels TRPV1 and TRPA1: how painful? Physiol Res 2014;63 (suppl 1):S205–213.
[6] Breiman L. Random forests. Mach Learn 2001;45:5–32.
[7] Breimann L, Friedman JH, Olshen RA, Stone CJ. Classification and regression trees. Boca Raton: Chapman and Hall, 1993.
[8] Campbell CM, Edwards RR, Carmona C, Uhart M, Wand G, Carteret A, Kim YK, Frost J, Campbell JN. Polymorphisms in the GTP cyclohydrolase gene (GCH1) are associated with ratings of capsaicin pain. PAIN 2009; 141:114–18.
[9] Clapham DE. TRP channels as cellular sensors. Nature 2003;426: 517–24.
[10] Cortes C, Vapnik V. Support-vector networks. Mach Learn 1995;20:273–97.
[11] Cover T, Hart P. Nearest neighbor pattern classification. IEEE Trans Inf Theor 1967;13:21–7.
[12] Craw S. Manhattan distance. In: Sammut C, Webb GI, editors. Encyclopedia of machine learning. Boston: Springer US, 2010. p. 639.
[13] Cross SA. Pathophysiology of pain. Mayo Clin Proc 1994;69:375–83.
[14] Culp M, Johnson K, Michailides G. ada: an R package for stochastic boosting. J Stat Softw 2006;17:27.
[15] Davis JB, Gray J, Gunthorpe MJ, Hatcher JP, Davey PT, Overend P, Harries MH, Latcham J, Clapham C, Atkinson K, Hughes SA, Rance K, Grau E, Harper AJ, Pugh PL, Rogers DC, Bingham S, Randall A, Sheardown SA. Vanilloid receptor-1 is essential for inflammatory thermal hyperalgesia. Nature 2000;405:183–7.
[16] Delaney A, Keighren M, Fleetwood-Walker SM, Jackson IJ. Involvement of the melanocortin-1 receptor in acute pain and pain of inflammatory but not neuropathic origin. PLoS One 2010;5:e12498.
[17] Demartines P, Hérault J. CCA: "Curvilinear component analysis". Proceedings of the Proc 15° Colloque sur le traitement du signal et des images, Vol. 199: GRETSI, Groupe d'Etudes du Traitement du Signal et des Images, 1995.
[18] Dhar V. Data science and prediction. Commun ACM 2013;56:64–73.
[19] Diatchenko L, Nackley AG, Tchivileva IE, Shabalina SA, Maixner W. Genetic architecture of human pain perception. Trends Genet 2007;23: 605–13.
[20] Fischer MJ, Balasuriya D, Jeggle P, Goetze TA, McNaughton PA, Reeh PW, Edwardson JM. Direct evidence for functional TRPV1/TRPA1 heteromers. Pflugers Arch 2014;466:2229–41.
[21] Fisher RA. On the interpretation of Chi square from contingency tables, and the calculation of P. J R Stat Soc 1922;85:87–94.
[22] Fung KL, Gottesman MM. A synonymous polymorphism in a common MDR1 (ABCB1) haplotype shapes protein function. Biochim Biophys Acta 2009;1794:860–71.
[23] Fung KL, Pan J, Ohnuma S, Lund PE, Pixley JN, Kimchi-Sarfaty C, Ambudkar SV, Gottesman MM. MDR1 synonymous polymorphisms alter transporter specificity and protein stability in a stable epithelial monolayer. Cancer Res 2014;74:598–608.
[24] Good PI. Resampling methods: a practical guide to data analysis. Boston: Birkhäuser, 2006.
[25] Graffelman J. Exploring diallelic genetic markers: the HardyWeinberg package. J Stat Softw 2015;64:1–23.
[26] Hardy GH. Mendelian proportions in a mixed population. Science 1908; 28:49–50.
[27] Herrmann L, Ultsch A. The architecture of ant-based clustering to improve topographic mapping. In: DorigoM, BirattariM, BlumC, ClercM, StützleT, WinfieldAFT, editors. Ant colony optimization and swarm intelligence. 6th International Conference, ANTS 2008, Brussels, Belgium, 22–24 September, 2008 Proceedings. Berlin: Springer Berlin Heidelberg, 2008. p. 379–86.
[28] Ho KW, Ward NJ, Calkins DJ. TRPV1: a stress response protein in the central nervous system. Am J Neurodegener Dis 2012;1:1–14.
[29] Horjales-Araujo E, Dahl JB. Is the experience of thermal pain genetics dependent? Biomed Res Int 2015;2015:349584.
[30] Julius D, Basbaum AI. Molecular mechanisms of nociception. Nature 2001;413:203–10.
[31] Juran JM. The non-pareto principle; Mea culpa. Qual Prog 1975;8:8–9.
[32] Karatzoglou A, Smola A, Hornik K, Zeileis A. Kernlab—an S4 package for kernel methods in R. J Stat Softw 2004;11:1–20.
[33] Kearns M, Valiant LG. Crytographic limitations on learning Boolean formulae and finite automata. Proceedings of the 21st annual ACM symposium on Theory of computing. Seattle, WA: ACM, 1989. p. 433–44.
[34] Khurana E, Fu Y, Chakravarty D, Demichelis F, Rubin MA, Gerstein M. Role of non-coding sequence variants in cancer. Nat Rev Genet 2016;17: 93–108.
[35] Kim H, Mittal DP, Iadarola MJ, Dionne RA. Genetic predictors for acute experimental cold and heat pain sensitivity in humans. J Med Genet 2006; 43:e40.
[36] Kim H, Neubert JK, San Miguel A, Xu K, Krishnaraju RK, Iadarola MJ, Goldman D, Dionne RA. Genetic influence on variability in human acute experimental pain sensitivity associated with gender, ethnicity and psychological temperament. PAIN 2004;109:488–96.
[37] Kremeyer B, Lopera F, Cox JJ, Momin A, Rugiero F, Marsh S, Woods CG, Jones NG, Paterson KJ, Fricker FR, Villegas A, Acosta N, Pineda-Trujillo NG, Ramirez JD, Zea J, Burley MW, Bedoya G, Bennett DL, Wood JN, Ruiz-Linares A. A gain-of-function mutation in TRPA1 causes familial episodic pain syndrome. Neuron 2010;66:671–80.
[38] Kringel D, Sisignano M, Zinn S, Lötsch J. Next-generation sequencing of the human TRPV1 gene and the regulating co-players LTB4R and LTB4R2 based on a custom AmpliSeq panel. PLoS One 2017;12: e0180116.
[39] Kringel D, Ultsch A, Zimmermann M, Jansen JP, Ilias W, Freynhagen R, Griessinger N, Kopf A, Stein C, Doehring A, Resch E, Lötsch J. Emergent biomarker derived from next-generation sequencing to identify pain patients requiring uncommonly high opioid doses. Pharmacogenomics J 2017;17:419–26.
[40] Lariviere WR, McBurney DH, Frot M, Balaban CD. Tonic, phasic, and integrator components of psychophysical responses to topical capsaicin account for differences of location and sex. J Pain 2005;6:777–81.
[41] Liaw A, Wiener M. Classification and regression by randomForest. R News 2002;2:18–22.
[42] Lötsch J, Dimova V, Hermens H, Zimmermann M, Geisslinger G, Oertel BG, Ultsch A. Pattern of neuropathic pain induced by topical capsaicin application in healthy subjects. PAIN 2015;156:405–14.
[43] Lötsch J, Dimova V, Lieb I, Zimmermann M, Oertel BG, Ultsch A. Multimodal distribution of human cold pain thresholds. PLoS One 2015; 10:e0125822.
[44] Lötsch J, Doehring A, Mogil JS, Arndt T, Geisslinger G, Ultsch A. Functional genomics of pain in analgesic drug development and therapy. Pharmacol Ther 2013;139:60–70.
[45] Lötsch J, Fluhr K, Neddermayer T, Doehring A, Geisslinger G. The consequence of concomitantly present functional genetic variants for the identification of functional genotype-phenotype associations in pain. Clin Pharmacol Ther 2009;85:25–30.
[46] Lötsch J, Geisslinger G, Heinemann S, Lerch F, Oertel BG, Ultsch A. QST response patterns to capsaicin- and UV-B-induced local skin hypersensitization in healthy subjects: a machine-learned analysis. PAIN 2018;159:11–24.

[47] Lötsch J, Kringel D. Use of computational functional genomics in drug discovery and repurposing for analgesic indications. Clin Pharmacol Ther 2017. doi: 10.1002/cpt.960. [Epub ahead of print].
[48] Lötsch J, Lippmann C, Kringel D, Ultsch A. Integrated computational analysis of genes associated with human hereditary insensitivity to pain. A drug repurposing perspective. Front Neurosci 2017;10:252.
[49] Lotsch J, Ultsch A. Machine learning in pain research. PAIN 2018;159: 623–30.
[50] Lötsch J, Ultsch A. A machine-learned knowledge discovery method for associating complex phenotypes with complex genotypes. Application to pain. J Biomed Inform 2013;46:921–8.
[51] Lötsch J, Ultsch A. Exploiting the structures of the U-matrix. In: Villmann T,Schleif FM, Kaden M, Lange M, editors. Advances in intelligent systems and computing. Vol. 295. Heidelberg: Springer, 2014. p. 248–57.
[52] Magerl W, Krumova EK, Baron R, Tolle T, Treede RD, Maier C. Reference data for quantitative sensory testing (QST): refined stratification for age and a novel method for statistical comparison of group data. PAIN 2010; 151:598–605.
[53] Mogil JS, Sorge RE, LaCroix-Fralish ML, Smith SB, Fortin A, Sotocinal SG, Ritchie J, Austin JS, Schorscher-Petcu A, Melmed K, Czerminski J, Bittong RA, Mokris JB, Neubert JK, Campbell CM, Edwards RR, Campbell JN, Crawley JN, Lariviere WR, Wallace MR, Sternberg WF, Balaban CD, Belfer I, Fillingim RB. Pain sensitivity and vasopressin analgesia are mediated by a gene-sex-environment interaction. Nat Neurosci 2011;14:1569–73.
[54] Murphy KP. Machine learning: a probabilistic perspective. Cambridge, MA: The MIT Press, 2012.
[55] Pareto V. Manuale di economia politica, Milan: Società editrice libraria, revised and translated into French as Manuel d'économie politique. Paris: Giard et Briére, 1909.
[56] Patapoutian A, Tate S, Woolf CJ. Transient receptor potential channels: targeting pain at the source. Nat Rev Drug Discov 2009;8:55–68.
[57] Pearson K. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. Philosophical Mag Series 5 1900;50:157–75.
[58] Petersen KL, Rowbotham MC. A new human experimental pain model: the heat/capsaicin sensitization model. Neuroreport 1999;10:1511–16.
[59] Pfau D, Klein T, Blunk JA, Geber C, Krumova E, Limbeck C, Magerl W, Maier C, Westermann A, Schuh-Hofer S, Tiede W, Treede RD. QST Quantitative sensorische Testung, Handanweisung für den Untersucher, Eine standardisierte Testbatterie für die Quantitative Sensorische Testung nach den Regeln des Deutschen Forschungsverbundes Neuropathischer Schmerz (DFNS). In: Rolke R, Andrews A, Magerl W, Treede RD, editors. Mannheim, Germany: Lehrstuhl für Neurophysiologie, Universitätsmedizin Mannheim, 2010.
[60] R Development Core Team. R: a language and environment for statistical computing. Vienna: 2008.
[61] Reubish D, Emerling D, Defalco J, Steiger D, Victoria C, Vincent F. Functional assessment of temperature-gated ion-channel activity using a real-time PCR machine. Biotechniques 2009;47:iii–ix.
[62] Ritchie GR, Dunham I, Zeggini E, Flicek P. Functional annotation of noncoding sequence variants. Nat Methods 2014;11:294–6.
[63] Rolke R, Baron R, Maier C, Tolle TR, Treede RD, Beyer A, Binder A, Birbaumer N, Birklein F, Botefur IC, Braune S, Flor H, Huge V, Klug R, Landwehrmeyer GB, Magerl W, Maihofner C, Rolko C, Schaub C, Scherens A, Sprenger T, Valet M, Wasserka B. Quantitative sensory testing in the German Research Network on Neuropathic Pain (DFNS): standardized protocol and reference values. PAIN 2006;123:231–43.
[64] Rolke R, Magerl W, Campbell KA, Schalber C, Caspari S, Birklein F, Treede RD. Quantitative sensory testing: a comprehensive protocol for clinical trials. Eur J Pain 2006;10:77–88.
[65] Sauna ZE, Kimchi-Sarfaty C. Understanding the contribution of synonymous mutations to human disease. Nat Rev Genet 2011;12: 683–91.
[66] Schapire RE, Freund Y. A short introduction to boosting. J Jpn Soc Artif Intelligence 1999;14:771–80.
[67] Schapire RE, Freund Y. Boosting: foundations and algorithms. Cambridge, MA: The MIT Press, 2012.
[68] Schütz M, Oertel BG, Heimann D, Doehring A, Walter C, Dimova V, Geisslinger G, Lötsch J. Consequences of a human TRPA1 genetic variant on the perception of nociceptive and olfactory stimuli. PLoS One 2014;9:e95592.
[69] Shannon CE. A mathematical theory of communication. Bell Syst Techn J 1951;30:50–64.
[70] Smirnov N. Table for estimating the goodness of fit of empirical distributions. Ann Math Statist 1948;19:279–81.
[71] Smolensky P. On the proper treatment of connectionism. Behav Brain Sci 2010;11:1–23.
[72] Story GM, Peier AM, Reeve AJ, Eid SR, Mosbacher J, Hricik TR, Earley TJ, Hergarden AC, Andersson DA, Hwang SW, McIntyre P, Jegla T, Bevan S, Patapoutian A. ANKTM1, a TRP-like channel expressed in nociceptive neurons, is activated by cold temperatures. Cell 2003;112: 819–29.
[73] Swets JA. The relative operating characteristic in psychology: a technique for isolating effects of response bias finds wide use in the study of perception and cognition. Science 1973;182:990–1000.
[74] Teslovich TM, Musunuru K, Smith AV, Edmondson AC, Stylianou IM, Koseki M, Pirruccello JP, Ripatti S, Chasman DI, Willer CJ, Johansen CT, Fouchier SW, Isaacs A, Peloso GM, Barbalic M, Ricketts SL, Bis JC, Aulchenko YS, Thorleifsson G, Feitosa MF, Chambers J, Orho-Melander M, Melander O, Johnson T, Li X, Guo X, Li M, Shin Cho Y, Jin Go M, Jin Kim Y, Lee JY, Park T, Kim K, Sim X, Twee-Hee Ong R, Croteau-Chonka DC, Lange LA, Smith JD, Song K, Hua Zhao J, Yuan X, Luan Ja, Lamina C, Ziegler A, Zhang W, Zee RYL, Wright AF, Witteman JCM, Wilson JF, Willemsen G, Wichmann HE, Whitfield JB, Waterworth DM, Wareham NJ, Waeber G, Vollenweider P, Voight BF, Vitart V, Uitterlinden AG, Uda M, Tuomilehto J, Thompson JR, Tanaka T, Surakka I, Stringham HM, Spector TD, Soranzo N, Smit JH, Sinisalo J, Silander K, Sijbrands EJG, Scuteri A, Scott J, Schlessinger D, Sanna S, Salomaa V, Saharinen J, Sabatti C, Ruokonen A, Rudan I, Rose LM, Roberts R, Rieder M, Psaty BM, Pramstaller PP, Pichler I, Perola M, Penninx BWJH, Pedersen NL, Pattaro C, Parker AN, Pare G, Oostra BA, O'Donnell CJ, Nieminen MS, Nickerson DA, Montgomery GW, Meitinger T, McPherson R, McCarthy MI, McArdle W, Masson D, Martin NG, Marroni F, Mangino M, Magnusson PKE, Lucas G, Luben R, Loos RJF, Lokki ML, Lettre G, Langenberg C, Launer LJ, Lakatta EG, Laaksonen R, Kyvik KO, Kronenberg F, König IR, Khaw KT, Kaprio J, Kaplan LM, Johansson A, Jarvelin MR, Janssens ACJW, Ingelsson E, Igl W, Kees Hovingh G, Hottenga JJ, Hofman A, Hicks AA, Hengstenberg C, Heid IM, Hayward C, Havulinna AS, Hastie ND, Harris TB, Haritunians T, Hall AS, Gyllensten U, Guiducci C, Groop LC, Gonzalez E, Gieger C, Freimer NB, Ferrucci L, Erdmann J, Elliott P, Ejebe KG, Döring A, Dominiczak AF, Demissie S, Deloukas P, de Geus EJC, de Faire U, Crawford G, Collins FS, Chen YD, Caulfield MJ, Campbell H, Burtt NP, Bonnycastle LL, Boomsma DI, Boekholdt SM, Bergman RN, Barroso I, Bandinelli S, Ballantyne CM, Assimes TL, Quertermous T, Altshuler D, Seielstad M, Wong TY, Tai ES, Feranil AB, Kuzawa CW, Adair LS, Taylor HA, Borecki IB, Gabriel SB, Wilson JG, Holm H, Thorsteinsdottir U, Gudnason V, Krauss RM, Mohlke KL, Ordovas JM, Munroe PB, Kooner JS, Tall AR, Hegele RA, Kastelein JJP, Schadt EE, Rotter JI, Boerwinkle E, Strachan DP, Mooser V, Stefansson K, Reilly MP, Samani NJ, Schunkert H, Cupples LA, Sandhu MS, Ridker PM, Rader DJ, van Duijn CM, Peltonen L, Abecasis GR, Boehnke M, Kathiresan S. Biological, clinical and population relevance of 95 loci for blood lipids. Nature 2010;466:707–13.
[75] Thrun MC. Projection-based clustering through self-organization and swarm intelligence: combining cluster analysis with the visualization of high-dimensional data. Wiesbaden, Germany: Springer Fachmedien Wiesbaden, 2018.
[76] Thrun MC, Lerch F, Lötsch J, Ultsch A. Visualization and 3D printing of multivariate data of biomarkers. Proceedings of the International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision. 2016. p. 7–16.
[77] Treutlein J, Cichon S, Ridinger M, Wodarz N, Soyka M, Zill P, Maier W, Moessner R, Gaebel W, Dahmen N, Fehr C, Scherbaum N, Steffens M, Ludwig KU, Frank J, Wichmann HE, Schreiber S, Dragano N, Sommer WH, Leonardi-Essmann F, Lourdusamy A, Gebicke-Haerter P, Wienker TF, Sullivan PF, Nothen MM, Kiefer F, Spanagel R, Mann K, Rietschel M. Genome-wide association study of alcohol dependence. Arch Gen Psychiatry 2009;66:773–84.
[78] Turner SD. qqman: an R package for visualizing GWAS results using Q-Q and Manhattan plots. bioRxiv 2014. Cold Spring Harbor Laboratory Press. http://biorxiv.org/content/early/2014/05/14/005165.
[79] Uhl GR, Walther D, Behm FM, Rose JE. Menthol preference among smokers: association with TRPA1 variants. Nicotine Tob Res 2011;13: 1311–15.
[80] Ultsch A. Visualisation and classification with artificial life. In: Kiers HAL, Rasson JP, Groenen PJF, Schader M, editors. Data analysis, classification, and related methods. Berlin: Springer Berlin Heidelberg, 2000. p. 229–34.
[81] Ultsch A. Pareto density estimation: a density estimation for knowledge discovery. In: BaierD, WerrneckeKD, editors. Innovations in classification, data science, and information Systems. Proceedings of the 27th Annual Conference of the German Classification Society (GfKL), Cottbus, March 12-14, 2003. Heidelberg, Berlin: Springer Verlag, 2003.

[82] Ultsch A, Kringel D, Kalso E, Mogil JS, Lötsch J. A data science approach to candidate gene selection of pain regarded as a process of learning and neural plasticity. PAIN 2016;157:2747–57.
[83] Ultsch A, Lötsch J. Computed ABC analysis for rational selection of most informative variables in multivariate data. PLoS One 2015;10:e0129767.
[84] Ultsch A, Lötsch J. Machine-learned cluster identification in high-dimensional data. J Biomed Inform 2017;66:95–104.
[85] Ultsch A, Sieman HP. Kohonen's self organizing feature maps for exploratory data analysis. Proceedings of the INNC'90, Int Neural Network Conference, International Neural Network Conference, July 9–13, 1990, Palais des Congres, Paris, France. Kluwer, 1990. p. 305–8.
[86] Ultsch A, Thrun MC, Hansen-Goos O, Lötsch J. Identification of molecular fingerprints in human heat pain thresholds by use of an interactive mixture model R toolbox (AdaptGauss). Int J Mol Sci 2015;16:25897–911.
[87] Valdes AM, De Wilde G, Doherty SA, Lories RJ, Vaughn FL, Laslett LL, Maciewicz RA, Soni A, Hart DJ, Zhang W, Muir KR, Dennison EM, Wheeler M, Leaverton P, Cooper C, Spector TD, Cicuttini FM, Chapman V, Jones G, Arden NK, Doherty M. The Ile585Val TRPV1 variant is involved in risk of painful knee osteoarthritis. Ann Rheum Dis 2011;70:1556–61.
[88] van der Maaten LJP, Hinton GE. Visualizing high-dimensional data using t-SNE. J Mach Learn Res 2008;9:2579–605.
[89] Venna J, Peltonen J, Nybo K, Aidos H, Kaski S. Information retrieval perspective to Nonlinear dimensionality reduction for data visualization. J Mach Learn Res 2010;11:451–90.
[90] Vriens J, Owsianik G, Hofmann T, Philipp Stephan E, Stab J, Chen X, Benoit M, Xue F, Janssens A, Kerselaers S, Oberwinkler J, Vennekens R, Gudermann T, Nilius B, Voets T. TRPM3 is a nociceptor channel involved in the detection of noxious heat. Neuron 2011;70:482–94.
[91] Walker SH, Duncan DB. Estimation of the probability of an event as a function of several independent variables. Biometrika 1967;54: 167–79.
[92] Ward JH Jr. Hierarchical grouping to optimize an objective function. J Am Stat Assoc 1963;58:236–44.
[93] Wild A. Best practice in inventory management. New York: Wiley, 1997.
[94] You FM, Deal KR, Wang J, Britton MT, Fass JN, Lin D, Dandekar AM, Leslie CA, Aradhya M, Luo MC, Dvorak J. Genome-wide SNP discovery in walnut with an AGSNP pipeline updated for SNP discovery in allogamous organisms. BMC Genomics 2012;13:354.

Erklärung zu den Autorenanteilen an der Publikation:

A machine-learned analysis of human gene polymorphisms modulating persisting pain points to major roles of neuroimmune processes (*printed*)

Name der Zeitschrift: European Journal of Pain

Beteiligte Autoren: D Kringel, C Lippmann, M J Parnham, E Kalso, A Ultsch und J Lötsch

Was hat der Promovierende bzw. was haben die Koautoren beigetragen?

(1) zu Entwicklung und Planung

Promovierender: 30%
Autor EK: 20%
Autor JL: 50%

(2) zur Durchführung der einzelnen Untersuchungen und Experimente

Promovierender: 100% (Literatur Recherche, DNA Extraktion, NGS-Panel Design, Sequenzierung, Varianten Annotation, Methoden Validierung)

(3) zur Erstellung der Datensammlung und Abbildungen

Promovierender: 30% (Datenerhebung, Run-Statistik)
Autor JL: 70% (Erstellung der Abbildungen)

(4) zur Analyse und Interpretation der Daten

Promovierender: 30% (Datenvorverarbeitung, Charakterisierung und Assoziation der Varianten, Dateninterpretation)
Autor JL: 70% (Datenvisualisierung, Programmierung der KI, Anwendung der Klassifikatoren, Datenanalyse, Dateninterpretation)

(5) zum Verfassen des Manuskripts

Promovierender: 35%
Autor CL: 5%
Autor MJP: 10%
Autor EK: 10%
Autor AU: 5%
Autor: JL: 35%

Zustimmende Bestätigungen der oben genannten Angaben:

_____ _____

Datum/Ort Unterschrift Promovend

_____ _____

Datum/Ort Unterschrift Betreuer

EJP
European Journal of Pain

ORIGINAL ARTICLE

A machine-learned analysis of human gene polymorphisms modulating persisting pain points to major roles of neuroimmune processes

D. Kringel[1], C. Lippmann[2], M.J. Parnham[2], E. Kalso[3,4], A. Ultsch[5], J. Lötsch[1,2]

1 Institute of Clinical Pharmacology, Goethe - University, Frankfurt am Main, Germany
2 Fraunhofer Institute for Molecular Biology and Applied Ecology IME, Branch for Translational Medicine and Pharmacology TMP, Frankfurt
3 Institute of Clinical Medicine, University of Helsinki, Pain Clinic, Helsinki University Central Hospital, Helsinki, Finland
4 Institute of Biomedicine, Pharmacology, University of Helsinki, Helsinki, Finland
5 DataBionics Research Group, University of Marburg, Germany

Correspondence
Jörn Lötsch
E-mail: j.loetsch@em.uni-frankfurt.de

Funding sources
This work was funded by the European Union Seventh Framework Programme (FP7/2007 - 2013) under grant agreement no. 602919 (EK, JL, GLORIA) and in addition by the Landesoffensive zur Entwicklung wissenschaftlich-ökonomischer Exzellenz (LOEWE) Zentrum for Translational Medicine and Pharmacology project 'Process pharmacology: A data science based approach to drug repurposing' (JL). The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Conflicts of interest
The authors have declared that no further conflicts of interest exist.

Accepted for publication
13 June 2018

doi:10.1002/ejp.1270

Abstract

Background: Human genetic research has implicated functional variants of more than one hundred genes in the modulation of persisting pain. Artificial intelligence and machine-learning techniques may combine this knowledge with results of genetic research gathered in any context, which permits the identification of the key biological processes involved in chronic sensitization to pain.
Methods: Based on published evidence, a set of 110 genes carrying variants reported to be associated with modulation of the clinical phenotype of persisting pain in eight different clinical settings was submitted to unsupervised machine-learning aimed at functional clustering. Subsequently, a mathematically supported subset of genes, comprising those most consistently involved in persisting pain, was analysed by means of computational functional genomics in the Gene Ontology knowledgebase.
Results: Clustering of genes with evidence for a modulation of persisting pain elucidated a functionally heterogeneous set. The situation cleared when the focus was narrowed to a genetic modulation consistently observed throughout several clinical settings. On this basis, two groups of biological processes, the immune system and nitric oxide signalling, emerged as major players in sensitization to persisting pain, which is biologically highly plausible and in agreement with other lines of pain research.
Conclusions: The present computational functional genomics-based approach provided a computational systems-biology perspective on chronic sensitization to pain. Human genetic control of persisting pain points to the immune system as a source of potential future targets for drugs directed against persisting pain. Contemporary machine-learned methods provide innovative approaches to knowledge discovery from previous evidence.
Significance: We show that knowledge discovery in genetic databases and contemporary machine-learned techniques can identify relevant biological processes involved in Persitent pain.

1. Introduction

Persisting pain has a high prevalence (Elliott et al., 1999; Breivik et al., 2006; van Hecke et al., 2013) affecting a significant proportion of the world's population (Breivik et al., 2006). Its pathophysiology is incompletely understood, which is reflected in the limited success of available treatment options (Moore et al., 2009, 2012; Derry et al., 2012) and has stimulated intense research on this topic (Kringel and Lötsch, 2015). In this context, the study of human gene polymorphisms that modulate the persisting pain phenotype is an accepted research approach which has been pursued for more than 50 years (Godinova, 1965). A genetic background to persisting pain is clearly reflected by a protective effect against persisting pain exerted, for example, by a haplotype of the guanosine-5′-triphosphate (GTP) cyclohydrolase 1 gene (*GCH1*), originally reported to be composed of 15 genetic variants (Tegeder et al., 2006, 2008), or by a reduction in the perceived intensity of pain exerted, for example, by a deletion/insertion variant in the serotonin transporter gene-linked polymorphic region (5-HTTLPR; gene: *SLC6A4*) reportedly reducing the perception of heat pain (Lindstedt et al., 2011; Hooten et al., 2013; Kunz et al., 2016). On the other hand, increased risk for persisting pain is conferred, for example, by the rs12584920 variant of the 5-hydroxytryptamine receptor 2A gene (*HTR2A*) (Nicholl et al., 2011) or the rs734784 polymorphism in the potassium voltage-gated channel modifier, subfamily S member 1, gene (*KCNS1*) (Costigan et al., 2010).

Human genetic research during the last decade has identified many common variants of more than a hundred different genes spread across the genome that modulate the phenotype of persisting pain in several different clinical settings (Table 1). Thanks to concomitant developments in computer science, including progress in artificial intelligence, machine-learning and knowledge discovery in databases (Ashburner et al., 2000), the analysis of fundamental, complex biological functions has become increasingly possible. This allows persisting pain to be approached at a functional genomics level by combining the information on genetic modulation acquired in clinical studies with current knowledge of the function of human genes. This active research topic has already led to the identification of candidate genes for further clinical genetic pain research (Lötsch et al., 2013) and highlighted key pathophysiological processes of pain which may be targeted for future pharmacological treatment options (Ultsch et al., 2016).

In the present analysis, information on genes, for which empirical evidence indicates the existence of variants that modulate the clinical phenotype of persisting pain, was analysed at a functional genomics level. In this way, the biochemical, cellular and/or physiological properties of each and every gene product can be investigated to gain an understanding of the relationship between the genome and the phenotype on a global genomewide scale (Gibson and Muse, 2009). Applying machine-learned techniques (Fig. 1), the genes currently known to have relevance to human persisting pain were analysed for functional patterns that may provide insight into its pathophysiology based on current research activities, applying a data-driven approach without using prior hypotheses about the most important biological functions characterizing persisting pain. By applying methods of precisely calculated item selection (Ultsch and Lötsch, 2015), the present analysis aimed to identify a subset of genes that were most consistently reported to be involved in the modulation of persisting pain with a subsequent analysis of the main biological functions exerted by the products of these genes.

2. Methods

2.1 Search strategy

A set of genes relevant to persisting pain, based on published associations of their variants with phenotypic differences in persisting pain patients, was obtained from (1) a PubMed database (accessed in August 2016), by searches for the string '(chronic OR persisting OR neuropathic OR back OR inflammatory OR musculoskeletal OR visceral OR widespread OR idiopathic OR fibromyalgia) AND pain AND (polymorphism OR variant) NOT review' and (2) publications starting from the year 2000, which is close to the first sequencing of the human genome (Lander et al., 2001; Venter et al., 2001) marking the beginning of a new area of genetic research, and (3) published overviews on pain genetics (e.g. Edwards, 2006; Tegeder and Lötsch, 2009; Mogil, 2012; Zorina-Lichtenwalter et al., 2016). To avoid redundancies, reports of positive associations of any gene variant were included only once per clinical setting in the present analysis. This implies that not every variant found to be functionally associated with a persisting pain phenotype was taken into account. The resulting information for each gene, thus, comprised (1) a positive report of a gene

Table 1 List of 110 genes with empirically supported relevance to persisting pain, based on published evidence that their genetic variants are associated with phenotypic differences in persisting pain patients in several clinical settings.

Gene	NCBI	Type of pain	References	Gene	NCBI	Type of pain	References	Gene	NCBI	Type of pain	References
ACAN	176	Musculoskeletal	Kirk et al. (2003)	*GCH1*	2643	Widespread	Kim et al. (2010a)	*MTHFD1*	4522	Musculoskeletal	Aneiros-Guerrero et al. (2011)
ACE	1636	Musculoskeletal	Rommel et al. (2006)	*GDF5*	8200	Musculoskeletal	Valdes et al. (2011c)	*MTRR*	4552	Musculoskeletal	Aneiros-Guerrero et al. (2011)
ADRA1A	148	Neuropathic	Herlyn et al. (2010)	*GNB3*	2784	Visceral	Oshima et al. (2010)	*MYT1L*	23,040	Widespread	Docampo et al. (2014)
ADRA1D	146	Visceral	Sugaya et al. (2002)	*GRK5*	2869	Miscellaneous	Smith et al. (2011)	*NCR3*	259,197	Neuropathic	Sato et al. (2002)
ADRA2A	150	Idiopathic	Kim et al. (2004)	*GSTM1*	2944	Visceral	Wu et al. (2000)	*NPY*	4852	Back pain	Herlyn et al. (2010)
ADRA2C	152	Idiopathic	Kim et al. (2004)	*GSTM1*	2944	Musculoskeletal	Aneiros-Guerrero et al. (2011)	*NR3C1*	2908	Widespread	Holliday et al. (2010)
ADRB2	154	Widespread	Hocking et al. (2010)	*GSTP1*	2950	Visceral	Woo et al. (2010)	*NRXN3*	9369	Widespread	Docampo et al. (2014)
ADRB2	154	Musculoskeletal	Diatchenko et al. (2006)	*GSTT1*	2952	Visceral	Woo et al. (2010)	*NTRK1*	4914	Miscellaneous	Shatzky et al. (2000)
ADRB2	154	Neuropathic	Herlyn et al. (2010)	*HFE*	3077	Musculoskeletal	Alizadeh et al. (2007)	*OPRM1*	4988	Miscellaneous	Cheng et al. (2010)
ADRB3	155	Visceral	Sugaya et al. (2002)	*HLA-A*	3105	Neuropathic	Sato et al. (2002)	*P2RX7*	5027	Musculoskeletal	Sorge et al. (2012)
ANP32A	8125	Musculoskeletal	Valdes et al. (2009)	*HLA-B*	3106	Neuropathic	de Rooij et al. (2009)	*P2RX7*	5027	Neuropathic	Sorge et al. (2012)
APOE	348	Widespread	Reeser et al. (2011)	*HLA-B*	3106	Inflammatory	Gullo et al. (1982)	*PCSK6*	5046	Musculoskeletal	Malfait et al. (2012)
AR	367	Visceral	Shaik et al. (2009)	*HLA-C*	3107	Neuropathic	Ozawa et al. (1999)	*PGK1*	5230	Visceral	Riley and Krieger, 2002)
ASPN	54,829	Musculoskeletal	Nakamura et al. (2007)	*HLA-DQA1*	3117	Neuropathic	de Rooij et al. (2009)	*PGR*	5241	Visceral	De Carvalho et al. (2007)
CACNA2D3	55,799	Back pain	Neely et al. (2010)	*HLA-DQB1*	3119	Neuropathic	de Rooij et al. (2009)	*POMC*	5443	Widespread	Holliday et al. (2010)
CACNG2	10,369	Neuropathic	Nissenbaum et al. (2010)	*HLA-DRB1*	3123	Neuropathic	Sato et al. (2002)	*PRSS1*	5644	Inflammatory	Midha et al. (2010)
CALCA	796	Neuropathic	Herlyn et al. (2010)	*HTR2A*	3356	Visceral	Pata et al. (2004)	*PTGS1*	5742	Visceral	Arisawa et al. (2008)
CAMK4	814	Miscellaneous	Smith et al. (2011)	*HTR3E*	285,242	Visceral	Kilpatrick et al. (2011)	*SCN5A*	6331	Visceral	Saito et al., 2009)
CASP9	842	Back pain	Guo et al. (2011)	*IFNG*	3458	Inflammatory	Noponen-Hietala et al. (2005)	*SCN5A*	6331	Idiopathic	Reimann et al. (2010)
CCT5	22,948	Widespread	Peters et al. (2013)	*IFNG*	3458	Inflammatory	Oen et al. (2005)	*SCN9A*	6335	Back pain	Reimann et al. (2010)
CFTR	1080	Inflammatory	Midha et al. (2010)	*IFRD1*	3475	Miscellaneous	Smith et al. (2011)	*SCN9A*	6335	Inflammatory	Reimann et al. (2010)

Table 1 (*Continued*)

Gene	NCBI	Type of pain	References	Gene	NCBI	Type of pain	References	Gene	NCBI	Type of pain	References
CRHBP	1393	Widespread	Holliday et al. (2010)	*IL-10*	3586	Back pain	Shoskes et al. (2002)	*SCN9A*	6335	Miscellaneous	Reimann et al. (2010)
CILP	8483	Back pain	Seki et al. (2005)	*IL-10*	3586	Inflammatory	Noponen-Hietala et al. (2005)	*SCN9A*	6335	Musculoskeletal	Valdes et al. (2011a)
CNR1	1268	Visceral	Park et al. (2011)	*IL-10*	3586	Musculoskeletal	Oen et al. (2005)	*SCN10A*	6336	Neuropathic	Faber et al. (2012)
COL1A1	1277	Back pain	Tilkeridis et al. (2005)	*IL-16*	3603	Visceral	Gan et al. (2010)	*SCN11A*	11,280	Widespread	Leipold et al., 2013)
COL6A4P1	344,875	Musculoskeletal	Miyamoto et al. (2008)	*IL-1A*	3552	Back pain	Solovieva et al. (2004)	*SERPINA1*	5265	Widespread	Blanco et al. (2006)
COL9A2	1298	Back pain	Ala-Kokko, 2002)	*IL-1B*	3553	Back pain	Zhang et al. (2002)	*SERPINA6*	866	Widespread	Holliday et al. (2010)
COL9A3	1299	Back pain	Kales et al. (2004)	*IL-1B*	3553	Miscellaneous	Jeremias et al. (2000)	*SHMT1*	6470	Idiopathic	Aneiros-Guerrero et al. (2011)
COMT	1312	Musculoskeletal	van Meurs et al. (2009)	*IL-1R2*	7850	Neuropathic	Stephens et al. (2014)	*SLC6A4*	6532	Idiopathic	Herken et al. (2001)
COMT	1312	Widespread	Cohen et al. (2009)	*IL-1RN*	3557	Back pain	Kim et al. (2010b)	*SMAD3*	4088	Musculoskeletal	Valdes et al. (2010)
COMT	1312	Back pain	Dai et al. (2010)	*IL-1RN*	3557	Musculoskeletal	Attur et al. (2010)	*SOD2*	6648	Idiopathic	Arisan et al. (2006)
COMT	1312	Visceral	Karling et al. (2011)	*IL-1RN*	3557	Idiopathic	Witkin et al. (2002)	*SPINK1*	6690	Inflammatory	Midha et al. (2010)
COMT	1312	Idiopathic	Tahara et al. (2008)	*IL-2*	3558	Inflammatory	Noponen-Hietala et al. (2005)	*TAAR1*	134,864	Widespread	Smith et al. (2012)
CRH	1392	Widespread	Holliday et al. (2010)	*IL-4*	3565	Visceral	Sugaya et al. (2002)	*TAC1*	6863	Back pain	Herlyn et al. (2010)
CRHBP	1393	Musculoskeletal	Linnstaedt et al. (2016)	*IL-4*	3565	Inflammatory	Noponen-Hietala et al. (2005)	*TACR1*	6869	Visceral	Renner et al. (2009)
CRHR1	1394	Widespread	Holliday et al. (2010)	*IL-4R*	3566	Visceral	Sugaya et al. (2002)	*TGFB1*	7040	Back pain	Herlyn et al. (2010)
CYP2D6	1565	Visceral	Wu et al. (2000)	*IL-4R*	3566	Inflammatory	Noponen-Hietala et al. (2005)	*TGFB1*	7040	Inflammatory	Shoskes et al. (2002)
DIO2	1734	Musculoskeletal	Meulenbelt et al. (2008)	*IL-6*	3569	Back pain	Herlyn et al. (2010)	*TGFB1*	7040	Musculoskeletal	Oen et al. (2005)
DRD4	1815	Musculoskeletal	Aneiros-Guerrero et al. (2011)	*IL-6*	3569	Idiopathic	Shoskes et al. (2002)	*TNF*	7124	Back pain	Herlyn et al. (2010)
DRD4	1815	Widespread	Buskila et al. (2004)	*IL-6*	3569	Inflammatory	Noponen-Hietala et al. (2005)	*TNF*	7124	Idiopathic	Shoskes et al. (2002)
LPAR1	1902	Musculoskeletal	Mototani et al. (2008)	*IL-6*	3569	Neuropathic	Oen et al. (2005)	*TNF*	7124	Inflammatory	Noponen-Hietala et al. (2005)
ESR1	2099	Idiopathic	Ribeiro-Dasilva et al. (2009)	*KCNJ6*	3763	Back pain	Bruehl et al. (2013)	*TNF*	7124	Musculoskeletal	Oen et al. (2005)

Table 1 (*Continued*)

Gene	NCBI	Type of pain	References	Gene	NCBI	Type of pain	References	Gene	NCBI	Type of pain	References
ESR1	2099	Visceral	Govindan et al. (2009)	*KCNS1*	3787	Neuropathic	Costigan et al. (2010)	*TP53*	7157	Visceral	Ribeiro Junior et al. (2009)
ESR1	2099	Musculoskeletal	Kang et al. (2007)	*MAOA*	4128	Musculoskeletal	Gursoy et al. (2008)	*TPH2*	121,278	Widespread	Nicholl et al. (2011)
FAM173B	134,145	Widespread	Peters et al. (2013)	*MBL2*	4153	Idiopathic	Babula et al. (2004)	*TPH2*	121,278	Musculoskeletal	Nicholl et al. (2011)
FKBP5	2289	Musculoskeletal	Bortsov et al. (2013)	*MC1R*	4157	Idiopathic	Foster et al. (2004)	*TRPA1*	8989	Neuropathic	Binder et al. (2011)
GBP1	2633	Widespread	Smith et al. (2012)	*MC2R*	4158	Widespread	Holliday et al. (2010)	*TRPM8*	79,054	Neuropathic	Binder et al. (2011)
GC	2638	Visceral	Faserl et al. (2011)	*MIF*	4282	Visceral	Arisawa et al. (2007)	*TRPV1*	7442	Musculoskeletal	Valdes et al. (2011b)
GCH1	2643	Back pain	Doehring et al. (2009)	*SOD2*	6648	Visceral	Arisan et al. (2006)	*TRPV1*	7442	Neuropathic	Binder et al. (2011)

The gene, NCBI number, clinical setting of persisting pain are shown together with a key reference in which this association was reported. The studies are given grouped for the relevant gene, however, in arbitrary order of clinical settings or publication year.

modulation in persisting pain and (2) the clinical setting of this finding.

2.2 Data analysis

Data were analysed using the R software package (version 3.3.2 for Linux; http://CRAN.R-project.org/; R Development Core Team, 2008) on an Intel Xeon® computer running on Ubuntu Linux 16.04.1 64-bit. The analysis employed several methods of machine-learning that, as described previously (Lötsch and Ultsch, 2017a), may be referred to as a set of methods that can automatically detect patterns in data and then use the uncovered patterns to predict or classify future data, to observe structures such as subgroups in the data or to extract information from the data suitable to derive new knowledge (Murphy, 2012; Dhar, 2013). More detailed descriptions including definitions of key concepts have been provided elsewhere (Lötsch and Ultsch, 2017a).

The analysis aimed at describing the functional genomics of persisting pain based on the biological roles of the genes that reportedly carry variants modulating that phenotype. The biological roles were assessed as biological processes in which the genes are involved, defined as series of events or molecular functions with a defined beginning and end (Ashburner et al., 2000) and queried from the Gene Ontology (GO) knowledgebase that provides the acquired knowledge about the biological functions of gene products, described with a controlled vocabulary of GO terms (Ashburner et al., 2000).

Thus, the basis on which the present functional picture of persisting pain was created consisted of the biological processes in which the genes carrying modulatory variants were reported to be involved. The functional picture of persisting pain was sought pursuing two different analytical paths (Fig. 1). In a *first* approach, functional subgroups were sought in the set of human genes, variants of which have been associated with modulation of the clinical phenotype of persisting pain. This was addressed by applying a clustering algorithm on the matrix given by the genes versus their annotated biological processes; an approach that previously proofed as suitable for gene function based classifications (Lötsch and Ultsch, 2016a). In a *second* approach, the hypothesis was pursued that the functional genomics of persisting pain will prevail across clinical settings irrespective of the disease that had originally triggered the process. Therefore, the most informative subsets of the genes were identified using a computed item categorization technique

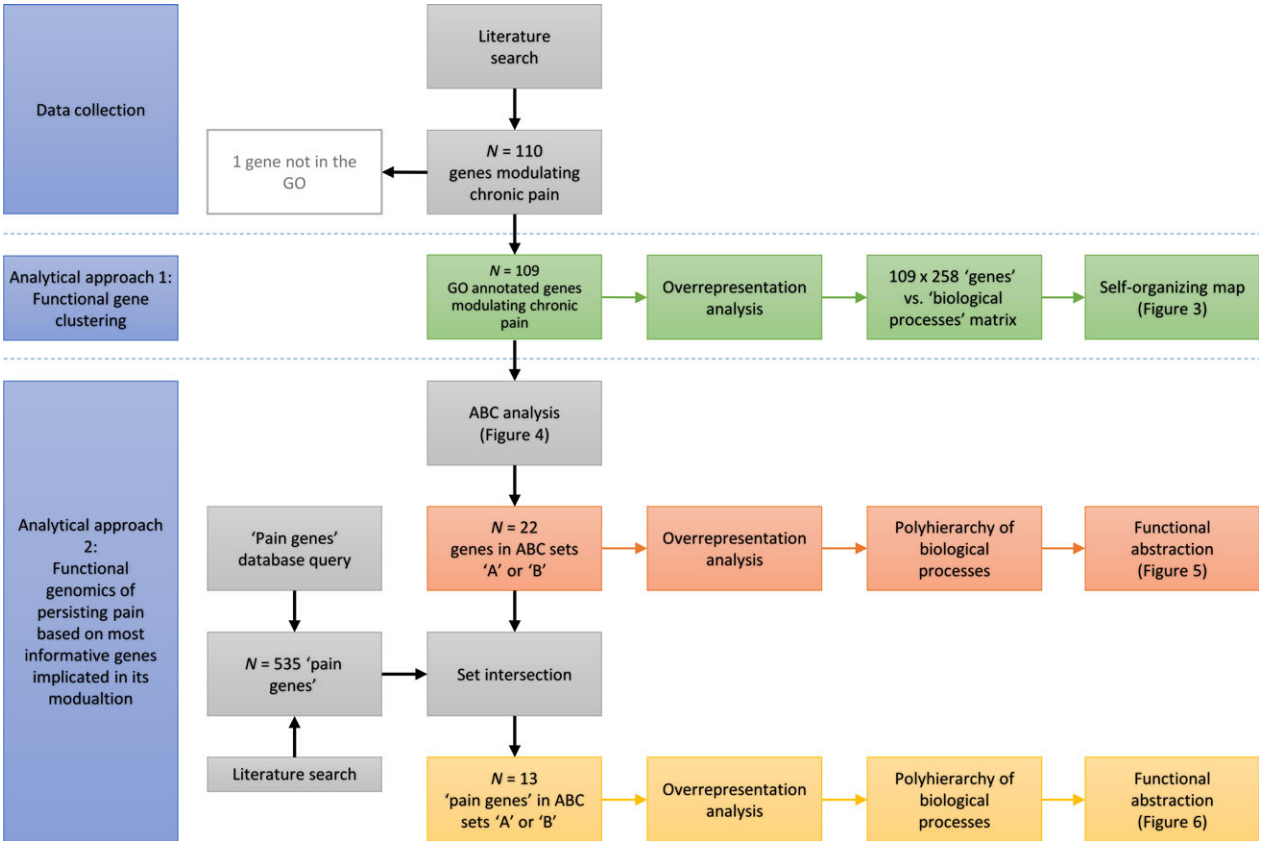


Figure 1 Flow chart of the data analysis. The figure provides an overview of the machine-learning and data science approach applied, specifying the data flow of the gene sets and the analyses applied to them. The figure follows the analytical flow (left column of blue rectangles, separated by horizontal dashed lines) that following data collection from the literature was implemented as two analytical approaches comprising (1) functional gene clustering and (2) a computational functional genomics analysis of the biological roles of the genes implicated in the modulation of persisting pain in various clinical settings. The coloured frames show the presented analysis, the grey frames the intermediate data flows.

(Ultsch and Lötsch, 2015) and subsequently, techniques of knowledge discovery were applied to identify the particular biological roles exerted by this set of genes as opposed to the biological functions exerted by a similarly sized random set of genes. This analysis was performed twice, once using the most informative genes resulting from above-mentioned analysis, and, to strengthen the evidence for pain reliance, again using the additional selection criterion that the genes should be listed among pain-relevant genes in the PainGenes database (Lacroix-Fralish et al., 2007). The analytical steps are described in detail in the following.

2.2.1 Functional clustering of genes

A first analytical approach to the functional genomics of persisting pain reflected in the set of genes that carry variants reported to modulate the phenotype aimed at finding functional clusters of genes.

2.2.2 Generation of a gene versus functional feature matrix

As a prerequisite for functional clustering of genes, a gene versus function matrix was created. Therefore, the biological functions in which the genes, or their products, are involved were queried from the GO knowledge base (http://www.geneontology.org/). The GO knowledge base is searchable for three major categories, consisting of biological process, cellular component and molecular function. As the most suitable representation of processes that are involved in the chronification of central sensitization to pain, the GO category biological process was selected as previously (Lötsch et al., 2013; Lötsch and Ultsch, 2016b; Ultsch et al., 2016). According to the GO knowledgebase, this category contains one or more ordered collections of molecular functions involving chemical or physical transformations such as cell growth and maintenance or signal transduction (Ashburner et al., 2000).

However, not all processes known to be influenced by the genes were sought, but only those that were annotated to the present set of genes more often than expected for any similarly sized random set of genes. Therefore, to capture biological processes that are particularly relevant to the present gene set, the set of genes was submitted to overrepresentation analysis (ORA; Backes et al., 2007). As intended, this compared the occurrence (as defined by its annotation term) of the particular set of genes covered by a GO term with the number of genes expected to be defined by this term. The significance of the association of a GO term with the expected list of genes was determined by means of a Fisher's exact test (Fisher, 1922). A p-value threshold, t_p, of 1×10^{-6} was applied and subsequently, the obtained results were additionally corrected for multiple testing according to Bonferroni (Hochberg, 1988). The result was the desired 'gene versus biological process' matrix (Fig. 2) that, rescaled as [0,1] indicating the absence or presence, respectively, of the involvement of a gene in a particular biological process, provided a filtered representation of the particularly important processes in which the analysed genes were involved while disregarding processes that would have been found by chance in any similarly sized gene set. This ORA-based filtering of gene functions was previously found to facilitate the functional analysis of gene sets including a context of pain and analgesia (Lötsch and Ultsch, 2016a; Lötsch et al., 2017).

2.2.3 Machine-learned cluster detection

Following the creation of the gene versus functional feature matrix, expressed as 'gene versus biological process' matrix, the feature space required for functional gene clustering was established as $D = \{(x_i) \mid x_i \in N^d, i = 1, \ldots, n\}$ comprising the d biological process to which the n genes in the analysed set were annotated. This feature space was searched for a cluster structure. Among several methods available for clustering, a method of unsupervised machine-learning shown recently to provide a viable unbiased clustering of high-dimensional biomedical data, outperforming classical clustering algorithms (Ultsch and Lötsch, 2017), was chosen. Specifically, the data space was projected from the high-dimensional feature space D onto a two-dimensional self-organizing map (SOM) of the Kohonen type (Kohonen, 1982). This map was composed of a toroid grid (Ultsch, 2003), i.e. a projection plane where opposite edges are connected. The grid had a size of 25×35 artificial neurons chosen according to the proposals of SOM size determination described previously (Ultsch and Lötsch, 2017). Each of the artificial neurons held a position vector, which carried the information about the biological processes associated with each gene, and a further parameter, which carried 'weights' of the same dimensions as the input dimensions. The weights were initially randomly drawn from the range of the data variables and subsequently adapted to the data during the learning phase of 25 epochs. The Euclidean distance was used for process (dis-)similarity; very general processes, such as 'biological process' that is the root term of the polyhierarchy carry the same value for all genes and therefore do not influence this distance. Following training of the neural network, an emergent SOM (ESOM; Lötsch and Ultsch, 2014; Ultsch and Sieman, 1990) was obtained that represented the genes as the localizations of their 'best matching units' (BMU), which are neurons that carried the vector most similar to a gene's data vector.

Following the projection of the data on the grid of neurons, an extension of the Kohonen map was applied to obtain clusters of genes. Specifically, the distance structure in the high-dimensional space was visualized using the so-called U-matrix (Ultsch and Sieman, 1990; Lötsch and Ultsch, 2014). The clusters became visible using a geographical map analogy where 'mountain ranges' represent large distances in the feature space that can be used to visually separate data clusters, whereas low 'valleys' represent sets of genes that are related to similar biological processes and therefore belong to the same cluster. The 'map' was further enhanced by calculating a so-called P-matrix (Ultsch, 2003), which displays the probability of a data point as $p(x) = \mid \{\text{data points } x_i \mid d(x_i, x) <= r\} \mid$ estimated as the number of data points in a sphere with radius r around x at each grid point on the ESOM's output grid. The calculations were performed using our R library 'Umatrix' (https://cran.r-project.org/package=Umatrix; Lötsch et al., 2018a).

2.2.4 Functional genomics analysis of most informative genes reported to modulate persisting pain

A second analytical approach at the functional genomics of persisting pain reflected in the set of genes that carry variants reported to modulate the phenotype aimed at identifying the biological roles of those genes that had been implicated most consistently in this context. Specifically, the present analysis aimed at the discovery of new knowledge about pain, rather than about the underlying disease, from

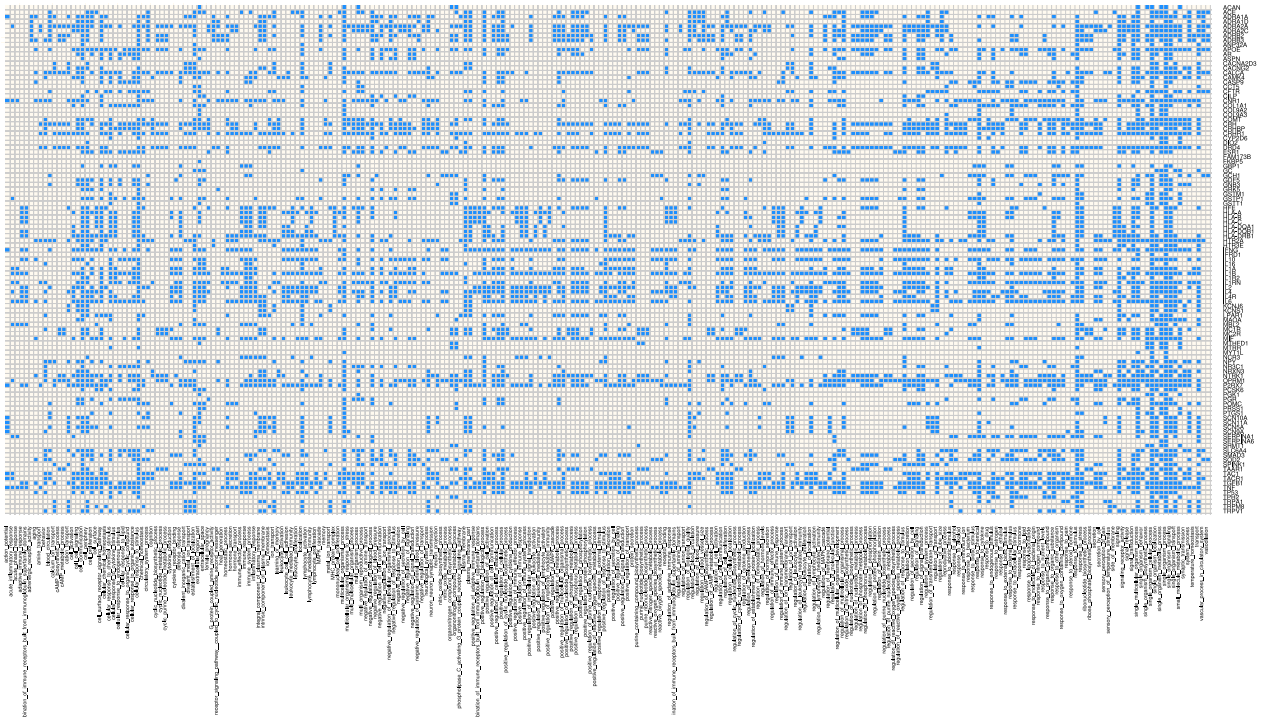


Figure 2 Matrix plot of the association of identified genes (rows; Table 1) with biological processes (columns), according to the annotations of the Gene Ontology (GO) knowledgebase (Ashburner et al., 2000), filtered for statistical significance in the present context by means of an overrepresentation analysis with a *p*-value threshold of 1×10^{-6} and correction for multiple testing according to Bonferroni. The matrix displays a yes/no scaling [1,0], colour-coded as blue or white, respectively. The figure has been created using the R software package (version 3.3.2 for Linux; http://CRAN.R-project.org/; R Development Core Team, 2008). Specifically, this plot was created using the 'heatmap.2' functions of the R package 'gplots' (Warnes G. R.; https://cran.r-project.org/package=gplots) with the build-in clustering of the plotting routine disabled (R switches 'Colv = FALSE, Rowv = FALSE'). For descriptions of the GO terms (abscissa), see the AmiGO search tool at http://amigo.geneontology.org/ (Carbon et al., 2009).

reports about a genetic modulation of a clinical symptom involving pain. This implies a distinction between the modulation of pain and the modulation of the disease-causing pain that has usually not been made in the included reports. For example, in rheumatic diseases a genetic variant could modulate the progress and severity of inflammatory processes and via this it could finally modulate pain, or a genetic variant could directly modulate the individual perception of pain and therefore, similar nociceptive stimuli produced by the inflammatory processes could cause different degrees of pain. However, in the interest of the present analysis, we assessed only the direct modulations of pain, which were expected to be provided most likely by those genes that had been implicated in modulation of persisting pain in more than one clinical setting.

2.2.5 Identification of genes most consistently reported to modulate persisting pain

To identify the most informative subset of genes reportedly modulating clinical persisting pain, a cut-off criterion had to be defined for the number of different clinical settings required for inclusion in further analyses. To avoid arbitrary criteria, the cut-off was obtained by applying an item categorization technique used to separate 'the important few' from the 'trivial many' (Juran, 1975). As a most suitable technique, because providing a mathematically based cut-off, computed ABC analysis was chosen for the present purpose, supported by previous demonstrations that it is suitable for item selection tasks in biomedical research (Ultsch and Lötsch, 2015; Lötsch and Ultsch, 2017b; Lötsch et al., 2018b).

ABC analysis requires a set of positive numbers, which was given by the column sums of the 'clinical settings versus genes' matrix (Fig. 4 top). The vector of the sums of clinical settings with positive reports of the modulatory involvement a gene's variant was submitted to computed ABC analysis (Ultsch and Lötsch, 2015), which aims to divide a set of positive data – here the set of genes scored according to their involvement in clinical settings

of persisting pain - into three disjointed subsets called 'A', 'B' and 'C'. Subset 'A' comprises 'the important few', subset 'C' comprises clearly non-profitable values, i.e. 'the trivial many' (Juran, 1975), whereas subset 'B' includes items that provide still a balance between effort and gain. Therefore, the limit separating subset 'C', i.e. the genes for which a modulation of persisting clinical pain provides the least relevant information, was chosen as the limit for the inclusion of genes in further analyses. These calculations were made using our R package 'ABCanalysis' (https://cran.r-project.org/package=ABCanalysis; Ultsch and Lötsch, 2015).

2.2.6 Functional genomics analysis of genes most consistently reported to modulate persisting pain

Following identification of the most informative genes as members of ABC sets 'A' or 'B', the functional genomics picture of persisting pain arising from these genes was analysed. This was obtained by applying ORA as described above, again using a p-value threshold of $t_p = 1 \times 10^{-6}$ with Bonferroni α-correction for multiple testing. The focus of this analysis was, however, the hierarchical representation of the complete knowledge on the biological roles of genes that carry polymorphisms observed to modulate persisting pain phenotypes. This was provided in a directed acyclic graph (DAG; Thulasiraman and Swamy, 1992). In this graph, the top-down, branching polyhierarchy of GO terms starts with the most broadly defined terms and progresses towards the branches, representing GO terms with the narrowest definition (details). These calculations were made using our R package 'dbtORA' (https://github.com/IME-TMP-FFM/dbtORA; Lippmann et al., 2018), which has been designed for knowledge discovery in the GO.

As the complete DAG usually contains many GO terms (e.g. 64 GO terms in the present ORA), the information was transformed into a more intelligible form using the method of 'functional abstraction' (Ultsch and Lötsch, 2014). This aims to reduce the numbers of GO terms using a heuristic search algorithm that identifies so-called functional areas (Ultsch and Lötsch, 2014), which are GO terms that qualify by their informational importance as headlines representing specific aspects (taxonomies) of the complete DAG with maximal coverage, precision, informational value and conciseness (Ultsch and Lötsch, 2014).

To narrow the focus even more to pain-relevant genes, the ORA was repeated using the set intersection of the most informative genes identified as described above by means of ABC analysis, with the genes listed among known pain-relevant genes in the PainGenes database (http://www.jbldesign.com/jmogil/enter.html; Lacroix-Fralish et al., 2007). These mainly include genes found, in at least three independent studies in transgenic mice, to contribute to the modulation of pain and identified using PubMed searches, with the addition of further genes (Lötsch et al., 2013) comprising those causally implicated in human hereditary diseases associated with extreme pain phenotypes (summarized in, e.g. Lötsch et al., 2017), and genes coding for the targets of approved analgesic drugs or of novel analgesics currently in clinical phases of development (Lötsch and Geisslinger, 2011). This provided a set of $n = 535$ 'pain genes' (Fig. 1 bottom).

3. Results

As a result of a literature search, a total of 110 unique genes were identified in eight different clinical settings of chronic central sensitization to pain, including back pain, inflammatory pain, musculoskeletal pain, neuropathic pain, visceral pain, widespread pain, idiopathic pain and miscellaneous pain, for which functional associations of genetic variants with differences in the phenotype of persisting pain had been reported (Table 1). Some of the included studies used a genomewide approach; however, many were candidate gene studies.

3.1 Functional clustering of genes carrying variants reportedly modulating persisting clinical pain

In a first analytical approach, functional subgroups were sought in the set of human genes, variants of which have been associated with modulation of the clinical phenotype of persisting pain. A filtered representation of the particularly important processes in which the analysed genes were involved while disregarding processes that would have been found by chance in any similarly sized gene set, was obtained by means of overrepresentation analysis (ORA) of the biological processes to which the genes were annotated in the GO knowledgebase. This identified $d = 258$ biological processes (GO terms), given the p-value threshold of 1×10^{-6} and the α-correction according to Bonferroni. One gene was neglected in this analysis, *COL6A4P1*, the collagen type VI alpha 4 pseudogene 1, because it was not referenced in the GO.

Subsequently, the 109 × 258 sized 'gene to biological process matrix' (Fig. 2) thus obtained was analysed for functional subgroups of genes. Using unsupervised machine-learning implemented as self-organizing artificial neuronal network of the Kohonen type enhanced by the U-matrix (i.e. an emergent self-organizing map, ESOM; Ultsch and Sieman, 1990; Lötsch and Ultsch, 2014), the high-dimensional data space was projected onto a two-dimensional toroid grid. On this grid, the U-matrix was visualized by applying a geographical landscape analogy (Fig. 3) providing a visual structure that could be employed for clustering of genes. The results indicated a large cluster comprising more than half of the genes (n = 58). However, this cluster was still composed of functionally very different genes, pointing towards a large heterogeneity of the genes chosen as candidate modulators of persisting pain in the different clinical studies, or was found with genomewide association studies (GWAS) without a focus on a particular gene. In addition, six smaller clusters were suggested, but their distinct separation was occasionally incomplete or, according to the U-matrix landscape analogy, they were not clearly presented as 'valleys' but consisted merely of 'mountain' zones separated by slightly higher ridges (Fig. 3). In ESOM/U-matrix based clustering, this indicates rather large intracluster distances.

As the present method has been shown to be well able to identify existent cluster structures while being unlikely to show false clusters (Ultsch, 2005; Ultsch and Lötsch, 2017), the main result of this analysis was that there is considerable heterogeneity among the genes reported to be involved in persisting pain without a clear functional focus on common general processes underlying this trait.

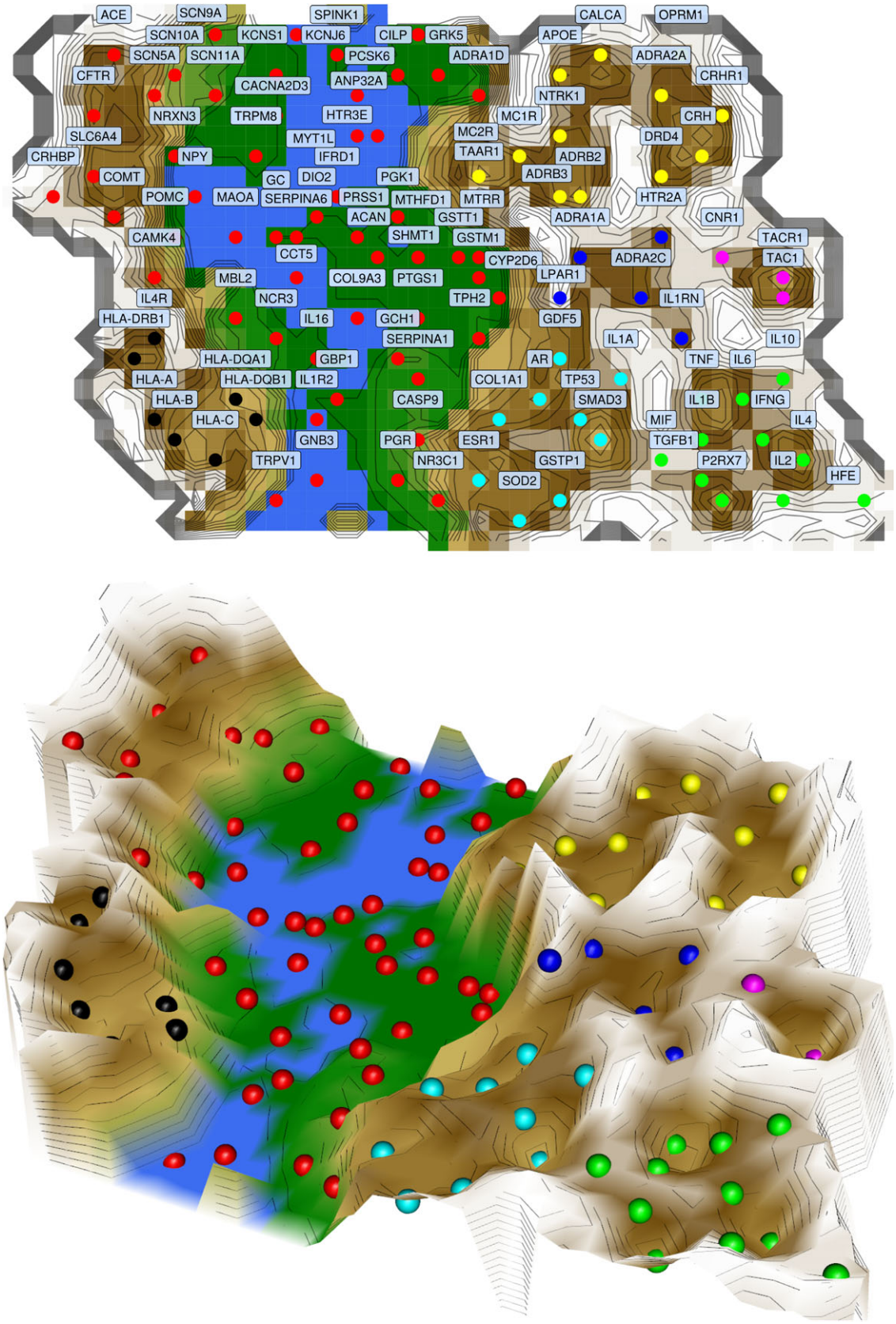
3.2 Functional genomics analysis of most informative genes reported to modulate persisting pain

In a second analytical approach, the hypothesis was pursued that the functional genomics of persisting pain will prevail across clinical settings irrespective of the disease that had originally triggered the process. However, the set intersection of the genes associated with any of the eight different clinical settings was empty. Nevertheless, the hypothesis that modulatory effects on chronic central sensitization to pain, rather than on the underlying disease-causing pain, are exerted by the same genes across several clinical settings could be pursued in several genes have been shown to be involved in more than one clinical setting of chronic central sensitization.

To identify how many clinical settings qualify as a cut-off, an ABC analysis was applied to the sum of settings in which each of the 110 genes was involved (column sums in the matrix plot in Fig. 4, top). This assigned 22 genes to ABC sets 'A' or 'B' (Table 2) that were included in further analysis as they could be regarded as best suited to perform an evidence-based functional genomics analysis of chronic central sensitization to pain, considering the difficulty of separation between a modulation of the pain-causing disease from a modulation of the perception and processing of chronic nociceptive input when a gene was involved only in a single clinical setting of persisting pain.

ORA identified 64 GO terms associated with this particular subset of genes more often than expected by chance, given the chosen p-value thresholds of 1×10^{-6} and correction for multiple testing according to Bonferroni. These terms provided a functional genomics perspective of the genes with variants

Figure 3 Result of a projection of the genes carrying variants reportedly modulating persistent pain onto a self-organizing map (SOM) of the Kohonen type (Kohonen, 1982). Following projection of the genes on the grid of neurons, based on their functional annotations in the Gene Ontology knowledgebase (Ashburner et al., 2000) (n = 109; one of the 110 genes in Table 1 was not annotated in the GO), the distance and density structures in the high-dimensional space were visualized using the so-called U*-matrix (Ultsch, 2003). Specifically, a trained SOM represents a topology preserving mapping of n high-dimensional data points $x_i \varepsilon D$, where D denotes the data space, onto a two-dimensional grid of neurons. A neuron n and the neurons in its neighbourhood $N(n)$ on the output grid of the SOM represent points in the data space. The sum of distances between n and $N(n)$ in the high-dimensional space, combined with the respective density probabilities, is shown on a U*-matrix as a height value (U-height) at neuron n. Large U-heights mean that there is a large gap in the data space. Low U-heights mean that the points are close to each other within the data space (Lötsch and Ultsch, 2014). The dots indicate the so-called best matching units (BMUs) of the SOM, which are those neurons whose weight vector is most similar to the input, i.e. the representation of the vector of the annotation of genes to GO terms. The BMUs are coloured according to the obtained clustering of the data space and labelled with the respective gene symbols. The cluster structure emerged from visualization of distances and density structures between neurons in the high-dimensional space by means of a U*-Matrix (Izenmann, 2009). Top: here, the genes represented by the BMUs are annotated. Bottom: 3D-display of the U-matrix in which the 'valleys', 'ridges' and 'basins' can be seen. Valleys indicate clusters of functionally similar genes based on the significant GO term annotations. The figure was created using the R software package (version 3.3.2 for Linux; http://CRAN.R-project.org/; R Development Core Team, 2008) using our R library 'Umatrix' (https://cran.r-project.org/package=Umatrix; Lötsch et al., 2018a).



shown to modulate the persisting pain sensitization phenotype in different clinical settings. Their graphical representation visualized their arrangement in the GO polyhierarchy (Fig. 5). Following application of functional abstraction (Ultsch and Lötsch, 2014), the main biological functions of the 22 gene carrying variants modulating persisting pain, were grouped around seven centres of biological activity or functional areas (Table 2; Fig. 5).

A first functional area to emerge was 'immune system process', represented in this particular gene set as an important common biological function (Fig. 5 middle left). The most significantly associated immune regulatory processes were 'regulation of immune effector processes' (GO:0002697, $p = 2.8 \times 10^{-8}$) and 'positive regulation of lymphocyte mediated immunity' (GO:0002708, $p = 2.9 \times 10^{-7}$). A second functional area centred on 'reactive oxygen species metabolic process' (GO:0072593, $p = 4.5 \times 10^{-11}$) and comprised mainly nitric oxide signalling related processes (Fig. 5 middle right) such as 'nitric oxide biosynthetic process' (GO:0006809, $p = 3.2 \times 10^{-12}$) and its regulation.

Further functional areas, however, mainly reflected processes known from previous research to contribute to pain (Lötsch et al., 2013; Ultsch et al., 2016). This included 'response to stimulus' with several subordinate terms, such as 'response to stress' (GO:0006950, $p = 5.2 \times 10^{-7}$), comprising the reaction of the body to several challenges such as 'response to chemical' (GO0042221, $p = 1.1 \times 10^{-10}$) and 'defense response' (GO:0006952, $p = 1.8 \times 10^{-7}$). Further subordinate areas included 'inflammatory response' (GO:0006954, $p = 2.8 \times 10^{-8}$) and 'response to other organism' (GO:0051707, $p = 2.5 \times 10^{-10}$) to which 'response to bacterium' (Fig. 5 middle left) was related. These response areas were also associated with the more general functional area 'regulation of multicellular organismal process' (GO: 51239, $p = 1.2 \times 10^{-7}$). In addition, the associated functional area 'transport', mainly comprised subordinate processes related to ion or transmitter transport such as 'regulation of ion transport' (GO:0043269, $p = 9 \times 10^{-9}$) or 'regulation of secretion by cell' (GO:1903530, $p = 4.1 \times 10^{-7}$). Finally, a functional area 'neurological system process' (GO:0050877, $p = 2.4 \times 10^{-7}$), as a more specific subordinate term to 'system process', reflected the expected involvement of nervous system related processes with pain.

The main results, i.e. the most specific functional areas pointing at immune processes and nitric oxide

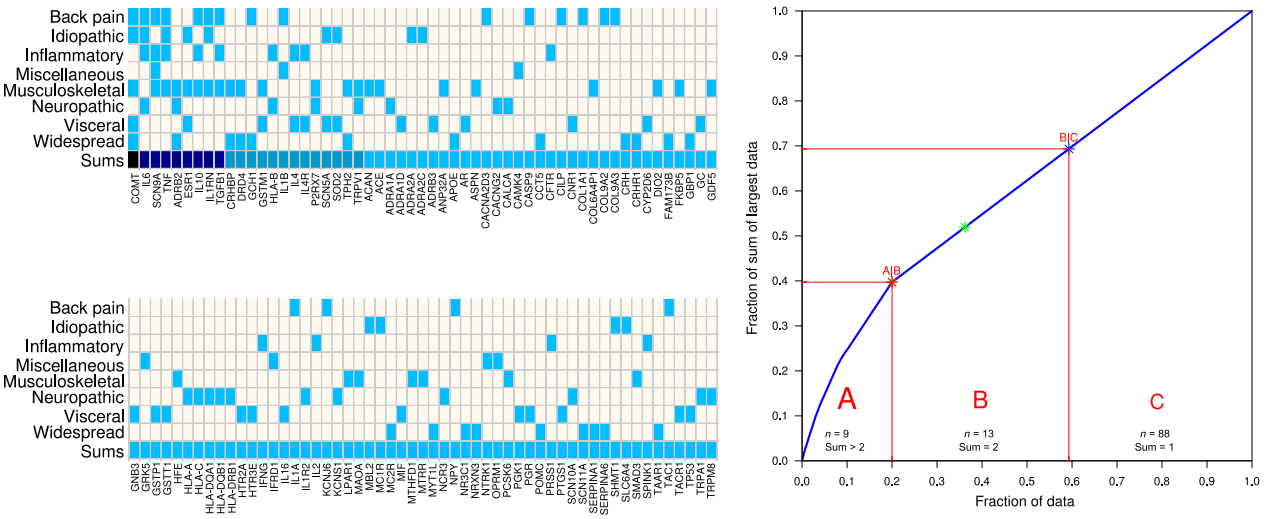


Figure 4 Number of clinical settings in which variants in the respective genes have been reported to be associated with modified phenotypes of chronic pain. Left: Matrix plot of the clinical settings (rows) versus the genes ($n = 110$, columns split in two halves, sorted for column sums in descending order). The column sums are displayed in the bottom row below the matrix. The numbers are displayed colour-coded with 0 = white, 1 = blue, >1 darker blue). Right: ABC plot of the cumulative distribution function of the sums of clinical settings in which the genes were reportedly involved (column sums in the top matrix; bottom line). The ABC set limits are indicated as red lines (for further details about an ABC analysis, see (Ultsch and Lötsch, 2015)). The figure was created using the R software package (version 3.3.2 for Linux; http://CRAN.R-project.org/; R Development Core Team, 2008). Specifically, the matrix plot was created using the 'heatmap.2' functions of the R package 'gplots' (Warnes G. R.; https://cran.r-project.org/package=gplots), with the build-in clustering of the plotting routine disabled (R switches 'Colv=FALSE, Rowv=FALSE'). The ABC curve was drawn using our R library 'ABCanalysis' (https://cran.r-project.org/package=ABCanalysis; Ultsch and Lötsch, 2015).

Table 2 Genes for which functional involvement of their variants has been reported in more than one clinical setting of persisting pain (number of these clinical settings given in the third column) were assigned to ABC sets 'A' or 'B' (Figure 4), comprising the most profitable information for an evidence-based functional genomics analysis of persisting pain.

			Functional areas						
Gene symbol	Gene name	Number of clinical settings of persisting pain	Response to stimulus	Immune system process	Reactive oxygen species metabolic process	Transport	Neurological system process	Regulation of multicellular organismal process	Multiorganism process
COMT	Catechol-O-methyltransferase	5	X	O	O	X	X	X	X
IL-6	Interleukin-6	4	X	X	X	X	X	X	X
SCN9A	Sodium channel, voltage-gated, type IX alpha subunit	4	X	O	O	X	X	O	O
TNF	Tumour necrosis factor	4	X	X	X	X	X	X	X
ADRB2	Adrenoceptor beta 2, surface	3	X	O	O	X	O	X	O
ESR1	Oestrogen receptor 1	3	X	O	X	O	O	X	X
IL-10	Interleukin-10	3	X	X	X	X	X	X	X
IL-1RN	Interleukin-1 receptor antagonist	3	X	X	O	X	X	X	X
TGFB1	Transforming growth factor, beta 1	3	X	X	X	X	O	X	X
CRHBP	Corticotropin-releasing hormone binding protein	2	X	O	O	X	X	X	X
DRD4	Dopamine receptor D4	2	X	O	O	X	X	X	X
GCH1	GTP cyclohydrolase 1	2	X	X	X	O	X	O	X
GSTM1	Glutathione S-transferase mu 1	2	X	O	O	O	O	O	O
HLA-B	Major histocompatibility complex, class I, B	2	X	X	O	O	O	X	X
IL-1B	Interleukin-1, beta	2	X	X	X	X	O	X	X
IL-4	Interleukin-4	2	X	X	X	X	O	X	X
IL-4R	Interleukin-4 receptor	2	X	X	O	X	O	X	X
P2RX7	Purinergic receptor P2X, ligand gated ion channel, 7	2	X	X	X	X	X	X	X
SCN5A	Sodium channel, voltage-gated, type V alpha subunit	2	X	O	O	X	X	X	O
SOD2	Superoxide dismutase 2, mitochondrial	2	X	X	X	O	X	O	X
TPH2	Tryptophan hydroxylase 2	2	X	O	O	O	O	O	O
TRPV1	Transient receptor potential cation channel, subfamily V, member 1	2	X	O	O	X	X	O	O
	Sum		22	12	10	16	13	16	16

The right part of the table displays the functional areas (Figure 5) or groups of biological functions in which the gene set is involved, together with their association with each gene (X = yes, O = No). The precise definition of the GO terms can be obtained using AmiGO search tool for GO at http://amigo.geneontology.org/ (Carbon et al., 2009).

signalling as key biological processes involved in persisting pain across more than a single clinical setting, prevailed when narrowing the gene set further on those that are also listed in the PainGenes database (Fig. 6). As the set intersection between the 22 genes identified above and the genes of the PainGenes database included only $n = 13$ genes, the ORA applying the same statistical thresholds resulted in fewer additional significant GO terms (Fig. 6).

4. Discussion

The present analysis used empirical evidence for functional human genetic variants to approach the genetic architecture of persisting pain. Although the evidence was collected from separate studies, its combination permitted a limited genomewide association analysis of the trait. Methods for data mining and machine-learned knowledge discovery were applied to publicly available databases in order to relate knowledge, acquired in the context of clinical pain research, on genes that modulate the phenotype of persisting pain with data on the biological functions of these human genes acquired in any context, without restriction to pain research (Lötsch et al., 2013).

The initial analysis of the whole set of genes showing positive results from clinical pain association studies indicated that, apart from a minority of genes that could be topically grouped such as interleukin or histocompatibility complex-related genes, most genes displayed very heterogeneous functions and the analysis did not illuminate the pathophysiology of persisting pain beyond the functions of the single genes. This was probably due to the fact that data were drawn mainly from candidate gene or GWAS approaches. This selection probably introduced a research bias by (1) addressing genetic modulators in the context of the underlying disease and (2) including a selection of genetic markers that mimic other successful reports of comparable studies.

The situation became clearer when the analytical focus was narrowed to genetic modulations consistently observed across several clinical conditions with potential underlying painful diseases. This reduced the analytical bias generated by genetic modulations responsible for a specific pain-causing disease and is in keeping with the contemporary approach to persisting pain as a distinct condition of central sensitization to pain and not merely a symptom of another underlying chronic disease. Consequently, it would be expected that the trait is modulated by specific genes which should be reflected by observations on its modulation in clinical research. The mathematically precise calculation provided by the ABC analysis (Ultsch and Lötsch, 2015), developed in order to select the most promising or profitable items from a larger set of items, resulted in identification of a set of 22 genes which could be then be assessed in a computational functional genomics analysis of persisting pain.

A major finding of this analysis of available evidence on genetic modulation of persisting pain was the particular importance of two groups of biological processes indicating involvement (1) of the immune system and of (2) nitric oxide signalling in persisting pain. Involvement of both processes is biologically highly plausible; however, their emergence as major process groups from a functional genomics analysis of data from clinical genetic research on persisting pain was not anticipated. Specific roles for the present subset of 22 genes, with repeated evidence for involvement in persisting pain, were exhibited by the 12 genes annotated as 'immune system process' (Table 1). This subset included interleukin (*IL-1B*, *IL-4*, *IL-6*, *IL-10*) (Dinarello, 1994; Choi and Reiser, 1998; Mocellin et al., 2004; Nemeth et al., 2004), interleukin receptor (*IL-1RN*, *IL-4R*) (Bittar and Bittar, 1996) and histocompatibility complex-related (*HLA-B*) genes (Dupont and Ceppellini, 1989), which have been shown to be involved in immunological mechanisms of pain (Sato et al., 2002; de Rooij et al., 2009). This is also supported by published evidence for the further genes in this list, such as, *TNF* (Vassalli, 1992; Franchimont et al., 1999), *TGFB1* (Li et al., 2006), *GCH1* (Schott et al., 1993), *P2RX7* (Schwartz et al., 2009) and *SOD2* (Wells et al., 2003). The second major process group emerging from the functional genomics analysis of the key evidence for genetic modulation of clinical persisting pain was nitric oxide signalling, in particular metabolic processes, summarized in this context under the GO term 'reactive oxygen species metabolic process' which includes the genes *IL-6* (Deakin et al., 1995), *TNF* (Deakin et al., 1995; Katusic et al., 1998), *ESR1* (Clapauch et al., 2014), *IL-10* (Cattaruzza et al., 2003), *TGFB1* (Saura et al., 2005), *GCH1* (Katusic et al., 1998; Zhang et al., 2007), *IL-1B* (Katusic et al., 1998), *IL-4* (Coccia et al., 2000), *P2RX7* (Gendron et al., 2003), *SOD2* (Fridovich, 1978). It is widely accepted that nitric oxide (NO) is critically involved in persisting pain (Chung, 2004). It has been shown that NO is produced in the spinal dorsal horn neurons in response to extensive nociceptive inputs and then it diffuses out and increases neurotransmitter release from primary afferent terminals, thereby contributing to central sensitization and persisting pain (Lin et al., 1999). Recent findings seem to indicate that not only NO is a mediator of persisting pain that accompanies inflammation, other reactive oxygen species like superoxide (SO) might also participate in persisting pain (Schwartz et al., 2008). Kim and colleagues found that NO and SO contribute to persisting pain via two separate and independent pathways and a recent study has shown that capsaicin-induced hyperalgesia is a consequence of superoxide build-up in spinal dorsal horn neurons. Superoxide dismutase (SOD-2) encoded by gene *SOD2* is a major

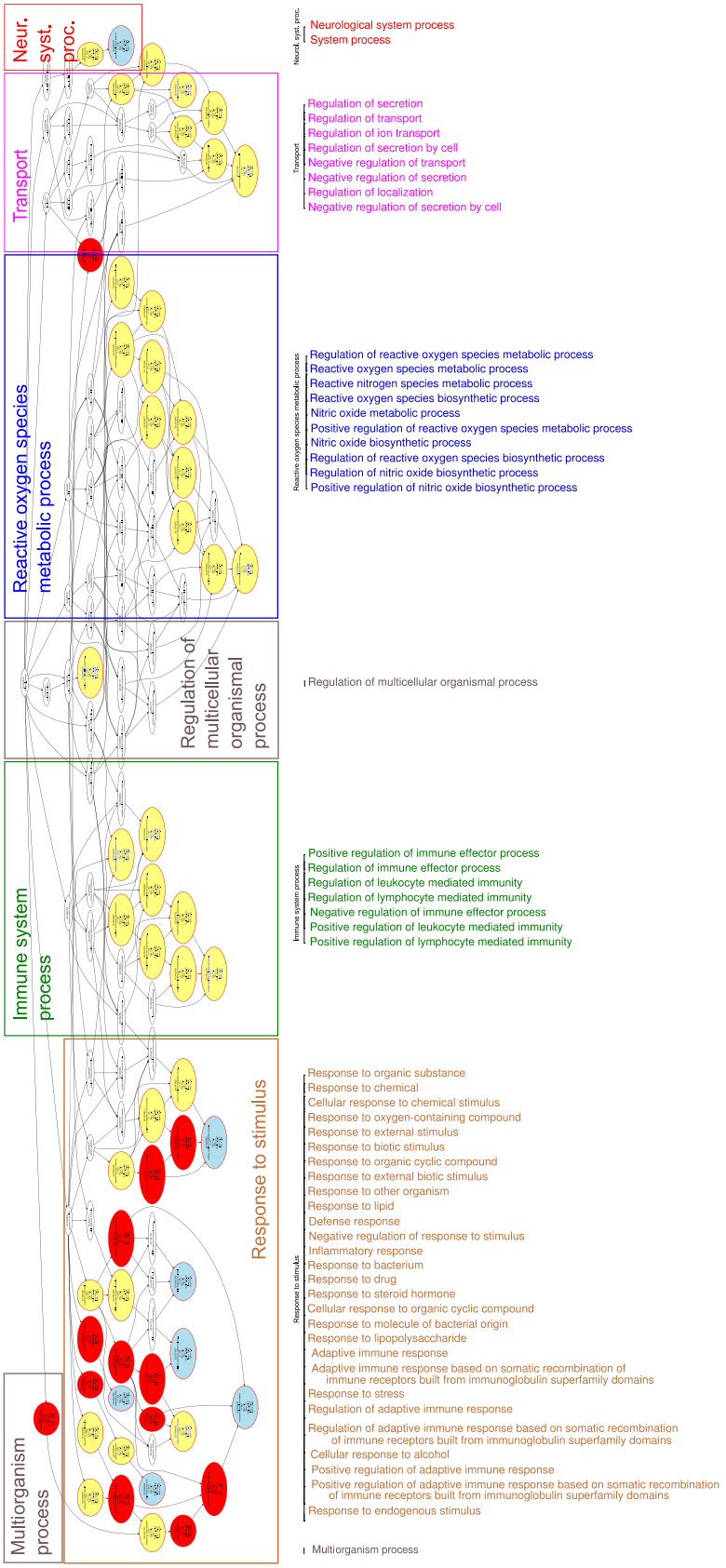


Figure 5 Results of an overrepresentation analysis (ORA; *p*-value threshold, $t_p = 1 \times 10^{-6}$ and Bonferroni α correction) of 22 genes (Table 2) carrying variants reported to modulate the persisting pain phenotypes in different clinical settings (ABC sets 'A' or 'B') versus all human genes. A top-down representation of the annotations (GO terms) is shown representing a systems-biology perspective of the biological processes modulated by this set of genes. Each ellipse represents a GO term. The graphical representation follows the standard of the GO knowledge base, where GO terms are related to each other by 'is-a', 'part-of', and 'regulates' relationships forming a branching polyhierarchy organized in a directed acyclic graph (DAG; Thulasiraman and Swamy, 1992). Top: Significant GO terms are shown as coloured ellipses with the number of member genes, the number of expected genes by chance and the significance of the deviation in the observed from the expected number of genes indicated. The biological processes in which the present $n = 22$ genes are involved can be summarized by seven primary 'functional areas' or headlines presenting particular aspects (taxonomies) of the complete polyhierarchy at maximum coverage, precision, informational value and conciseness (Ultsch and Lötsch, 2014). The ellipses are colour-coded using yellow for a 'headline'; i.e. a GO term that by its location in the polyhierarchy may serve as headlines for a branch of the hierarchy, red for significantly overrepresented terms and white for non-significant terms that need to be displayed to preserve the polyhierarchical structure of the DAG. Blue vertices or blue labels are the most specific terms (leaves of the DAG) at the end of a taxonomy (branch) in the polyhierarchy. Bottom: The GO terms (biological processes) taken from the functional areas, shown above in the DAG, are shown with larger fonts for better readability. The figure was created using the R software package (version 3.3.2 for Linux; http://CRAN.R-project.org/; R Development Core Team, 2008) and our R package 'dbtORA' (https://github.com/IME-TMP-FFM/dbtORA; Lippmann et al., 2018).

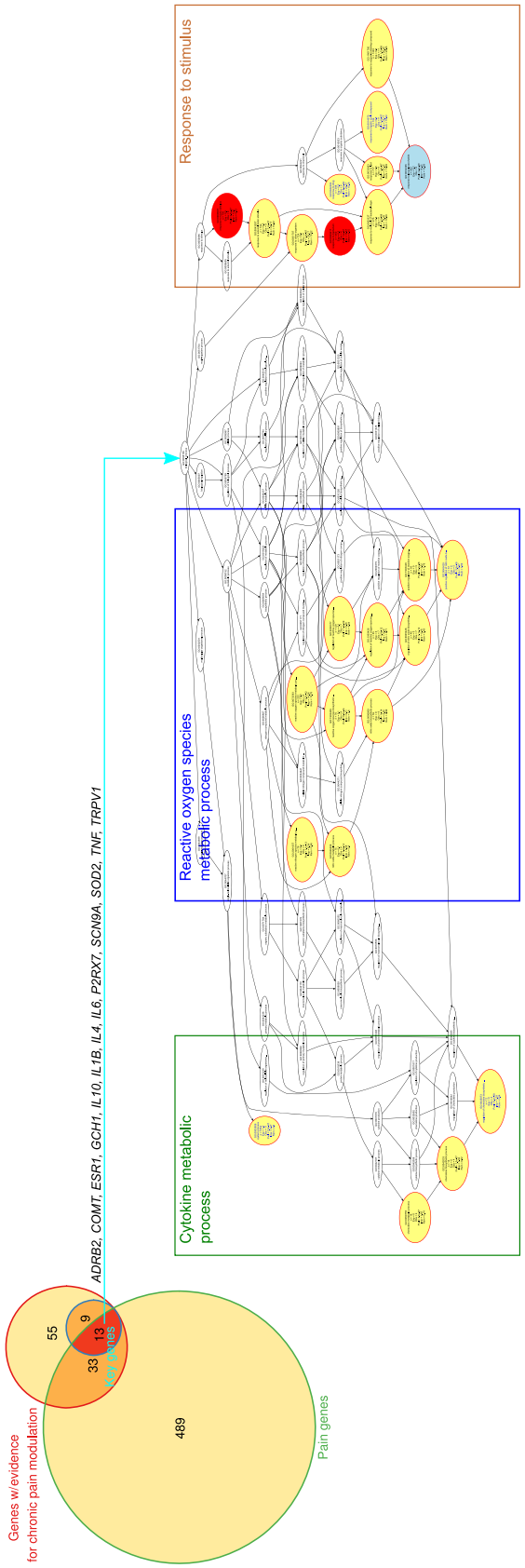


Figure 6 Overrepresentation analysis (ORA) of 13 genes (Table 2), versus all human genes, that (1) carry variants reported to modulate the persisting pain phenotypes, (2) belonging to the subset of 22 of these genes that are supported by evidence that this modulation applies to several clinical settings (ABC sets 'A' or 'B') and (3) were also part of a set of 535 genes, based on current pain research data, that can be considered, with their products, to be involved in the pathophysiology of pain (Ultsch et al., 2016). Left part: The Venn diagram (Venn, 1880) shows the set intersections on which this selection of 13 genes is based. Right part: Top-down representation of the annotations (GO terms) representing the systems-biology perspective of the biological processes modulated by this set of genes organized in a branching polyhierarchy forming a directed acyclic graph (DAG; Thulasiraman and Swamy, 1992). The figure represents the results of an overrepresentation analysis with parameters for p-value threshold, $t_p = 1 \times 10^{-6}$ and Bonferroni α correction. The biological processes in which the $n = 13$ genes are involved can be summarized by three primary 'functional areas' or headlines presenting particular aspects (taxonomies) of the complete polyhierarchy at maximum coverage, precision, informational value and conciseness (Ultsch and Lötsch, 2014). The ellipses are colour-coded using yellow for a 'headline'; i.e. a GO term that by its location in the polyhierarchy may serve as headlines for a branch of the hierarchy, red for significantly overrepresented terms located in the polyhierarchy below a headline and white for nonsignificant terms that need to be displayed to preserve the polyhierarchical structure of the DAG. Blue vertices or blue labels are the most specific terms (leaves of the DAG) at the end of a taxonomy (branch) in the polyhierarchy. Bottom: The GO terms (biological processes) taken from the functional areas, shown above in the DAG, are shown with larger fonts for better readability. The figure was created using the R software package (version 3.3.2 for Linux; http://CRAN.R-project.org/; R Development Core Team, 2008) and our R package 'dbtORA' (https://github.com/IME-TMP-FFM/dbtORA; Lippmann et al., 2018). The Venn diagram was drawn using the R library 'VennerableDevel' (Swinton J., https://r-forge.r-project.org/R/?group_id=474).

determinant suggesting a therapeutic potential for the manipulation of spinal SOD-2 activity in pain conditions (Schwartz et al., 2009).

The role of cytokines was further highlighted by further restricting the gene subset most consistently related to persisting pain by identifying the appearance of the relevant gene in an independently created list of 535 so-called pain genes (Ultsch et al., 2016). These were genes relevant to pain listed by several sources: mainly the Pain Genes Database (http://www.jbldesign.com/jmogil/enter.html (Lacroix-Fralish et al., 2007). The overlap of the set of the 22 genes, with repeated records of variants that modulate the clinical phenotype of persisting pain, with the 535 'pain genes' comprised 13 genes (*ADRB2*, *COMT*, *ESR1*, *GCH1*, *IL-10*, *IL-1B*, *IL-4*, *IL-6*, *P2RX7*, *SCN9A*, *SOD2*, *TNF*, *TRPV1*). After performing a similar ORA (Fig. 6) as that described above, these 13 genes were found to be mainly involved in cytokine production, covered by the significant GO terms 'chemokine metabolic process' (GO:0050755, $p = 3.3 \times 10^{-7}$), 'nitrogen compound metabolic process' and again 'response to stimulus', the most significant term being 'nitric oxide biosynthetic process' (GO:0006809, $p = 1 \times 10^{-14}$).

The involvement of the immune system in persisting pain is plausible from a biological perspective. One of the sites of interaction of the immune system with persisting pain has been identified as neuroimmune crosstalk at the glial–opioid interface (Tian et al., 2012). Previous research has shown that glial and immune cells, including astrocytes, microglia/macrophages, as well as T lymphocytes, are key cells activated during persisting pain, which contribute to pain persistence (Calvo et al., 2012; von Hehn et al., 2012). A role for local production of cytokines in the central nervous system during inflammatory conditions associated with persisting pain, such as rheumatoid arthritis (Lampa et al., 2012) or fibromyalgia (Kadetoff et al., 2012), as well as evidence for central nervous system sensitization by cytokines (Aden et al., 2010) also suggests such an immune system interaction with persisting pain states. Indeed, it has recently been suggested that the predominance of pain sensitization during chronic diseases in women is closely linked to the effects of female sex hormones on the neuroimmune system (Rosen et al., 2017).

The present results clearly support the modulation of neuroimmune system processes as a promising strategy in the development of novel analgesic drugs against persisting pain. This may be possible along several lines. For example, involvement of glial cells in opiate actions has been shown recently (Chen et al., 2010; Boue et al., 2012). Consequently, the elucidation of pain- and opioid-induced mechanisms at the level of glial and immune cells could lead to improvement of pain management. In an animal model, ibudilast, a nonselective phosphodiesterase-inhibiting, anti-inflammatory drug that also blocks glial activation probably via antagonism at the Toll-like receptor 4 (Jia et al., 2012), restored morphine-induced antinociception following tolerance development (Lilius et al., 2009). Similarly, minocycline, a tetracycline that inhibits microglial activation and proliferation, also seems to attenuate morphine tolerance in mouse models of neuropathic pain (Chen et al., 2010). Hence, increasing evidence points towards the immune system as a potential source of future targets for analgesic drugs directed against persisting pain.

By gathering the relevant reports, the present analysis centres on the current evidence about a genetic modulation of persisting pain on a gene level, without going into the details of single nucleotide polymorphism level as usually applied in review papers. However, the approach was centred on machine-learned knowledge discovery from the gathered evidence and was based on published evidence gathered from studies in which the authors had used a candidate gene approach or had performed a GWAS without a gene-specific hypothesis. Therefore, the present analysis implies a research bias given by the original hypotheses or on the inclusion of frequently addressed genetic variants in the analysed studies. The question addressed with the present analysis was about the greater functional perspective emerging from successful clinical studies of the genetic modulation of persisting pain. Importantly, while the analysis is biased with respect to the gene selection, made by the authors of the included studies, its results are not biased for a particular functional genomics perspective as this had not been an gene inclusion criterion in the analysed studies. Nevertheless, the present selection of repeatedly reported associations implies an advantage of frequently included genes that have been attracted research interest through the last several years, such as *OPRM1*, *GCH1* or *COMT*.

With the caution advised by the implicit research bias regarding the gene selection, the results of the present analysis were (1) unexpected considering that hypotheses about the involvement of immune system processes or of nitric oxide signalling were not preformulated for the present analysis and (2) biologically plausible and completely compatible with current research activities on persisting pain in the light of increasingly acknowledgement of an involvement of immune processes that has attracted

concerted research activities (Kringel and Lötsch, 2015), including that on the role of the glial–opioid interface in persisting pain (http://gloria.helsinki.fi). Hence, although the evidence was generated in separate studies, the combination of positive findings of a genetic modulation of persisting pain allows a limited yet valid genomewide analysis of the trait, providing a more comprehensive picture of functional background of persisting pain from genomics perspective than associations of single genotypes.

5. Conclusions

While many studies have focused on particular genes, the present analysis pursued the question whether their combined results may provide more complex insights into the pathophysiology of persisting pain in humans. In keeping with the contemporary trend towards 'big data' analysis in biomedical research, the current empirical data on modulation of persisting pain via human genetic polymorphisms have been subjected here to a computational functional genomics analysis. This evidence was then analysed for emergent, principal pathophysiological processes that characterize persisting sensitization to pain. Analysis of 110 unique genes, with variants that have been reported to modulate the clinical phenotype of persisting pain, led to the selection of functionally heterogeneous genes. By focusing on genes that have been repeatedly associated with modulation of persisting pain phenotypes in several clinical settings, a clearer picture emerged of the main processes identified by current human genetics research on persisting pain. A mathematically supported, precise selection of a subset of genes was possible using a computational functional genomics approach. On this basis, including the research bias of current clinical genetic association studies, the evidence gathered so far points to the view of persisting pain as a trait resulting from alterations in the immune system and/or in nitric oxide signalling, a concept that is biologically highly plausible and agrees with other lines of pain research. While analysing existing evidence and therefore limited to previously shown functional pathways, the present computational functional genomics-based approach provides a computational systems-biology perspective on chronic sensitization to pain by summarizing the empirical evidence gathered in many separate studies. Moreover, human genetic research on persisting pain emphasizes the immune system as a potential source of important future targets for analgesic drugs directed against persisting pain and demonstrates that contemporary machine-learned methods offer innovative approaches to knowledge discovery from previous evidence.

Author contributions

JL, DK and EK conceived and designed the analysis. JL, CL and DK analysed the data. JL, DK, EK and MJP wrote the article. JL, AU, DK and EK involved in discussion of methods and results.

References

Aden, U., Favrais, G., Plaisant, F., Winerdal, M., Felderhoff-Mueser, U., Lampa, J., Lelievre, V., Gressens, P. (2010). Systemic inflammation sensitizes the neonatal brain to excitotoxicity through a pro-/anti-inflammatory imbalance: Key role of TNFalpha pathway and protection by etanercept. *Brain Behav Immun* 24, 747–758.

Ala-Kokko, L. (2002). Genetic risk factors for lumbar disc disease. *Ann Med* 34, 42–47.

Alizadeh, B.Z., Njajou, O.T., Hazes, J.M., Hofman, A., Slagboom, P.E., Pols, H.A., van Duijn, C.M. (2007). The H63D variant in the HFE gene predisposes to arthralgia, chondrocalcinosis and osteoarthritis. *Ann Rheum Dis* 66, 1436–1442.

Aneiros-Guerrero, A., Lendinez, A.M., Palomares, A.R., Perez-Nevot, B., Aguado, L., Mayor-Olea, A., Ruiz-Galdon, M., Reyes-Engel, A. (2011). Genetic polymorphisms in folate pathway enzymes, DRD4 and GSTM1 are related to temporomandibular disorder. *BMC Med Genet* 12, 75.

Arisan, E.D., Arisan, S., Kiremit, M.C., Tigli, H., Caskurlu, T., Palavan-Unsal, N., Ergenekon, E. (2006). Manganese superoxide dismutase polymorphism in chronic pelvic pain syndrome patients. *Prostate Cancer Prostatic Dis* 9, 426–431.

Arisawa, T., Tahara, T., Shibata, T., Nagasaka, M., Nakamura, M. et al. (2007). Genetic polymorphisms of molecules associated with inflammation and immune response in Japanese subjects with functional dyspepsia. *Int J Mol Med* 20, 717–723.

Arisawa, T., Tahara, T., Shibata, T., Nagasaka, M., Nakamura, M. et al. (2008). Genetic polymorphisms of cyclooxygenase-1 (COX-1) are associated with functional dyspepsia in Japanese women. *J Womens Health (Larchmt)* 17, 1039–1043.

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H. et al. (2000). Gene ontology: Tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* 25, 25–29.

Attur, M., Wang, H.Y., Kraus, V.B., Bukowski, J.F., Aziz, N. et al. (2010). Radiographic severity of knee osteoarthritis is conditional on interleukin 1 receptor antagonist gene variations. *Ann Rheum Dis* 69, 856–861.

Babula, O., Danielsson, I., Sjoberg, I., Ledger, W.J., Witkin, S.S. (2004). Altered distribution of mannose-binding lectin alleles at exon I codon 54 in women with vulvar vestibulitis syndrome. *Am J Obstet Gynecol* 191, 762–766.

Backes, C., Keller, A., Kuentzer, J., Kneissl, B., Comtesse, N. et al. (2007). GeneTrail–advanced gene set enrichment analysis. *Nucleic Acids Res* 35, W186–W192.

Binder, A., May, D., Baron, R., Maier, C., Tölle, T.R. et al. (2011). Transient receptor potential channel polymorphisms are associated with the somatosensory function in neuropathic pain patients. *PLoS ONE* 6, e17387.

Bittar, E.E., Bittar, N. (1996). *Immunobiology* (Greenwich, Conn.: JAI Press).

Blanco, I., Arbesú, D., Kassam, D.A., de Serres, F.J., Fernández-Bustillo, E., Rodríguez, C. (2006). Alpha1-antitrypsin polymorphism in fibromyalgia syndrome patients from the Asturias province in Northern Spain: A significantly higher prevalence of the PI*Z deficiency allele in patients than in the general population. *J Musculoskelet Pain* 14, 5–12.

Bortsov, A.V., Smith, J.E., Diatchenko, L., Soward, A.C., Ulirsch, J.C. et al. (2013). Polymorphisms in the glucocorticoid receptor co-chaperone FKBP5 predict persistent musculoskeletal pain after traumatic stress exposure. *Pain* 154, 1419–1426.

Boue, J., Blanpied, C., Djata-Cabral, M., Pelletier, L., Vergnolle, N., Dietrich, G. (2012). Immune conditions associated with CD4 + T effector-induced opioid release and analgesia. *Pain* 153, 485–493.

Breivik, H., Collett, B., Ventafridda, V., Cohen, R., Gallacher, D. (2006). Survey of chronic pain in Europe: Prevalence, impact on daily life, and treatment. *Eur J Pain* 10, 287–333.

Bruehl, S., Denton, J.S., Lonergan, D., Koran, M.E., Chont, M. et al. (2013). Associations between KCNJ6 (GIRK2) gene polymorphisms and pain-related phenotypes. *Pain* 154, 2853–2859.

Buskila, D., Cohen, H., Neumann, L., Ebstein, R.P. (2004). An association between fibromyalgia and the dopamine D4 receptor exon III repeat polymorphism and relationship to novelty seeking personality traits. *Mol Psychiatry* 9, 730–731.

Calvo, M., Dawes, J.M., Bennett, D.L. (2012). The role of the immune system in the generation of neuropathic pain. *Lancet Neurol* 11, 629–642.

Camon, E., Magrane, M., Barrell, D., Binns, D., Fleischmann, W. et al. (2003). The Gene Ontology Annotation (GOA) project: Implementation of GO in SWISS-PROT, TrEMBL, and InterPro. *Genome Res* 13, 662–672.

Camon, E., Magrane, M., Barrell, D., Lee, V., Dimmer, E. et al. (2004). The Gene Ontology Annotation (GOA) Database: Sharing knowledge in Uniprot with Gene Ontology. *Nucleic Acids Res* 32, D262–D266.

Carbon, S., Ireland, A., Mungall, C.J., Shu, S., Marshall, B., Lewis, S., Ami, G.O.H., Web Presence Working Group (2009). AmiGO: Online access to ontology and annotation data. *Bioinformatics* 25, 288–289.

Cattaruzza, M., Slodowski, W., Stojakovic, M., Krzesz, R., Hecker, M. (2003). Interleukin-10 induction of nitric-oxide synthase expression attenuates CD40-mediated interleukin-12 synthesis in human endothelial cells. *J Biol Chem* 278, 37874–37880.

Chen, S., Hui, H., Zhang, D., Xue, Y. (2010). The combination of morphine and minocycline may be a good treatment for intractable post-herpetic neuralgia. *Med Hypotheses* 75, 663–665.

Cheng, K.I., Lin, S.R., Chang, L.L., Wang, J.Y., Lai, C.S. (2010). Association of the functional A118G polymorphism of OPRM1 in diabetic patients with foot ulcer pain. *J Diabetes Complications* 24, 102–108.

Choi, P., Reiser, H. (1998). IL-4: Role in disease and regulation of production. *Clin Exp Immunol* 113, 317–319.

Chung, J.M. (2004). The role of reactive oxygen species (ROS) in persistent pain. *Mol Interv* 4, 248–250.

Clapauch, R., Mourao, A.F., Mecenas, A.S., Maranhao, P.A., Rossini, A., Bouskela, E. (2014). Endothelial function and insulin resistance in early postmenopausal women with cardiovascular risk factors: Importance of ESR1 and NOS3 polymorphisms. *PLoS ONE* 9, e103444.

Coccia, E.M., Stellacci, E., Marziali, G., Weiss, G., Battistini, A. (2000). IFN-gamma and IL-4 differently regulate inducible NO synthase gene expression through IRF-1 modulation. *Int Immunol* 12, 977–985.

Cohen, H., Neumann, L., Glazer, Y., Ebstein, R.P., Buskila, D. (2009). The relationship between a common catechol-O-methyltransferase (COMT) polymorphism val(158) met and fibromyalgia. *Clin Exp Rheumatol* 27, S51–S56.

Costigan, M., Belfer, I., Griffin, R.S., Dai, F., Barrett, L.B. et al. (2010). Multiple chronic pain states are associated with a common amino acid-changing allele in KCNS1. *Brain* 133, 2519–2527.

Dai, F., Belfer, I., Schwartz, C.E., Banco, R., Martha, J.F. et al. (2010). Association of catechol-O-methyltransferase genetic variants with outcome in patients undergoing surgical treatment for lumbar degenerative disc disease. *Spine J* 10, 949–957.

De Carvalho, C.V., Nogueira-De-Souza, N.C., Costa, A.M., Baracat, E.C., Girao, M.J., D'Amora, P., Schor, E., da Silva, I.D. (2007). Genetic polymorphisms of cytochrome P450cl7alpha (CYP17) and progesterone receptor genes (PROGINS) in the assessment of endometriosis risk. *Gynecol Endocrinol* 23, 29–33.

Deakin, A.M., Payne, A.N., Whittle, B.J., Moncada, S. (1995). The modulation of IL-6 and TNF-alpha release by nitric oxide following stimulation of J774 cells with LPS and IFN-gamma. *Cytokine* 7, 408–416.

Derry, S., Gill, D., Phillips, T., Moore, R.A. (2012). Milnacipran for neuropathic pain and fibromyalgia in adults. *Cochrane Database Syst Rev* 3, CD008244.

Dhar, V. (2013). Data science and prediction. *Commun ACM* 56, 64–73.

Diatchenko, L., Anderson, A.D., Slade, G.D., Fillingim, R.B., Shabalina, S.A. et al. (2006). Three major haplotypes of the beta2 adrenergic receptor define psychological profile, blood pressure, and the risk for development of a common musculoskeletal pain disorder. *Am J Med Genet B Neuropsychiatr Genet* 141B, 449–462.

Dinarello, C.A. (1994). The biological properties of interleukin-1. *Eur Cytokine Netw* 5, 517–531.

Docampo, E., Escaramis, G., Gratacos, M., Villatoro, S., Puig, A. et al. (2014). Genome-wide analysis of single nucleotide polymorphisms and copy number variants in fibromyalgia suggest a role for the central nervous system. *Pain* 155, 1102–1109.

Doehring, A., Freynhagen, R., Griessinger, N., Zimmermann, M., Sittl, R., Hentig, N., Geisslinger, G., Lötsch, J. (2009). Cross-sectional assessment of the consequences of a GTP cyclohydrolase 1 haplotype for specialized tertiary outpatient pain care. *Clin J Pain* 25, 781–785.

Dupont, B., Ceppellini, R. (1989). *Immunobiology of HLA* (New York: Springer-Verlag).

Edwards, R.R. (2006). Genetic predictors of acute and chronic pain. *Curr Rheumatol Rep* 8, 411–417.

Elliott, A.M., Smith, B.H., Penny, K.I., Smith, W.C., Chambers, W.A. (1999). The epidemiology of chronic pain in the community. *Lancet* 354, 1248–1252.

Faber, C.G., Lauria, G., Merkies, I.S., Cheng, X., Han, C. et al. (2012). Gain-of-function Nav1.8 mutations in painful neuropathy. *Proc Natl Acad Sci USA* 109, 19444–19449.

Faserl, K., Golderer, G., Kremser, L., Lindner, H., Sarg, B., Wildt, L., Seeber, B. (2011). Polymorphism in vitamin D-binding protein as a genetic risk factor in the pathogenesis of endometriosis. *J Clin Endocrinol Metab* 96, E233–E241.

Fisher, R.A. (1922). On the interpretation of chi square from contingency tables, and the calculation of P. *J Roy Stat Soc* 85, 87–94.

Foster, D.C., Sazenski, T.M., Stodgell, C.J. (2004). Impact of genetic variation in interleukin-1 receptor antagonist and melanocortin-1 receptor genes on vulvar vestibulitis syndrome. *J Reprod Med* 49, 503–509.

Franchimont, D., Martens, H., Hagelstein, M.T., Louis, E., Dewe, W., Chrousos, G.P., Belaiche, J., Geenen, V. (1999). Tumor necrosis factor alpha decreases, and interleukin-10 increases, the sensitivity of human monocytes to dexamethasone: Potential regulation of the glucocorticoid receptor. *J Clin Endocrinol Metab* 84, 2834–2839.

Fridovich, I. (1978). The biology of oxygen radicals. *Science* 201, 875–880.

Gan, X.L., Lin, Y.H., Zhang, Y., Yu, T.H., Hu, L.N. (2010). Association of an interleukin-16 gene polymorphism with the risk and pain phenotype of endometriosis. *DNA Cell Biol* 29, 663–667.

Gendron, F.P., Chalimoniuk, M., Strosznajder, J., Shen, S., Gonzalez, F.A., Weisman, G.A., Sun, G.Y. (2003). P2X7 nucleotide receptor activation enhances IFN gamma-induced type II nitric oxide synthase activity in BV-2 microglial cells. *J Neurochem* 87, 344–352.

Gibson, G., Muse, S.V. (2009). *A Primer of Genome Science* (Sunderland, Massachusetts: Sinauer Associates).

Godinova, A.M. (1965). Genetic analysis of migraine. *Zh Nevropatol Psikhiatr Im S S Korsakova* 65, 1132–1138.

Govindan, S., Shaik, N.A., Vedicherla, B., Kodati, V., Rao, K.P., Hasan, Q. (2009). Estrogen receptor-alpha gene (T/C) Pvu II polymorphism in endometriosis and uterine fibroids. *Dis Markers* 26, 149–154.

Gullo, L., Tabacchi, P.L., Corazza, G.R., Calanca, F., Campione, O., Labo, G. (1982). HLA-B13 and chronic calcific pancreatitis. *Dig Dis Sci* 27, 214–216.

Guo, T.M., Liu, M., Zhang, Y.G., Guo, W.T., Wu, S.X. (2011). Association between Caspase-9 promoter region polymorphisms and discogenic low back pain. *Connect Tissue Res* 52, 133–138.

Gursoy, S., Erdal, E., Sezgin, M., Barlas, I.O., Aydeniz, A., Alasehirli, B., Sahin, G. (2008). Which genotype of MAO gene that the patients have are likely to be most susceptible to the symptoms of fibromyalgia? *Rheumatol Int* 28, 307–311.

van Hecke, O., Torrance, N., Smith, B.H. (2013). Chronic pain epidemiology and its clinical relevance. *Br J Anaesth* 111, 13–18.

von Hehn, C.A., Baron, R., Woolf, C.J. (2012). Deconstructing the neuropathic pain phenotype to reveal neural mechanisms. *Neuron* 73, 638–652.

Herken, H., Erdal, E., Mutlu, N., Barlas, O., Cataloluk, O., Oz, F., Guray, E. (2001). Possible association of temporomandibular joint pain and dysfunction with a polymorphism in the serotonin transporter gene. *Am J Orthod Dentofacial Orthop* 120, 308–313.

Herlyn, P., Muller-Hilke, B., Wendt, M., Hecker, M., Mittlmeier, T., Gradl, G. (2010). Frequencies of polymorphisms in cytokines, neurotransmitters and adrenergic receptors in patients with complex regional pain syndrome type I after distal radial fracture. *Clin J Pain* 26, 175–181.

Hochberg, Y. (1988). A sharper bonferroni procedure for multiple tests of significance. *Biometrika* 75, 800–802.

Hocking, L.J., Smith, B.H., Jones, G.T., Reid, D.M., Strachan, D.P., Macfarlane, G.J. (2010). Genetic variation in the beta2-adrenergic receptor but not catecholamine-O-methyltransferase predisposes to chronic pain: Results from the 1958 British Birth Cohort Study. *Pain* 149, 143–151.

Holliday, K.L., Nicholl, B.I., Macfarlane, G.J., Thomson, W., Davies, K.A., McBeth, J. (2010). Genetic variation in the hypothalamic-pituitary-adrenal stress axis influences susceptibility to musculoskeletal pain: Results from the EPIFUND study. *Ann Rheum Dis* 69, 556–560.

Hooten, W.M., Hartman, W.R., Black, J.L., Laures, H.J., Walker, D.L. (2013). Associations between serotonin transporter gene polymorphisms and heat pain perception in adults with chronic pain. *BMC Med Genet* 14, 78.

Izenmann, A. (2009). *Modern Multivariate Statistical Techniques* (Berlin: Springer).

Jeremias, J., Ledger, W.J., Witkin, S.S. (2000). Interleukin 1 receptor antagonist gene polymorphism in women with vulvar vestibulitis. *Am J Obstet Gynecol* 182, 283–285.

Jia, Z.J., Wu, F.X., Huang, Q.H., Liu, J.M. (2012). Toll-like receptor 4: The potential therapeutic target for neuropathic pain. *Zhongguo Yi Xue Ke Xue Yuan Xue Bao* 34, 168–173.

Juran, J.M. (1975). The non-Pareto principle; Mea culpa. *Quality Progress* 8, 8–9.

Kadetoff, D., Lampa, J., Westman, M., Andersson, M., Kosek, E. (2012). Evidence of central inflammation in fibromyalgia-increased cerebrospinal fluid interleukin-8 levels. *J Neuroimmunol* 242, 33–38.

Kales, S.N., Linos, A., Chatzis, C., Sai, Y., Halla, M., Nasioulas, G., Christiani, D.C. (2004). The role of collagen IX tryptophan polymorphisms in symptomatic intervertebral disc disease in Southern European patients. *Spine (Phila Pa 1976)* 29, 1266–1270.

Kang, S.C., Lee, D.G., Choi, J.H., Kim, S.T., Kim, Y.K., Ahn, H.J. (2007). Association between estrogen receptor polymorphism and pain susceptibility in female temporomandibular joint osteoarthritis patients. *Int J Oral Maxillofac Surg* 36, 391–394.

Karling, P., Danielsson, A., Wikgren, M., Soderstrom, I., Del-Favero, J., Adolfsson, R., Norrback, K.F. (2011). The relationship between the val158met catechol-O-methyltransferase (COMT) polymorphism and irritable bowel syndrome. *PLoS ONE* 6, e18035.

Katusic, Z.S., Stelter, A., Milstien, S. (1998). Cytokines stimulate GTP cyclohydrolase I gene expression in cultured human umbilical vein endothelial cells. *Arterioscler Thromb Vasc Biol* 18, 27–32.

Kilpatrick, L.A., Labus, J.S., Coveleskie, K., Hammer, C., Rappold, G. et al. (2011). The HTR3A polymorphism c. -42C>T is associated with amygdala responsiveness in patients with irritable bowel syndrome. *Gastroenterology* 140, 1943–1951.

Kim, H.J., Camilleri, M., Carlson, P.J., Cremonini, F., Ferber, I. et al. (2004). Association of distinct alpha(2) adrenoceptor and serotonin transporter polymorphisms with constipation and somatic symptoms in functional gastrointestinal disorders. *Gut* 53, 829–837.

Kim, D.H., Dai, F., Belfer, I., Banco, R.J., Martha, J.F. et al. (2010a). Polymorphic variation of the guanosine triphosphate cyclohydrolase 1 gene predicts outcome in patients undergoing surgical treatment for lumbar degenerative disc disease. *Spine (Phila Pa 1976)* 35, 1909–1914.

Kim, D.H., Lee, S.H., Kim, K.T., Yu, S.D. (2010b). Association of interleukin-1 receptor antagonist gene polymorphism with response to conservative treatment of lumbar herniated nucleus pulposus. *Spine (Phila Pa 1976)* 35, 1527–1531.

Kirk, K.M., Doege, K.J., Hecht, J., Bellamy, N., Martin, N.G. (2003). Osteoarthritis of the hands, hips and knees in an Australian twin sample–evidence of association with the aggrecan VNTR polymorphism. *Twin Res* 6, 62–66.

Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. *Biol Cybernet* 43, 59–69.

Kringel, D., Lötsch, J. (2015). Pain research funding by the European Union Seventh Framework Programme. *Eur J Pain* 19, 595–600.

Kunz, M., Hennig, J., Karmann, A.J., Lautenbacher, S. (2016). Relationship of 5-HTTLPR polymorphism with various factors of pain processing: Subjective experience. Motor responsiveness and catastrophizing. *PLoS ONE* 11, e0153089.

Lacroix-Fralish, M.L., Ledoux, J.B., Mogil, J.S. (2007). The Pain Genes Database: An interactive web browser of pain-related transgenic knockout studies. *Pain* 131(3), e1–e4.

Lampa, J., Westman, M., Kadetoff, D., Agreus, A.N., Le Maitre, E. et al. (2012). Peripheral inflammatory disease associated with centrally activated IL-1 system in humans and mice. *Proc Natl Acad Sci USA* 109, 12728–12733.

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C. et al. (2001). Initial sequencing and analysis of the human genome. *Nature* 409, 860–921.

Leipold, E., Liebmann, L., Korenke, G.C., Heinrich, T., Giesselmann, S. et al. (2013). A de novo gain-of-function mutation in SCN11A causes loss of pain perception. *Nat Genet* 45, 1399–1404.

Li, M.O., Wan, Y.Y., Sanjabi, S., Robertson, A.K., Flavell, R.A. (2006). Transforming growth factor-beta regulation of immune responses. *Annu Rev Immunol* 24, 99–146.

Lilius, T.O., Rauhala, P.V., Kambur, O., Kalso, E.A. (2009). Modulation of morphine-induced antinociception in acute and chronic opioid treatment by ibudilast. *Anesthesiology* 111, 1356–1364.

Lin, Q., Palecek, J., Palecková, V., Peng, Y.B., Wu, J., Cui, M., Willis, W.D. (1999). Nitric oxide mediates the central sensitization of primate spinothalamic tract neurons. *J Neurophysiol* 81, 1075–1085.

Lindstedt, F., Lonsdorf, T.B., Schalling, M., Kosek, E., Ingvar, M. (2011). Perception of thermal pain and the thermal grill illusion is associated with polymorphisms in the serotonin transporter gene. *PLoS ONE* 6, e17752.

Linnstaedt, S.D., Bortsov, A.V., Soward, A.C., Swor, R., Peak, D.A. et al. (2016). CRHBP polymorphisms predict chronic pain development following motor vehicle collision. *Pain* 157, 273–279.

Lippmann, C., Kringel, D., Ultsch, A., Lötsch, J. (2018). Computational functional genomics-based approaches in analgesic drug discovery and repurposing. *Pharmacoeconomics* 19, 783–797.

Lötsch, J., Geisslinger, G. (2011). Pharmacogenetics of new analgesics. *Br J Pharmacol* 163, 447–460.

Lötsch, J., Ultsch, A. (2017a). Machine learning in pain research. *Pain* 159, 623–630.

Lötsch, J., Ultsch, A. (2014). Exploiting the structures of the U-matrix. In *Advances in Intelligent Systems and Computing*, Villmann, T., Schleif, F.-M., Kaden, M., Lange, M., eds. (Heidelberg: Springer) 248–257.

Lötsch, J., Ultsch, A. (2016a). A machine-learned computational functional genomics-based approach to drug classification. *Eur J Clin Pharmacol* 72, 1449–1461.

Lötsch, J., Ultsch, A. (2016b). Process pharmacology: A pharmacological data science approach to drug development and therapy. *CPT Pharmacometrics Syst Pharmacol* 5, 192–200.

Lötsch, J. and Ultsch, A. (2017b). Random forests followed by ABC analysis as a feature selection method for machine-learning. In *Conference of the International Federation of Classification Societies (Tokyo)*, p. 170.

Lötsch, J., Doehring, A., Mogil, J.S., Arndt, T., Geisslinger, G., Ultsch, A. (2013). Functional genomics of pain in analgesic drug development and therapy. *Pharmacol Ther* 139, 60–70.

Lötsch, J., Lippmann, C., Kringel, D., Ultsch, A. (2017). Integrated computational analysis of genes associated with human hereditary insensitivity to pain. A drug repurposing perspective. *Front Neurosci* 10, 252.

Lötsch, J., Lerch, F., Djaldetti, R., Tegeder, I., Ultsch, A. (2018a). Identification of disease-distinct complex biomarker patterns by means of unsupervised machine-learning using an interactive R toolbox (Umatrix). *BMC Big Data Analytics* 3, pp. 5 https://doi.org/10.1186/s41044-018-0032-1

Lötsch, J., Sipilä, R., Dimova, V., Kalso, E. (2018b). Machine-learned selection of psychological questionnaire items relevant to the development of persistent pain after breast cancer surgery. *Br J Anaesth* [in press].

Malfait, A.M., Seymour, A.B., Gao, F., Tortorella, M.D., Le Graverand-Gastineau, M.P. et al. (2012). A role for PACE4 in osteoarthritis pain: Evidence from human genetic association and null mutant phenotype. *Ann Rheum Dis* 71, 1042–1048.

Meulenbelt, I., Min, J.L., Bos, S., Riyazi, N., Houwing-Duistermaat, J.J. et al. (2008). Identification of DIO2 as a new susceptibility locus for symptomatic osteoarthritis. *Hum Mol Genet* 17, 1867–1875.

van Meurs, J.B., Uitterlinden, A.G., Stolk, L., Kerkhof, H.J., Hofman, A., Pols, H.A., Bierma-Zeinstra, S.M. (2009). A functional polymorphism in the catechol-O-methyltransferase gene is associated with osteoarthritis-related pain. *Arthritis Rheum* 60, 628–629.

Midha, S., Khajuria, R., Shastri, S., Kabra, M., Garg, P.K. (2010). Idiopathic chronic pancreatitis in India: Phenotypic characterisation and strong genetic susceptibility due to SPINK1 and CFTR gene mutations. *Gut* 59, 800–807.

Miyamoto, Y., Shi, D., Nakajima, M., Ozaki, K., Sudo, A. et al. (2008). Common variants in DVWA on chromosome 3p24.3 are associated with susceptibility to knee osteoarthritis. *Nat Genet* 40, 994–998.

Mocellin, S., Marincola, F., Rossi, C.R., Nitti, D., Lise, M. (2004). The multifaceted relationship between IL-10 and adaptive immunity: Putting together the pieces of a puzzle. *Cytokine Growth Factor Rev* 15, 61–76.

Mogil, J.S. (2012). Pain genetics: Past, present and future. *Trends Genet* 28, 258–266.

Moore, R.A., Straube, S., Wiffen, P.J., Derry, S., McQuay, H.J. (2009). Pregabalin for acute and chronic pain in adults. *Cochrane Database Syst Rev* CD007076. https://doi.org/10.1002/14651858

Moore, R.A., Derry, S., Aldington, D., Cole, P., Wiffen, P.J. (2012). Amitriptyline for neuropathic pain and fibromyalgia in adults. *Cochrane Database Syst Rev* 12, CD008242.

Mototani, H., Iida, A., Nakajima, M., Furuichi, T., Miyamoto, Y. et al. (2008). A functional SNP in EDG2 increases susceptibility to knee osteoarthritis in Japanese. *Hum Mol Genet* 17, 1790–1797.

Murphy, K.P. (2012). *Machine Learning: A Probabilistic Perspective* (Cambridge, MA, USA: The MIT Press).

Nakamura, T., Shi, D., Tzetis, M., Rodriguez-Lopez, J., Miyamoto, Y. et al. (2007). Meta-analysis of association between the ASPN D-repeat and osteoarthritis. *Hum Mol Genet* 16, 1676–1681.

Neely, G.G., Hess, A., Costigan, M., Keene, A.C., Goulas, S. et al. (2010). A genome-wide Drosophila screen for heat nociception identifies alpha2delta3 as an evolutionarily conserved pain gene. *Cell* 143, 628–638.

Nemeth, E., Rivera, S., Gabayan, V., Keller, C., Taudorf, S., Pedersen, B.K., Ganz, T. (2004). IL-6 mediates hypoferremia of inflammation by inducing the synthesis of the iron regulatory hormone hepcidin. *J Clin Invest* 113, 1271–1276.

Nicholl, B.I., Holliday, K.L., Macfarlane, G.J., Thomson, W., Davies, K.A. et al. (2011). Association of HTR2A polymorphisms with chronic widespread pain and the extent of musculoskeletal pain: Results from two population-based cohorts. *Arthritis Rheum* 63, 810–818.

Nissenbaum, J., Devor, M., Seltzer, Z., Gebauer, M., Michaelis, M. et al. (2010). Susceptibility to chronic pain following nerve injury is genetically affected by CACNG2. *Genome Res* 20, 1180–1190.

Noponen-Hietala, N., Virtanen, I., Karttunen, R., Schwenke, S., Jakkula, E. et al. (2005). Genetic variations in IL6 associate with intervertebral disc disease characterized by sciatica. *Pain* 114, 186–194.

Oen, K., Malleson, P.N., Cabral, D.A., Rosenberg, A.M., Petty, R.E., Nickerson, P., Reed, M. (2005). Cytokine genotypes correlate with pain and radiologically defined joint damage in patients with juvenile rheumatoid arthritis. *Rheumatology (Oxford)* 44, 1115–1121.

Oshima, T., Nakajima, S., Yokoyama, T., Toyoshima, F., Sakurai, J. et al. (2010). The G-protein beta3 subunit 825 TT genotype is associated with epigastric pain syndrome-like dyspepsia. *BMC Med Genet* 11, 13.

Ozawa, A., Sasao, Y., Iwashita, K., Miyahara, M., Sugai, J. et al. (1999). HLA-A33 and -B44 and susceptibility to postherpetic neuralgia (PHN). *Tissue Antigens* 53, 263–268.

Park, J.M., Choi, M.G., Cho, Y.K., Lee, I.S., Kim, S.W., Choi, K.Y., Chung, I.S. (2011). Cannabinoid receptor 1 gene polymorphism and irritable bowel syndrome in the Korean population: A hypothesis-generating study. *J Clin Gastroenterol* 45, 45–49.

Pata, C., Erdal, E., Yazc, K., Camdeviren, H., Ozkaya, M., Ulu, O. (2004). Association of the -1438 G/A and 102 T/C polymorphism of the 5-Ht2A receptor gene with irritable bowel syndrome 5-Ht2A gene polymorphism in irritable bowel syndrome. *J Clin Gastroenterol* 38, 561–566.

Peters, M.J., Broer, L., Willemen, H.L., Eiriksdottir, G., Hocking, L.J. et al. (2013). Genome-wide association study meta-analysis of chronic widespread pain: Evidence for involvement of the 5p15.2 region. *Ann Rheum Dis* 72, 427–436.

R Development Core Team. (2008). R: A Language and Environment for Statistical Computing.

Reeser, J.C., Payne, E., Kitchner, T., McCarty, C.A. (2011). Apolipoprotein e4 genotype increases the risk of being diagnosed with posttraumatic fibromyalgia. *PM R* 3, 193–197.

Reimann, F., Cox, J.J., Belfer, I., Diatchenko, L., Zaykin, D.V. et al. (2010). Pain perception is altered by a nucleotide polymorphism in SCN9A. *Proc Natl Acad Sci USA* 107, 5148–5153.

Renner, S.P., Ekici, A.B., Maihofner, C., Oppelt, P., Thiel, F.C. et al. (2009). Neurokinin 1 receptor gene polymorphism might be correlated with recurrence rates in endometriosis. *Gynecol Endocrinol* 25, 726–733.

Ribeiro Junior, C.L., Arruda, J.T., Silva, C.T., Moura, K.K. (2009). Analysis of p53 codon 72 gene polymorphism in Brazilian patients with endometriosis. *Genet Mol Res* 8, 494–499.

Ribeiro-Dasilva, M.C., Peres Line, S.R., Godoy, L., dos Santos, M.C., Arthuri, M.T., Hou, W., Fillingim, R.B., Rizzatti Barbosa, C.M. (2009). Estrogen receptor-alpha polymorphisms and predisposition to TMJ disorder. *J Pain* 10, 527–533.

Riley, D.E., Krieger, J.N. (2002). X Chromosomal short tandem repeat polymorphisms near the phosphoglycerate kinase gene in men with chronic prostatitis. *Biochim Biophys Acta* 1586, 99–107.

Rommel, O., Kley, R.A., Dekomien, G., Epplen, J.T., Vorgerd, M., Hasenbring, M. (2006). Muscle pain in myophosphorylase deficiency (McArdle's disease): The role of gender, genotype, and pain-related coping. *Pain* 124, 295–304.

de Rooij, A.M., Florencia Gosso, M., Haasnoot, G.W., Marinus, J., Verduijn, W., Claas, F.H., van den Maagdenberg, A.M., van Hilten, J.J. (2009). HLA-B62 and HLA-DQ8 are associated with Complex Regional Pain Syndrome with fixed dystonia. *Pain* 145, 82–85.

Rosen, S., Ham, B., Mogil, J.S. (2017). Sex differences in neuroimmunity and pain. *J Neurosci Res* 95, 500–508.

Saito, Y.A., Strege, P.R., Tester, D.J., Locke, G.R. III, Talley, N.J. et al. (2009). Sodium channel mutation in irritable bowel syndrome: Evidence for an ion channelopathy. *Am J Physiol Gastrointest Liver Physiol* 296, G211–G218.

Sato, M., Ohashi, J., Tsuchiya, N., Kashiwase, K., Ishikawa, Y. et al. (2002). Association of HLA-A*3303-B*4403-DRB1*1302 haplotype, but not of TNFA promoter and NKp30 polymorphism, with postherpetic neuralgia (PHN) in the Japanese population. *Genes Immun* 3, 477–481.

Saura, M., Zaragoza, C., Herranz, B., Griera, M., Diez-Marques, L., Rodriguez-Puyol, D., Rodriguez-Puyol, M. (2005). Nitric oxide regulates transforming growth factor-beta signaling in endothelial cells. *Circ Res* 97, 1115–1123.

Schott, K., Gutlich, M., Ziegler, I. (1993). Induction of GTP-cyclohydrolase I mRNA expression by lectin activation and

interferon-gamma treatment in human cells associated with the immune response. *J Cell Physiol* 156, 12–16.
Schwartz, E.S., Lee, I., Chung, K., Chung, J.M. (2008). Oxidative stress in the spinal cord is an important contributor in capsaicin-induced mechanical secondary hyperalgesia in mice. *Pain* 138, 514–524.
Schwartz, E.S., Kim, H.Y., Wang, J., Lee, I., Klann, E., Chung, J.M., Chung, K. (2009). Persistent pain is dependent on spinal mitochondrial antioxidant levels. *J Neurosci* 29, 159–168.
Seki, S., Kawaguchi, Y., Chiba, K., Mikami, Y., Kizawa, H. et al. (2005). A functional SNP in CILP, encoding cartilage intermediate layer protein, is associated with susceptibility to lumbar disc disease. *Nat Genet* 37, 607–612.
Shaik, N.A., Govindan, S., Kodati, V., Rao, K.P., Hasan, Q. (2009). Polymorphic (CAG)n repeats in the androgen receptor gene: A risk marker for endometriosis and uterine leiomyomas. *Hematol Oncol Stem Cell Ther* 2, 289–293.
Shatzky, S., Moses, S., Levy, J., Pinsk, V., Hershkovitz, E. et al. (2000). Congenital insensitivity to pain with anhidrosis (CIPA) in Israeli-Bedouins: Genetic heterogeneity, novel mutations in the TRKA/NGF receptor gene, clinical findings, and results of nerve conduction studies. *Am J Med Genet* 92, 353–360.
Shoskes, D.A., Albakri, Q., Thomas, K., Cook, D. (2002). Cytokine polymorphisms in men with chronic prostatitis/chronic pelvic pain syndrome: Association with diagnosis and treatment response. *J Urol* 168, 331–335.
Smith, S.B., Maixner, D.W., Greenspan, J.D., Dubner, R., Fillingim, R.B. et al. (2011). Potential genetic risk factors for chronic TMD: Genetic associations from the OPPERA case control study. *J Pain* 12, T92–T101.
Smith, S.B., Maixner, D.W., Fillingim, R.B., Slade, G., Gracely, R.H. et al. (2012). Large candidate gene association study reveals genetic risk factors and therapeutic targets for fibromyalgia. *Arthritis Rheum* 64, 584–593.
Solovieva, S., Leino-Arjas, P., Saarela, J., Luoma, K., Raininko, R., Riihimaki, H. (2004). Possible association of interleukin 1 gene locus polymorphisms with low back pain. *Pain* 109, 8–19.
Sorge, R.E., Trang, T., Dorfman, R., Smith, S.B., Beggs, S. et al. (2012). Genetically determined P2X7 receptor pore formation regulates variability in chronic pain sensitivity. *Nat Med* 18, 595–599.
Stephens, K., Cooper, B.A., West, C., Paul, S.M., Baggott, C.R. et al. (2014). Associations between cytokine gene variations and severe persistent breast pain in women following breast cancer surgery. *J Pain* 15, 169–180.
Sugaya, K., Nishijima, S., Yamada, T., Miyazato, M., Hatano, T., Ogawa, Y. (2002). Molecular analysis of adrenergic receptor genes and interleukin-4/interleukin-4 receptor genes in patients with interstitial cystitis. *J Urol* 168, 2668–2671.
Tahara, T., Arisawa, T., Shibata, T., Nakamura, M., Wang, F., Hirata, I. (2008). COMT gene val158met polymorphism in patients with dyspeptic symptoms. *Hepatogastroenterology* 55, 979–982.
Tegeder, I., Lötsch, J. (2009). Current evidence for a modulation of low back pain by human genetic variants. *J Cell Mol Med* 13(8B), 1605–19. https://doi.org/10.1111/j.1582-4934.2009.00703.x. Epub 2009 Feb 17.
Tegeder, I., Costigan, M., Griffin, R.S., Abele, A., Belfer, I. et al. (2006). GTP cyclohydrolase and tetrahydrobiopterin regulate pain sensitivity and persistence. *Nat Med* 12, 1269–1277.
Tegeder, I., Adolph, J., Schmidt, H., Woolf, C.J., Geisslinger, G., Lötsch, J. (2008). Reduced hyperalgesia in homozygous carriers of a GTP cyclohydrolase 1 haplotype. *Eur J Pain* 12, 1069–1077.
Thulasiraman, K., Swamy, M.N.S. (1992). *Graphs: Theory and Algorithms* (New York, NY: Wiley).
Tian, L., Ma, L., Kaarela, T., Li, Z. (2012). Neuroimmune crosstalk in the central nervous system and its significance for neurological diseases. *J Neuroinflammation* 9, 155.
Tilkeridis, C., Bei, T., Garantziotis, S., Stratakis, C.A. (2005). Association of a COL1A1 polymorphism with lumbar disc disease in young military recruits. *J Med Genet* 42, e44.
Ultsch, A. (2003). Maps for visualization of high-dimensional data spaces. In *WSOM* (Kyushu, Japan), pp. 225–230.
Ultsch, A. (2005). Clustering with SOM: U*C. In *Workshop on Self-Organizing Maps (Paris)*, pp. 75–82.
Ultsch, A., Lötsch, J. (2014). Functional abstraction as a method to discover knowledge in gene ontologies. *PLoS ONE* 9, e90191.
Ultsch, A., Lötsch, J. (2015). Computed ABC analysis for rational selection of most informative variables in multivariate data. *PLoS ONE* 10, e0129767.
Ultsch, A., Lötsch, J. (2017). Machine-learned cluster identification in high-dimensional data. *J Biomed Inform* 66, 95–104.
Ultsch, A., Sieman, H.P. (1990). Kohonen's self organizing feature maps for exploratory data analysis. *INNC'90, Int Neural Network Conference (Dordrecht* (Netherlands: Kluwer)) 305–308.
Ultsch, A., Kringel, D., Kalso, E., Mogil, J.S., Lötsch, J. (2016). A data science approach to candidate gene selection of pain regarded as a process of learning and neural plasticity. *Pain* 157, 2747–2757.
Valdes, A.M., Lories, R.J., van Meurs, J.B., Kerkhof, H., Doherty, S. et al. (2009). Variation at the ANP32A gene is associated with risk of hip osteoarthritis in women. *Arthritis Rheum* 60, 2046–2054.
Valdes, A.M., Spector, T.D., Tamm, A., Kisand, K., Doherty, S.A. et al. (2010). Genetic variation in the SMAD3 gene is associated with hip and knee osteoarthritis. *Arthritis Rheum* 62, 2347–2352.
Valdes, A.M., Arden, N.K., Vaughn, F.L., Doherty, S.A., Leaverton, P.E. et al. (2011a). Role of the Nav1.7 R1150W amino acid change in susceptibility to symptomatic knee osteoarthritis and multiple regional pain. *Arthritis Care Res (Hoboken)* 63, 440–444.
Valdes, A.M., De Wilde, G., Doherty, S.A., Lories, R.J., Vaughn, F.L. et al. (2011b). The Ile585Val TRPV1 variant is involved in risk of painful knee osteoarthritis. *Ann Rheum Dis* 70, 1556–1561.
Valdes, A.M., Evangelou, E., Kerkhof, H.J., Tamm, A., Doherty, S.A. et al. (2011c). The GDF5 rs143383 polymorphism is associated with osteoarthritis of the knee with genome-wide statistical significance. *Ann Rheum Dis* 70, 873–875.
Vassalli, P. (1992). The pathophysiology of tumor necrosis factors. *Annu Rev Immunol* 10, 411–452.
Venn, J. (1880). On the Diagrammatic and Mechanical Representation of Propositions and Reasonings. *Dublin Philos Mag J Sci* 9, 1–18.
Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J. et al. (2001). The sequence of the human genome. *Science* 291, 1304–1351.
Wells, C.A., Ravasi, T., Faulkner, G.J., Carninci, P., Okazaki, Y. et al. (2003). Genetic control of the innate immune response. *BMC Immunol* 4, 5.
Witkin, S.S., Gerber, S., Ledger, W.J. (2002). Differential characterization of women with vulvar vestibulitis syndrome. *Am J Obstet Gynecol* 187, 589–594.
Woo, H.Y., Kim, K.H., Lim, S.W. (2010). Estrogen receptor 1, glutathione S-transferase P1, glutathione S-transferase M1, and glutathione S-transferase T1 genes with dysmenorrhea in Korean female adolescents. *Korean J Lab Med* 30, 76–83.
Wu, D., Wang, X., Chen, D., Niu, T., Ni, J., Liu, X., Xu, X. (2000). Metabolic gene polymorphisms and risk of dysmenorrhea. *Epidemiology* 11, 648–653.
Zhang, X., Llamado, L., Pillay, I., Price, P., Will, R. (2002). Interleukin-1 gene polymorphism disease activity and bone mineral metabolism in rheumatoid arthritis. *Chin Med J (Engl)* 115, 46–49.
Zhang, L., Rao, F., Zhang, K., Khandrika, S., Das, M. et al. (2007). Discovery of common human genetic variants of GTP cyclohydrolase 1 (GCH1) governing nitric oxide, autonomic activity, and cardiovascular risk. *J Clin Invest* 117, 2658–2671.
Zorina-Lichtenwalter, K., Meloto, C.B., Khoury, S., Diatchenko, L. (2016). Genetic predictors of human chronic pain conditions. *Neuroscience* 338, 36–62.

Erklärung zu den Autorenanteilen an der Publikation:

Development of an AmpliSeq™ Panel for Next-Generation Sequencing of a Set of Genetic Predictors of Persisting Pain (*printed*)

Name der Zeitschrift: Frontiers in Pharmacology

Beteiligte Autoren: D Kringel, M A Kaunisto, C Lippmann, E Kalso und J Lötsch

Was hat der Promovierende bzw. was haben die Koautoren beigetragen?

(1) zu Entwicklung und Planung

Promovierender: 50%
Autor JL: 50%

(2) zur Durchführung der einzelnen Untersuchungen und Experimente

Promovierender: 100% (Literatur Recherche, DNA Extraktion, NGS-Panel Design, Sequenzierung, Varianten Annotation, Methoden Validierung)

(3) zur Erstellung der Datensammlung und Abbildungen

Promovierender: 60 % (Datenerhebung, Run-Statistik, Erstellung der Abbildung Chip-Beladung und Validierung)
Autor JL: 40 (Erstellung der Abbildungen Genetische Vektoren, Panel Intersektionen und GO-Terme)

(4) zur Analyse und Interpretation der Daten

Promovierender: 60% (Datenvorverarbeitung, Charakterisierung der Varianten, klinische Assoziation, Dateninterpretation)
Autor JL: 40% (Datenanalyse, GO-Term Auswertung, Dateninterpretation)

(5) zum Verfassen des Manuskripts

Promovierender: 40%
Autor MAK: 5%
Autor CL: 5%
AutorEK: 10%
Autor JL: 40%

Zustimmende Bestätigungen der oben genannten Angaben:

_____ _____

Datum/Ort Unterschrift Promovend

_____ _____

Datum/Ort Unterschrift Betreuer

frontiers in Pharmacology

ORIGINAL RESEARCH
published: 19 September 2018
doi: 10.3389/fphar.2018.01008

Development of an AmpliSeq™ Panel for Next-Generation Sequencing of a Set of Genetic Predictors of Persisting Pain

*Dario Kringel[1], Mari A. Kaunisto[2], Catharina Lippmann[3], Eija Kalso[4] and Jörn Lötsch[1,3]**

[1] Institute of Clinical Pharmacology, Goethe-University, Frankfurt, Germany, [2] Institute for Molecular Medicine Finland, HiLIFE, University of Helsinki, Helsinki, Finland, [3] Fraunhofer Institute for Molecular Biology and Applied Ecology – Project Group Translational Medicine and Pharmacology, Frankfurt, Germany, [4] Division of Pain Medicine, Department of Anesthesiology, Intensive Care and Pain Medicine, University of Helsinki and Helsinki University Hospital, Helsinki, Finland

OPEN ACCESS

Edited by:
Ulrich M. Zanger, Dr. Margarete Fischer-Bosch-Institut für Klinische Pharmakologie (IKP), Germany

Reviewed by:
Theodora Katsila, University of Patras, Greece
Cheryl D. Cropp, Samford University, United States

***Correspondence:**
Jörn Lötsch
j.loetsch@em.uni-frankfurt.de

Specialty section:
This article was submitted to Pharmacogenetics and Pharmacogenomics, a section of the journal Frontiers in Pharmacology

Received: *24 May 2018*
Accepted: *17 August 2018*
Published: *19 September 2018*

Citation:
Kringel D, Kaunisto MA, Lippmann C, Kalso E and Lötsch J (2018) Development of an AmpliSeq™ Panel for Next-Generation Sequencing of a Set of Genetic Predictors of Persisting Pain. Front. Pharmacol. 9:1008. doi: 10.3389/fphar.2018.01008

Background: Many gene variants modulate the individual perception of pain and possibly also its persistence. The limited selection of single functional variants is increasingly being replaced by analyses of the full coding and regulatory sequences of pain-relevant genes accessible by means of next generation sequencing (NGS).

Methods: An NGS panel was created for a set of 77 human genes selected following different lines of evidence supporting their role in persisting pain. To address the role of these candidate genes, we established a sequencing assay based on a custom AmpliSeq™ panel to assess the exomic sequences in 72 subjects of Caucasian ethnicity. To identify the systems biology of the genes, the biological functions associated with these genes were assessed by means of a computational over-representation analysis.

Results: Sequencing generated a median of $2.85 \cdot 10^6$ reads per run with a mean depth close to 200 reads, mean read length of 205 called bases and an average chip loading of 71%. A total of 3,185 genetic variants were called. A computational functional genomics analysis indicated that the proposed NGS gene panel covers biological processes identified previously as characterizing the functional genomics of persisting pain.

Conclusion: Results of the NGS assay suggested that the produced nucleotide sequences are comparable to those earned with the classical Sanger sequencing technique. The assay is applicable for small to large-scale experimental setups to target the accessing of information about any nucleotide within the addressed genes in a study cohort.

Keywords: pain, data science, knowledge discovery, functional genomics, next generation sequencing (NGS)

INTRODUCTION

Persisting pain has been proposed to result from a gene environment interaction where nerve injuries or inflammatory processes act as triggers while the clinical symptoms develop only in a minority of subjects (Lee and Tracey, 2013). A role of the genetic background in pain is supported by evidence of many variants modulating the individual perception of pain and the development of its persistence (Diatchenko et al., 2005; Lötsch et al., 2009b; Mogil, 2012). Genetic variants have been reported to confer protection against pain such as the rs1799971 variant in the μ-opioid receptor gene (*OPRM1*) (Lötsch et al., 2006), or to increase the risk for persisting pain such as the rs12584920 variant of the 5-hydroxytryptamine receptor 2A gene (*HTR2A*) (Nicholl et al., 2011) or the rs734784 polymorphism in the voltage-gated potassium ion channel modifier, subfamily S member 1, gene (*KCNS1*) (Costigan et al., 2010). Nevertheless, the genetic background of persisting pain is still incompletely understood (Mogil, 2009; Lötsch and Geisslinger, 2010) and under intense discussion.

Until recently, research focused on the role of selected functional genetic variants as protective or risk factors of persisting pain. This has changed with the broader availability of next generation sequencing (NGS) (Metzker, 2010). To make use of these technical advancements, we developed a custom AmpliSeq™ library and sequencing assay for efficient detection of genetic variants possibly associated with persisting pain. We propose an assay of a set of 77 genes supported by evidence of an involvement in pain and its development toward persistence. The set size fully uses the technical specifications of the AmpliSeq™ gene sequencing library technique.

MATERIALS AND METHODS

Selection of Genes Relevant for Persisting Pain

A set of candidate genes with shown or biologically plausible relevance to persisting pain was created by applying a combination of criteria, which provided three different genetic subsets. **Subset 1** was chosen exclusively on the basis of computational functional genomics based on a recently published analysis of persisting pain regarded as displaying systemic features of learning and neuronal plasticity (Mansour et al., 2014). As discussed previously (Ultsch et al., 2016), the view of chronic pain as a dysregulation in biological processes of learning and neuronal plasticity (Alvarado et al., 2013) seems to be captured by the controlled vocabulary (Camon et al., 2004) of the Gene Ontology (GO) knowledge base by the GO terms "learning or memory" (GO:0007611)[1] and "nervous system development" (GO:0007399)[2]. An intersection of the genes annotated to these GO terms with a set of 539 "pain genes" identified empirically as relevant to pain provided the first subset of 34 genes described in detail previously (Ultsch et al., 2016). Briefly, the intersecting set of so-called "pain genes" consists of a combination of (i) genes listed in the PainGenes database (Lacroix-Fralish et al., 2007)[3], (ii) genes causally involved in human hereditary diseases associated with extreme pain phenotypes, (iii) genes found to be associated with chronic pain in at least three human studies, and (iv) genes coding for targets of novel analgesics under clinical development (Lötsch et al., 2013).

Subset 2 consisted of genes that were reported to carry variants modulating the risk or the phenotypic symptoms in at least two different clinical settings of persisting pain. They were obtained using (i) a PubMed database search for the string "(chronic OR persisting OR neuropathic OR back OR inflammatory OR musculoskeletal OR visceral OR widespread OR idiopathic OR fibromyalgia) AND pain AND (polymorphism OR variant) NOT review," to which genes highlighted in overviews on pain genetics (e.g., Edwards, 2006) were added. The intersection of the queried genes with the set of 539 "pain genes" (see above) provided a subset of 13 genes (**Table 1**).

Finally, **subset 3** comprised genes that have consistently been included in human pain research projects over the last several years. One of them is the *OPRM1* gene that codes for the human μ-opioid receptor and which has been shown to modulate the time course of persisting cancer pain by delaying the necessity of opioid treatment (Lötsch et al., 2010). However, further genes were added such as the *GDNF* gene coding for the glial cell derived neurotrophic factor, which has been shown to be involved in a glia-dependent mechanism of neuropathic pain (Wang et al., 2014) although no modulating human genetic variants have been reported so far. Following expert counseling within the EU-funded "glial-opioid interface in chronic pain, GLORIA" research consortium (Kringel and Lötsch, 2015)[4], a subset of 30 genes (**Table 1**) was identified. Thus, the complete set as the union of the three subsets comprised 43 + 13 + 30 = 77 genes that are proposed to be included in an NGS panel of human persisting pain.

DNA Sample Origin

Due to the costs of assay development (for details, see second paragraph of the Discussion), the Ampliseq™ panel was established in a limited number of n = 72 DNA samples. This corresponds to the number of samples used in comparable recent studies for NGS assay establishment and validation (Bruera et al., 2018; De Luca et al., 2018; Mustafa et al., 2018; Shah et al., 2018). To further limit the project costs, the Ampliseq™ panel was established in a subset of samples originating from a clinical cohort of 1,000 women who had undergone breast cancer surgery (Kaunisto et al., 2013; Lötsch et al., 2018). The study followed the Declaration of Helsinki and was approved by the Coordinating Ethics Committee of the Helsinki University Hospital. Each participating subject had provided a written informed consent including genetic studies.

Specifically, for the presently reported method establishment, a subsample of 72 women (age 58.4 ± 8 years, mean ± standard deviation, weight 69.3 ± 11 kg), was drawn from the clinical

[1]http://amigo.geneontology.org/amigo/term/GO:0007611

[2]http://amigo.geneontology.org/amigo/term/GO:0007399

[3]http://www.jbldesign.com/jmogil/enter.html

[4]http://gloria.helsinki.fi

TABLE 1 | Genes included in the proposed NGS panel of persisting pain, combined from three subsets included on different bases.

Gene symbol	NCBI	Gene description	Reference
Subset #1			
ADCY1	107	Adenylate cyclase 1	Vadakkan et al., 2006
BDNF	627	Brain-derived neurotrophic factor	Obata and Noguchi, 2006
CDK5	1020	Cyclin-dependent kinase 5	Yang et al., 2014
CHRNB2	1141	Cholinergic receptor, nicotinic, beta 2	Dineley et al., 2015
CNR1	1268	Cannabinoid receptor 1 (brain)	Smith et al., 1998
DLG4	1742	Disks, large homolog 4 (Drosophila)	Florio et al., 2009
DRD1	1812	Dopamine receptor D1	Onojjighofia et al., 2014
DRD2	1813	Dopamine receptor D2	Onojjighofia et al., 2014
DRD3	1814	Dopamine receptor D3	Potvin et al., 2009
EGR1	1958	Early growth response 1	Ko et al., 2005
FOS	2353	Cellular oncogene FOS	Abbadie et al., 1994
FYN	2534	Src family tyrosine kinase	Liu et al., 2014
GABRA5	2558	GABA A receptor, alpha 5	Bravo-Hernández et al., 2016
GALR2	8811	Galanin receptor 2	Hulse et al., 2012
GRIN1	2902	Glutamate receptor, NMDA 1	Petrenko et al., 2003
GRIN2A	2903	Glutamate receptor, NMDA 2A	Petrenko et al., 2003
GRIN2B	2904	Glutamate receptor, NMDA 2B	Petrenko et al., 2003
GRM5	2915	Glutamate receptor, metabotropic 5	Walker et al., 2001
HRH3	11255	Histamine receptor H3	Huang et al., 2007
KIT	3815	Tyrosine kinase KIT	Sun et al., 2009
NF1	4763	Neurofibromin 1	Wolters et al., 2015
NGF	4803	Nerve growth factor	Kumar and Mahal, 2012
NTF4	4909	Neurotrophin 4	Kumar and Mahal, 2012
NTRK1	4914	Neurotrophic tyrosine kinase 1	Kumar and Mahal, 2012
OXT	5020	Oxytocin prepropeptide	Goodin et al., 2015
PLCB1	23236	Phospholipase C, beta 1	Shi T.-J.S. et al., 2008
PRKCG	5582	Protein kinase C, gamma	Sluka and Audette, 2006
PRNP	5621	Prion protein	Gadotti and Zamponi, 2011
PTN	5764	Pleiotrophin	Gramage and Herradon, 2010
PTPRZ1	5803	Protein tyrosine phosphatase Z 1	Ultsch et al., 2016
RELN	5649	Reelin	Buchheit et al., 2012
S100B	6285	S100 calcium binding protein B	Zanette et al., 2014
SLC6A4	6532	Serotonin transporter	Offenbaecher et al., 1999
TH	7054	Tyrosine hydroxylase	Bravo et al., 2014
Subset #2			
ADRB2	154	Adrenoceptor beta 2	Hocking et al., 2010
COMT	1312	Catechol-*O*-methyltransferase	Feng et al., 2013
ESR1	2099	Extrogen Receptor 1	Ribeiro-Dasilva et al., 2009
GCH1	2643	GTP cyclohydrolase 1	Tegeder et al., 2006
IL1B	3553	Interleukin 1B	Loncar et al., 2013
IL4	3565	Interleukin 4	Sugaya et al., 2002
IL6	3569	Interleukin 6	Shoskes et al., 2002
IL10	3586	Interleukin 10	Stephens et al., 2014
P2RX7	5027	Purinergic Receptor P2X7	Sorge et al., 2012
SCN9A	6335	Sodium voltage-gated alpha subunit 9	Reimann et al., 2010
SOD2	6648	Superoxide dismutase 2	Schwartz et al., 2009
TNF	7124	Tumor necrosis factor	Leung and Cahill, 2010
TRPV1	7442	Transient receptor potential cation channel, subfamily V, member 1	Bourinet et al., 2014
Subset #3			
ABHD12	26090	Abhydrolase domain containing 12	Kim, 2015
ABHD16A	7920	Abhydrolase domain containing 16A	Kim, 2015
ABHD6	57406	Abhydrolase domain containing 6	Kim, 2015

(Continued)

TABLE 1 | Continued

Gene symbol	NCBI	Gene description	Reference
CACNG2	10369	Calcium voltage-gated channel auxiliary subunit gamma 2	Nissenbaum et al., 2010
CSF1	1435	Colony stimulating factor 1	Thuault, 2016
DRD4	1815	Dopamine receptor D4	Buskila et al., 2004
FAAH	2166	Fatty acid amide hydrolase	Jayamanne et al., 2006
FKBP5	2289	Fk506 binding protein 5	Fujii et al., 2014
GDNF	2668	Glial cell derived neurotrophic factor	Sah et al., 2005
GFRA1	2674	GDNF family receptor alpha 1	Yamamoto et al., 2003
GPR132	29933	G protein-coupled receptor 132	Hohmann et al., 2017
HCN2	610	Hyperpolarization-activated cyclic nucleotide-gated	Tsantoulas et al., 2016
HLA-DQB1	3119	Major histocompatibility complex, class II, DQ beta 1	Dominguez et al., 2013
HLA-DRB1	3123	Major histocompatibility complex, class II, DR beta 1	Dominguez et al., 2013
HTR1A	3350	5-hydroxytryptamine (serotonin) receptor 1A	Lindstedt et al., 2012
HTR2A	3356	5-hydroxytryptamine (serotonin) receptor 2A	Nicholl et al., 2011
IL1R2	7850	Interleukin 1 receptor type 2	Stephens et al., 2014
KCNS1	3787	Potassium voltage-gated channel, modifier subfamily S, member 1	Costigan et al., 2010
LTB4R	1241	Leukotriene b4 receptor	Zinn et al., 2017
LTB4R2	56413	Leukotriene b4 receptor 2	Zinn et al., 2017
OPRD1	4985	Opioid receptor delta 1	Law et al., 2013
OPRK1	4986	Opioid receptor kappa 1	Guerrero et al., 2010
OPRM1	4988	Opioid receptor mu 1	Lötsch and Geisslinger, 2005
RET	5979	RET receptor tyrosine kinase	Snider and McMahon, 1998
RUNX1	861	Runt related transcription factor 1	Chen et al., 2006
TLR4	7099	Toll like Receptor 4	Hutchinson et al., 2010
TRPA1	8989	Transient receptor potential cation channel, subfamily A, member 1	Bourinet et al., 2014
TRPM8	79054	Transient receptor potential cation channel, subfamily M, member 8	Bourinet et al., 2014
TRPV4	59341	Transient receptor potential cation channel, subfamily V, member 4	Bourinet et al., 2014
TSPO	706	Translocator protein	Loggia et al., 2015

Subset #1 comprises d = 34 genes that had resulted from a computational functional genomics analysis (Ultsch et al., 2016) pursuing the hypothesis that persisting pain displays systemic features of learning and neuronal plasticity (Mansour et al., 2014). Hence, from a set of genes identified empirically as relevant to pain and listed in the PainGenes database (http://www.jbldesign.com/jmogil/enter.html, Lacroix-Fralish et al., 2007), those were selected that are annotated to the Gene Ontology (Ashburner et al., 2000) terms "learning or memory" and "nervous system development." The references are those found to provide evidence for an association with pain, except for PTPRZ1 that was a novel finding in (Ultsch et al., 2016). Subset #2 comprises d = 13 genes identified empirically as relevant to pain and listed in the PainGenes database (http://www.jbldesign.com/jmogil/enter.html, Lacroix-Fralish et al., 2007) and reported to carry variants that modulated the risk or the symptomatology in at least two different clinical settings of persisting paint. Subset #3 comprises d = 30 genes repeatedly shown during the last several years to play a role in the human genetics of persisting pain or recently reported as novel players.

subgroup not having developed persisting pain during the observation period. This was believed to come closer to a random sample than a mixture of patients with persisting and without persisting pain. This limitation of the sample selection has probably affected which and how many variants were identified. However, it is unlikely to have jeopardized the general applicability of the gene selection heuristics, assay establishment and validation, and of the functional analysis of the selected subset of genes.

DNA Template Preparation and Amplification

A multiplex PCR amplification strategy for the coding gene sequences was accomplished online (Ion Ampliseq™ Designer)[5] to amplify the target region specified above (for primer sequences, see **Supplementary Table 1**) with 25 base pair exon padding. After a comparison of several primer design options, the design providing the maximum target sequence coverage was chosen. The ordered 1,953 amplicons covered approximately 97.5% of the target sequence (**Supplementary Table 2**). A total of 10 ng DNA per sample was used for the target enrichment by a multiplex PCR and each DNA pool was amplified with the Ion Ampliseq™ Library Kit in conjunction with the Ion Ampliseq™ "custom Primer Pool"-protocols according to the manufacturer's procedures (Life Technologies, Darmstadt, Germany).

After each pool had undergone 18 PCR cycles, the PCR primers were removed with FuPa Reagent and the amplicons were ligated to the sequencing adaptors with short stretches of index sequences (barcodes) that enabled sample multiplexing for subsequent steps (Ion Xpress™ Barcode Adapters Kit; Life Technologies). After purification with AMPure XP beads (Beckman Coulter, Krefeld, Germany), the barcoded libraries were quantified with a Qubit® 2.0 Fluorimeter (Life Technologies, Darmstadt, Germany) and normalized for DNA concentration to a final concentration of 20 pmol/l using the Ion Library Equalizer™ Kit (Life Technologies, Darmstadt, Germany).

[5]http://www.ampliseq.com

Equalized barcoded libraries from seven to eight samples at a time were pooled. To clonally amplify the library DNA onto the Ion Sphere Particles (ISPs; Life Technologies, Darmstadt, Germany), the library pool was subjected to emulsion PCR by using an Ion PGM HI-Q View Template Kit on an PGM OneTouch system (Life Technologies, Darmstadt, Germany) following the manufacturer's protocol.

Sequencing

Enriched ISPs which carried many copies of the same DNA fragment were subjected to sequencing on an Ion 318 Chip to sequence pooled libraries with seven to eight samples. During this process, bases are inferred from light intensity signals, a process commonly referred to as base-calling (Ledergerber and Dessimoz, 2011). The number of combined libraries that can be accommodated in a single sequencing run depends on the size of the chip, the balance of barcoded library concentration, and the coverage required. The high-capacity 318 chip was chosen (instead of the low-capacity 314 or the medium-capacity 316 chip) to obtain a high sequencing depth of coverage for a genomic DNA library with >95% of bases at 30x. Sequencing was performed using the sequencing kit (Ion PGM Hi-Q Sequencing Kit; Life Technologies, Darmstadt, Germany) as per the manufacturer's instructions with the 200 bp single-end run configuration. This kit contained the most advanced sequencing chemistry available to users of the Ion PGM System (Life Technologies, Darmstadt, Germany).

Data Analysis

Bioinformatics Generation of Sequence Information

The raw data (unmapped BAM-files) from the sequencing runs were processed using Torrent Suite Software (Version 5.2.2, Life Technologies, Darmstadt, Germany) to generate read alignments which were filtered by the software into mapped BAM-files using the reference genomic sequence (hg19) of target genes. Variant calling was performed with the Torrent Variant Caller Plugin using as key parameters: minimum allele frequency = 0.15, minimum quality = 10, minimum coverage = 20 and minimum coverage on either strand = 3.

The annotation of called variants was done using the Ion Reporter Software (Version 4.4; Life Technologies, Darmstadt, Germany) for the VCF files that contained the nucleotide reads and the GenomeBrowse® software (Version 2.0.4, Golden Helix, Bozeman, MT, United States) to map the sequences to the reference sequences GRCh37 hg19 (dated February 2009). The SNP and Variation Suite software (Version 8.4.4; Golden Helix, Bozeman, MT, United States) was used for the analysis of sequence quality, coverage and for variant identification.

Based on the observed allelic frequency, the expected number of homozygous and heterozygous carriers of the respective SNP (single nucleotide polymorphism) was calculated using the Hardy-Weinberg equation. Only variants within the Hardy-Weinberg equilibrium as assessed using Fisher's exact test (Emigh, 1980) were retained. The SNP and Variation Suite software (Version 8.4.4; Golden Helix, Bozeman, MT, United States) was used for the analysis of sequence quality, coverage and for variant identification.

Assay Validation

Method validation was accomplished by means of Sanger sequencing (Sanger and Coulson, 1975; Sanger et al., 1977) in an independent external laboratory (Eurofins Genomics, Ebersberg, Germany). As performed previously with different AmpliSeq^{TM} panels (Kringel et al., 2017) and other genotyping assays (Skarke et al., 2004, 2005), four DNA samples have been chosen randomly from an independent cohort of healthy subjects and sequenced with the current NGS panel. For the detected variant type, single nucleotide polymorphisms from five different genomic regions for which clinical associations have been reported (**Table 2**), i.e., rs324420 (*FAAH*), rs333970 (*CSF1*), rs4986790 (*TLR4*), rs4633 (*COMT*), and rs17151558 (*RELN*) were chosen for external sequencing. Amplification of the respective DNA segments was done using PCR primer pairs (forward, reverse) of (i) 5′-TTTCTTAAAAAGGCCAGCCTCCT-3′ and 5′-AATGACCCAAGATGCAGAGCA-3′ (ii) 5′-GCCTT CAACCCCGGGATGG-3′ and 5′-CTCCGATCCCTGGTGC TCCTC-3′ (iii) 5′-TTTATTGCACAGACTTGCGGGTTC-3′ and 5′-AGCCTTTTGAGAGATTTGAGTTTCA-3′ (iv) 5′-CC TTATCGGCTGGAACGAGTT-3′ and 5′-GTAAGGGCTTT GATGCCTGGT-3′ (v) 5′-GTTATTCCTCTGTAAGCAGCTGCC T-3′ and 5′-TGTTTGTTTTAGATTGTGGTGGGTT-3′. Results of Sanger sequencing were aligned with the genomic sequence and analyzed using Chromas Lite® (Version 2.1.1, Technelysium Pty Ltd, South Brisbane, QLD, Australia) and the GenomeBrowse® (Version 2.0.4, Golden Helix, Bozeman, MT, United States) was used to compare the sequences obtained with NGS or Sanger techniques.

RESULTS

The NGS assay of the proposed set of 77 human genes relevant to persisting pain was established in 72 genomic DNA samples. As applied previously (Kringel et al., 2017), only exons including 25 bases of padding around all targeted coding regions for which the realized read-depths for each nucleotide was higher than 20 were contemplated as successfully analyzed. With this acceptance criterion the whole or almost whole coverage of the relevant sequences was obtained (**Table 1**; for details on missing variants, see **Supplementary Table 3**). The NGS sequencing process of the whole patient cohort required ten separate runs, each with samples of $n = 7$ or $n = 8$ patients. Coverage statistics were analogous between all runs and matched the scope of accepted quality levels [20–22]. A median of $2.85 \cdot 10^6$ reads per run was produced. The mean depth was close to 200 reads, the mean read length of called bases resulted in 205 bases and average chip loading was 71% (**Figure 1A**). To establish a sequencing output with a high density of ISPs on a sequencing chip, the chip loading value should exceed 60% (Life Technologies, Carlsbad, United States). The generated results of all NGS runs matched with the results obtained with Sanger sequencing of random samples (**Figure 1B**), meaning the accordance of nucleotide sequences between NGS and Sanger sequencing was 100% in all validated samples.

TABLE 2 | A list of coding human variants in the 77 putative chronic pain genes, found in the present random sample of 72 subjects of Caucasian ethnicity, for which clinical associations have been reported.

Gene	Variant	dbSNP# accession number	Known clinical association	Reference
Pain context				
FAAH	1:46870761-SNV	rs324420	Effect of endocannabinoid degradation on pain	Cajanus et al., 2016
FAAH	1:46870761-SNV	rs324420	Cold and heat pain sensitivity	Kim et al., 2006b
CSF1	1:110466338-SNV	rs333970	Rheumatoid arthritis	Solus et al., 2015
NGF	1:115829313-SNV	rs6330	Procedural pain	Ersig et al., 2017
NGF	1:115829313-SNV	rs6330	Susceptibility to migraine	Coskun et al., 2016
IL1B	2:113590966-SNV	rs1143634	Adverse effects in postoperative pain	Somogyi et al., 2016
IL1B	2:113590966-SNV	rs1143634	Low back pain	Feng et al., 2016
SCN9A	2:167099158-SNV	rs6746030	Pain susceptibility in Parkinson disease	Greenbaum et al., 2012
SCN9A	2:167099158-SNV	rs6746030	Congenital insensitivity to pain	Klein et al., 2013
SCN9A	2:167099158-SNV	rs6746030	Basal Pain Sensitivity	Duan et al., 2015
SCN9A	2:167145122-SNV	rs188798505	Altered pain perception	Reimann et al., 2010
DRD3	3:113890815-SNV	rs6280	Acute pain in sickle cell disease	Jhun et al., 2014
DRD3	3:113890815-SNV	rs6280	Higher prevalence of migraine	Hu et al., 2014
ADRB2	5:148206646-SNV	rs1042717	Musculoskeletal pain	Diatchenko et al., 2006
ADRB2	5:148206885-SNV	rs1800888	Migraine	Schurks et al., 2009
ESR1	6:152129077-SNV	rs2077647	Migraine	Schürks et al., 2010
ESR1	6:152129077-SNV	rs2077647	Musculoskeletal pain	Wise et al., 2009
OPRM1	6:154360797-SNV	rs1799971	Pain of various origins	Lötsch et al., 2009c
SOD2	6:160113872-SNV	rs4880	Migraine	Palmirotta et al., 2015
IL6	7:22771039-SNV	rs13306435	Low back pain	Eskola et al., 2010
OPRK1	8:54142157-SNV	rs702764	Neuropathic pain	Garassino et al., 2013
TLR4	9:120475302-SNV	rs4986790	Musculoskeletal pain	Gębura et al., 2017
TH	11:2188238-SNV	rs6357	Widespread Pain	Jhun et al., 2015
TH	11:2190951-SNV	rs6356	Migraine	Corominas et al., 2009
BDNF	11:27679916-SNV	rs6265	Widespread Pain	Ersig et al., 2017
DRD2	11:113283459-SNV	rs6277	Post-surgical pain	Kim et al., 2006a
DRD2	11:113283477-SNV	rs6275	Migraine	Onaya et al., 2013
P2RX7	12:121600253-SNV	rs208294	Cold pain sensitivity	Ide et al., 2014
P2RX7	12:121605355-SNV	rs7958311	Neuropathic pain	Ursu et al., 2014
HTR2A	13:47409034-SNV	rs6314	Migraine susceptibility	Yücel et al., 2016
TRPV1	17:3480447-SNV	rs8065080	Neuropathic pain	Doehring et al., 2011
KCNS1	20:43723627-SNV	rs734784	Neuropathic pain	Doehring et al., 2011
COMT	22:19950235-SNV	rs4633	Postoperative pain	Khalil et al., 2017
COMT	22:19950263-SNV	rs6267	Widespread Pain	Lin et al., 2017
COMT	22:19951271-SNV	rs4680	Altered pain perception	Wang et al., 2015
Other context				
CSF1	1:110466466-SNV	rs1058885	Periodontitis	Chen et al., 2014
CSF1	1:110466555-SNV	rs2229165	Carcinogenesis/breast cancer	Savas et al., 2006
NTRK1	1:156846233-SNV	rs6334	Nephropathy	Hahn et al., 2011
NTRK1	1:156848946-SNV	rs6339	Acute myeloid leukemia	Schweinhardt et al., 2008
SCN9A	2:167143050-SNV	rs41268673	Erythromelalgia	Klein et al., 2013
TRPM8	2:234854550-SNV	rs11562975	Hyperresponsiveness in bronchial asthma	Naumov et al., 2015
TRPM8	2:234905078-SNV	rs11563208	Anthropometric parameters	Potapova et al., 2014
DRD3	3:113890789-SNV	rs3732783	Phenotypic traits relevant to anorexia nervosa	Root et al., 2011
KIT	4:55593464-SNV	rs3822214	Cancer risk	Pelletier and Weidhaas, 2010
KIT	4:55602765-SNV	rs3733542	Glandular odontogenic cyst	Siqueira et al., 2017
HTR1A	5:63257483-SNV	rs1799921	Bipolar disorders	Goodyer et al., 2010
ADRB2	5:148206646-SNV	rs1042717	Cognitive dysfunction in opioid-treated patients with cancer	Kurita et al., 2016
DRD1	5:174868840-SNV	rs155417	Alcohol dependence	Hack et al., 2011
HLA-DQB1	6:32629920-SNV	rs41544112	Ulcerative colitis	Achkar et al., 2012
FKBP5	6:35544942-SNV	rs34866878	Clinical response in pediatric acute myeloid leukemia	Mitra et al., 2011

(Continued)

TABLE 2 | Continued

Gene	Variant	dbSNP[#] accession number	Known clinical association	Reference
CNR1	6:88853635-SNV	rs1049353	Bone mineral density	Woo et al., 2015
CNR1	6:88853635-SNV	rs1049353	Alcohol dependence	Marcos et al., 2012
CNR1	6:88853635-SNV	rs1049353	Nicotine dependence	Chen et al., 2008
CNR1	6:88853635-SNV	rs1049353	Obesity	Schleinitz et al., 2010
CNR1	6:88853635-SNV	rs1049353	Psychiatric disorders	Hillard et al., 2012
ESR1	6:152129077-SNV	rs2077647	Breast cancer susceptibility	Li et al., 2016
ESR1	6:152129077-SNV	rs2077647	Prostate cancer development	Jurečeková et al., 2015
ESR1	6:152129077-SNV	rs2077647	Osteoporosis	Sonoda et al., 2012
ESR1	6:152129308-SNV	rs746432	Mood disorders	Mill et al., 2008
ESR1	6:152201875-SNV	rs4986934	Endometrial cancer risk	Wedrén et al., 2008
OPRM1	6:154360508-SNV	rs6912029	Irritable bowel syndrome	Camilleri et al., 2014
OPRM1	6:154360797-SNV	rs1799971	Schizophrenia	Serý et al., 2010
OPRM1	6:154414573-SNV	rs562859	Depressive disorder	Garriock et al., 2010
OPRM1	6:154414563-SNV	rs675026	Treatment response for opiate dependence	Al-Eitan et al., 2012
SOD2	6:160113872-SNV	rs4880	Development of type 2 diabetes mellitus	Li et al., 2015
SOD2	6:160113872-SNV	rs4880	Breast cancer susceptibility	Rodrigues et al., 2014
SOD2	6:160113872-SNV	rs4880	Asthma	Yucesoy et al., 2012
ADCY1	7:45703971-SNV	rs1042009	Bipolar disorder	Shi J. et al., 2008
RELN	7:103124207-SNV	rs1062831	Attention deficit hyperactivity disorder	Kwon et al., 2016
RELN	7:103251161-SNV	rs362691	Childhood epilepsy	Dutta et al., 2011
OPRK1	8:54142154-SNV	rs16918875	Susceptibility to addiction	Kumar et al., 2012
TRPV1	8:72948588-SNV	rs13280644	Perception olfactory stimuli	Schütz et al., 2014
TLR4	9:120475602-SNV	rs4986791	Breast cancer susceptibility	Milne et al., 2014
GRIN1	9:140051238-SNV	rs6293	Schizophrenia	Georgi et al., 2007
RET	10:43610119-SNV	rs1799939	Hirschsprung's disease	Vaclavikova et al., 2014
RET	10:43615094-SNV	rs1800862	Medullary thyroid carcinoma	Ceolin et al., 2012
GFRA1	10:117884950-SNV	rs2245020	Age-related macular degeneration	Schmidt et al., 2006
DRD4	11:637537-Del	rs587776842	Acousticous neurinoma	Nöthen et al., 1994
BDNF	11:27720937-SNV	rs66866077	Irritable bowel syndrome-diarrhea	Camilleri et al., 2014
DRD2	11:113283484-SNV	rs1801028	Neurologic disorders	Doehring et al., 2009
GRIN2B	12:13717508-SNV	rs1806201	Alzheimer's disease	Andreoli et al., 2014
TRPV4	12:110252547-SNV	rs3742030	Hyponatremia	Tian et al., 2009
P2RX7	12:121592689-SNV	rs17525809	Multiple sclerosis	Oyanguren-Desez et al., 2011
HTR2A	13:47466622-SNV	rs6305	Susceptibility to substance abuse	Herman and Balogh, 2012
LTB4R	14:24785092-SNV	rs34645221	Asthma susceptibility	Tulah et al., 2012
GABRA5	15:27182357-SNV	rs140682	Autism-spectrum disorders	Hogart et al., 2007
GRIN2A	16:9943666-SNV	rs2229193	Hyperactivity disorder	Kim et al., 2017
DLG4	17:7099811-SNV	rs17203281	Schizophrenia	Tsai et al., 2007
SLC6A4	17:28530193-SNV	rs6352	Autism-spectrum disorders	Prasad et al., 2009
NF1	17:29553485-SNV	rs2285892	Neurofibromatosis	Maertens et al., 2007
HCN2	19:607984-SNV	rs3752158	Risk of depression	McIntosh et al., 2012
PRKCG	19:54394965-SNV	rs3745396	Osteosarcoma susceptibility	Lu et al., 2015
PRNP	20:4680251-SNV	rs1799990	Creutzfeldt-Jakob disease	Mead et al., 2009
HRH3	20:60791422-SNV	rs3787430	Risk of chronic heart failure	He et al., 2016
S100B	21:48022230-SNV	rs1051169	Schizophrenia	Liu et al., 2005

The selection is restricted to one or two publications per variant, and it focuses on a pain context corresponding to the main aim of the present NGS gene panel; however, functional variants highlighted in another clinical context are additionally provided in the lower part of the table. [#]*Database of Single Nucleotide Polymorphisms (dbSNP). Bethesda (MD, United States): National Center for Biotechnology Information, National Library of Medicine. Available from: http://www.ncbi.nlm.nih.gov/SNP/ (Sherry et al., 2001).*

Following elimination of nucleotides agreeing with the standard human genome sequence GRCh37 g1k (dated February 2009), the result of the NGS consisted of a vector of nucleotide information about the $d = 77$ genes for each individual DNA sample (**Figure 2**). This vector had a length equaling the set union of the number of chromosomal positions in which a non-reference nucleotide had been found in any probe of the actual cohort. Specifically, a total of 3,185 genetic variants was found, of

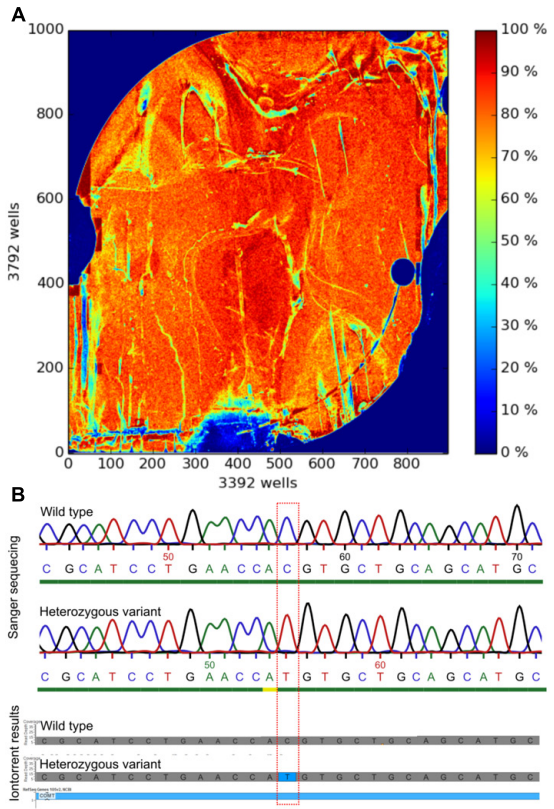


FIGURE 1 | Assay establishment and validation. **(A)** Pseudo-color image of the Ion 318^{TM} v2 Chip plate showing percent loading across the physical surface. This sequencing run had a 76% loading, which ensures a high Ion Sphere Particles (ISP) density. Every 318 chip contains 11 million wells and the color scale on the right side conduces as a loading indicator. Deep red coloration stays for a 100% loading, which means that every well in this area contains an ISP (templated and non-templated) whereas deep blue coloration implies that the wells in this area are empty. **(B)** Alignment of a segment of the ion torrent sequence of the *COMT* gene as a Golden Helix Genome Browse® readout versus the same sequence according to an externally predicted Sanger electropherogram. Highlighted is the *COMT* variant rs4633 (COMT c.186C>T → p.His62 =) as a heterozygous mutation and a non-mutated wild type. The SNP is part of the functional *COMT* haplotype comprising rs4633, rs4818 and rs4680, which showed >11-fold difference in expressed enzyme activity and was reported to be associated with different phenotypes of pain sensitivity (Diatchenko et al., 2005).

which 659 were located in coding parts of the genes, 1,241 were located in introns and 1,285 in the 3′-UTR, 5′-UTR, upstream or downstream regions. The coding variants for which a clinical or phenotypic association have been reported are listed in **Table 2** together with an example of each variant. Most of the observed variants were single nucleotide polymorphisms ($d = 571$) whereas mixed polymorphisms ($d = 26$), nucleotide insertions ($d = 18$) or nucleotide deletions ($d = 44$) were more rarely found.

DISCUSSION

In this report, development and validation of a novel Ampliseq^{TM} NGS assay for the coding regions and boundary parts of $d = 77$ genes qualifying as candidate modulators of persisting pain is described. The NGS assay produced nucleotide sequences that corresponded, with respect to the selected validation probes, to the results of classical Sanger sequencing. However, the NGS assay substantially reduced the laboratory effort to obtain the genetic information and provides the perquisites to be used in high throughput environments. In particular, the presented NGS assay is convenient for small up to large-scale setups. As mentioned in the methods section, a limitation of the present results applies to the identified genetic variants as only samples from Caucasian women were included. By contrast, the validity of gene selection and assay establishment is unlikely to be reduced by this selection chosen to remain within the financial limits of the present project.

Specifically, as observed previously (Kringel et al., 2017), the comprehensive genetic information and the high throughput are reflected in the assay costs. Specifically, sequencing of the 77 genes in 72 DNA samples required approximately € 18,000 for the AmpliSeq^{TM} custom panel, € 5,500 for library preparation, € 700 for template preparation and € 700 for sequencing. Ten 318 sequencing chips cost around € 7,000 and in addition and basic consumables and laboratory supplies issued approximately € 800. With 7–8 barcoded samples loaded on ten chips, the expense to analyses the gene sequence for a single patient were around € 325. While NGS costs are likely to decrease in the near future (Lohmann and Klein, 2014), present assay establishment was therefore applied in DNA samples planned for future genotype versus phenotype association analysis, which required using DNA from patients of a pain-relevant cohort instead from a true random sample of healthy subjects.

As a result of the present assay development, a set of $d = 77$ genes was chosen as potentially relevant to persisting pain. The chosen set of genes differs from alternative proposals aiming at similar phenotypes (Mogil, 2012; Zorina-Lichtenwalter et al., 2016). However, when analyzing these alternatives for mutual agreement, only limited overlap could be observed (**Figure 3**). This emphasizes that the genetic architecture of persisting pain is incompletely understood, and several independent lines of research can be pursued. Of note, the present set showed the largest agreement with a set of $d = 539$ genes identified empirically as relevant to pain and listed in the PainGenes database (Lacroix-Fralish et al., 2007)[6] or recognized as causing human hereditary diseases associated with extreme pain phenotypes (Lötsch et al., 2013; Ultsch et al., 2016). Combining all proposals into a large panel was not an option due to the technical limitations of the IonTorrent restricting the panel size to 500 kb (pipeline version 5.6.2); therefore, further genes would need to be addressed in separate panels.

In the present study sample, selected with a certain bias by using, as explained above for cost saving, clinical samples from only women and only Caucasians, a total of 659 genetic coding variants were found. Regardless of the sample preselection, 105 clinical associations (**Table 2**) could be queried for the observed variants from openly obtainable data sources comprising (i) the

[6]http://www.jbldesign.com/jmogil/enter.html

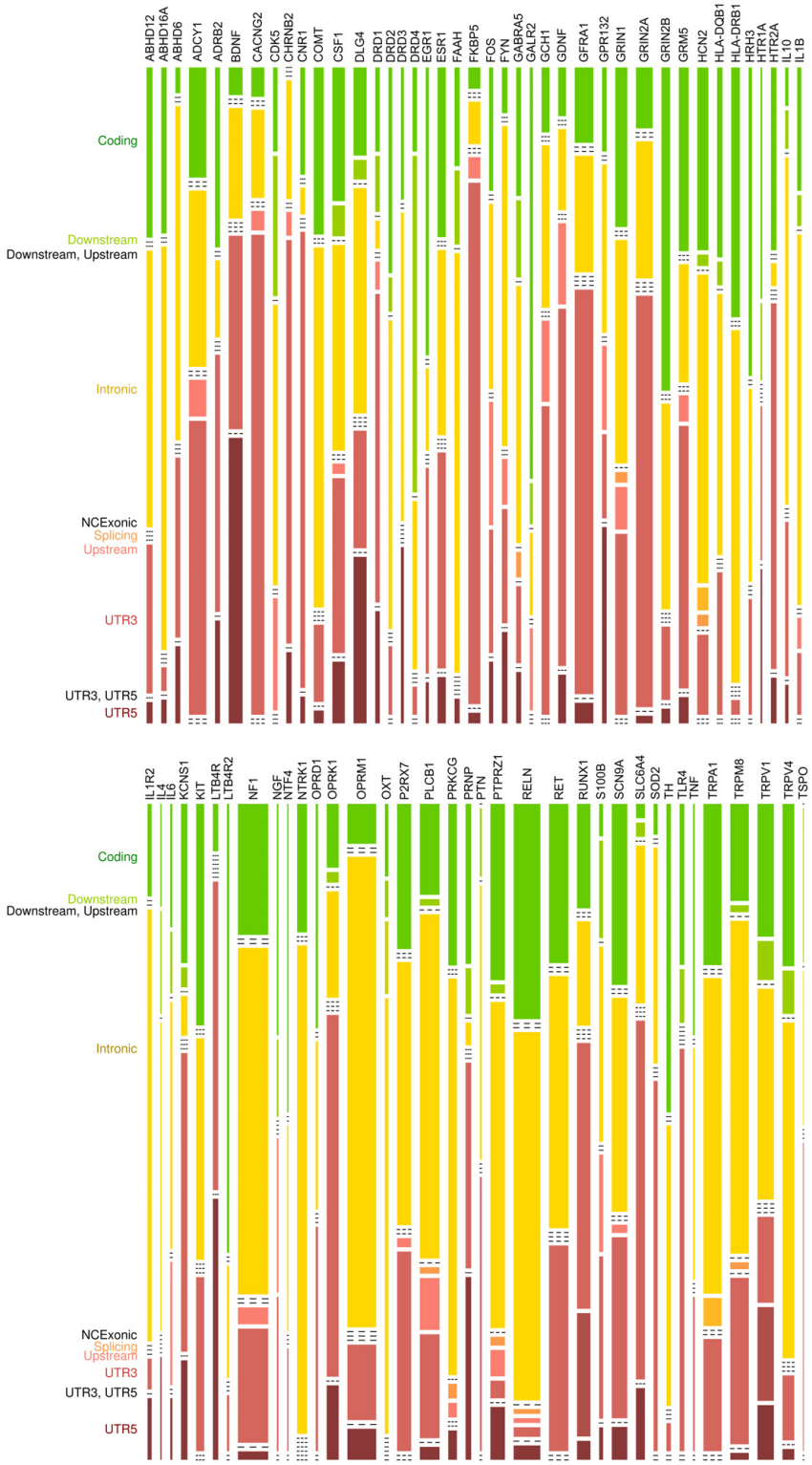


FIGURE 2 | Mosaic plot representing a contingency table of the types of genetic variants detected by means of the present AmpliSeq^{TM} panel versus the genes included in the assay. The vertical size of the cells is proportional to the number of variants of a particular type; the horizontal size of the cells is proportional to the number of variants found in the respective gene. The location of the variants is indicated at the left of the mosaic plot in letters colored similarly to the respective bars in the mosaic plot. Variants were not found at all possible locations of each gene, which causes the reduction of several bars to dashed lines drawn as placeholders and indicating that at the particular location no variant has been found in the respective gene. The figure has been created using the R software package (version 3.4.2 for Linux; http://CRAN.R-project.org/, R Development Core Team, 2008). UTR: untranslated region. NCExonic: Non-coding exonic.

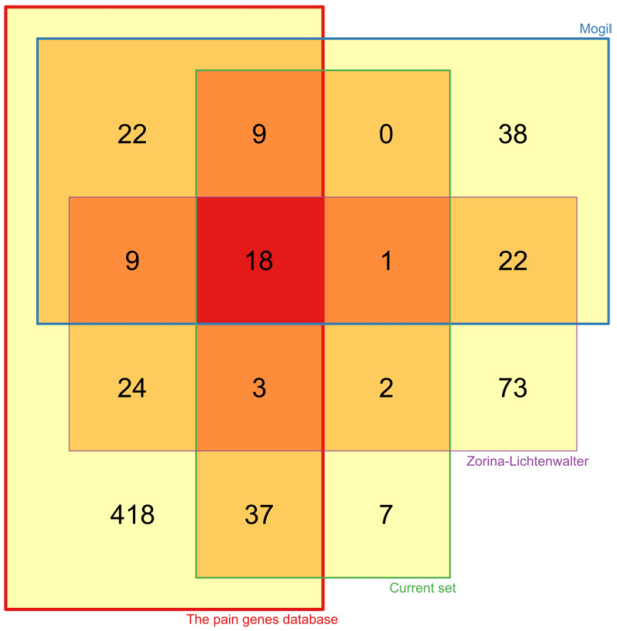


FIGURE 3 | Venn diagram (Venn, 1880) visualizing the intersections between the presently proposed set of human genes involved in modulating the risk or the clinical course of persisting pain ("Current set," green frame), and two alternative proposals ["Mogil" (Mogil, 2012), blue frame and "Zorina-Lichtenwalter" (Zorina-Lichtenwalter et al., 2016), violet frame]. In addition, a set of $d = 539$ genes identified empirically as relevant to pain and either listed in the PainGenes database (http://www.jbldesign.com/jmogil/enter.html, Lacroix-Fralish et al., 2007) or added because recognized as causing human hereditary diseases associated with extreme pain phenotypes, found to be regulated in chronic pain in at least three studies including human association studies, or being targets of novel analgesics. The number of shared genes between data sets is numerically shown in the respective intersections of the Venn diagram. The figure has been created using the R software package (version 3.4.2 for Linux; http://CRAN.R-project.org/, R Development Core Team, 2008) with the particular package "Vennerable" (Swinton J., https://r-forge.r-project.org/R/?group_id=474).

Online Mendelian Inheritance in Man (OMIM®) database[7], (ii) the NCBI gene index database[8], the GeneCards database[9] [27] and the "1000 Genomes Browser"[10] (all accessed in December 2017). The observation of functional variants in the present cohort preselected for the absence of pain persistence is plausible as (i) variants can exert protective effects against chronic pain and (ii) most genetic variants identified so far exert only small effects on pain and the individual result of their functional modulations depends on their combined effects or from the sum of positive and negative effects on pain perception (Lötsch et al., 2009a).

The selection of genes (**Table 1**) relied on empirical evidence of their involvement in pain. For subset #1 ($d = 34$), this had been shown for 33 genes in the original paper (Ultsch et al., 2016). As the hypothesis that persisting pain displays systemic features of learning and of neuronal plasticity (Mansour et al., 2014) could be substantiated at a computational functional genomics level, the further gene (*PTPRZ1*, protein tyrosine phosphatase Z 1) can also be regarded as supported by prior knowledge to be included in the present set. The subset comprised, for example, genes associated with the mesolimbic dopaminergic system, i.e., *DRD1, DRD2, DRD3*, which code for dopamine receptors, and *TH*, which is the coding gene for the tyrosine hydroxylase, a metabolic restricting enzyme in dopaminergic pathways, which have been implicated in promoting chronic back pain (Hagelberg et al., 2003, 2004; Jaaskelainen et al., 2014; Martikainen et al., 2015). Further 14 genes were involved in the circadian rhythm recognized as a modulatory factor in various pain conditions such as arthritis (Haus et al., 2012; Gibbs and Ray, 2013) and neuropathic pain (Gilron and Ghasemlou, 2014). The subset further included three NMDA receptor genes (*GRIN1, GRIN2A,* and *GRIN2B*) known to be major players in a number of essential physiological functions including neuroplasticity (Coyle and Tsai, 2004). In addition, metabotropic glutamate receptors (mGluR) have been implemented in several chronic pain conditions. One subtype, mGluR5, coded by *GRM5*, is of particular interest in the context of pain conditions as recent studies showed a pro-nociceptive role of mGluR5 in models of chronic pain (Walker et al., 2001; Crock et al., 2012). Furthermore, genes associated with histaminergic signaling such as *HRH3* have been implicated in pain transmission (Hough and Rice, 2011) and analgesia (Huang et al., 2007).

The second subset of genes relied on a new PubMed search rather than on a previously published and hypothesis-based selection of candidate genes. A computational functional genomics analysis of this subset (details not shown) suggested its involvement in (i) immune processes and (ii) nitric oxide signaling. The genes annotated to the GO term "immune system process" included interleukin (IL1B, IL4, IL6, IL10) (Dinarello, 1994; Choi and Reiser, 1998; Mocellin et al., 2004; Nemeth et al., 2004) and histocompatibility complex related (HLA-B) genes (Dupont and Ceppellini, 1989), which have been shown to be involved in immunological mechanisms of pain (Sato et al., 2002; de Rooij et al., 2009). This is also supported by published evidence for the further genes in this list, such as, *TNF* (Vassalli, 1992; Franchimont et al., 1999), *GCH1* (Schott et al., 1993) and *P2RX7* (Chen and Brosnan, 2006). The second major process group emerging from the functional genomics analysis of the key evidence for genetic modulation of clinical chronic pain was nitric oxide signaling, in particular metabolic processes, summarized in this context under the GO term "reactive oxygen species metabolic process" which includes the genes *IL6* (Deakin et al., 1995), *TNF* (Deakin et al., 1995; Katusic et al., 1998), *ESR1* (Clapauch et al., 2014), *IL10* (Cattaruzza et al., 2003), *GCH1* (Katusic et al., 1998; Zhang et al., 2007), *IL1B* (Katusic et al., 1998), *IL4* (Coccia et al., 2000), *P2RX7* (Gendron et al., 2003), *SOD2* (Fridovich, 1978). Furthermore, catecholamines including noradrenaline, adrenaline and dopamine have multiple functions in the brain and spinal cord including pain perception and processing (D'Mello and Dickenson, 2008). Catechol-*O*-methyltransferase, encoded by the *COMT* gene, is one of several enzymes that degrade dopamine, noradrenaline and adrenaline

[7]http://www.ncbi.nlm.nih.gov/omim
[8]http://www.ncbi.nlm.nih.gov/gene
[9]http://www.genecards.org
[10]https://www.ncbi.nlm.nih.gov/variation/tools/1000genomes

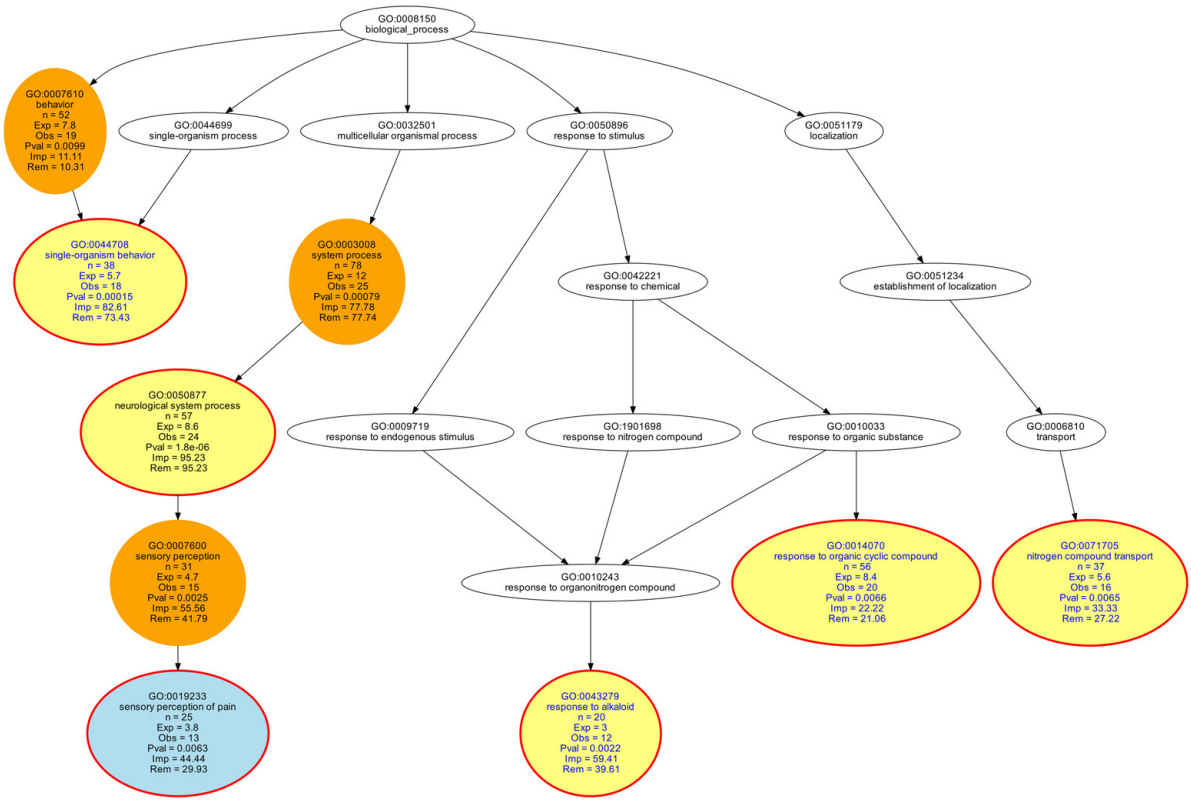


FIGURE 4 | Top–down representation of the annotations (GO terms) representing the taxonomy of the functional differences between the set of $d = 77$ genes included in the proposed NGS panel of persisting pain and two alternative proposals of genes modulating persisting pain in humans (Mogil, 2012; Zorina-Lichtenwalter et al., 2016). The figure represents the results of an over-representation analysis of the present set of $d = 77$ genes against the reference comprising the set intersection of the alternative gene lists. A p-value threshold of 0.01 and Bonferroni α-correction were applied. Significant terms are shown as colored circles with the number of member genes, the number of expected genes by change and the significance of the deviation of the observed from the expected number of genes indicated (yellow = headline, red = significant term, blue = significant term located as a leave at the end of a taxonomy in the polyhierarchy). The graphical representation follows the standard of the GO knowledgebase, where GO terms are related to each other by "is-a," "part-of," and "regulates" relationships forming a polyhierarchy organized in a directed acyclic graph (DAG, Thulasiraman and Swamy, 1992). The figure has been created using our R library "dbtORA" (https://github.com/IME-TMP-FFM/dbtORA, Lippmann et al., 2018) on the R software package (version 3.4.2 for Linux; http://CRAN.R-project.org/, R Development Core Team, 2008) and the freely available graph visualization software GraphViz (http://www.graphviz.org, Gansner and North, 2000).

and has become one of the most frequently addressed genes in pain research (Nackley et al., 2006).

Finally, subset #3 ($d = 30$) consists of genes repeatedly shown to play a role in the genetic modulation of persisting pain in humans or, by contrast, included a few novel items only recently published in the context of pain. This included members of the transient receptor potential (TRP) family (*TRPA1, TRPM8, TRPV4*) that are expressed at nociceptors and which are well established players in the perception of pain via their excitation by chemical, thermal or mechanical stimuli (Clapham, 2003). This similarly applies to the opioidergic system represented by the inclusion of the genes coding for the major opioid receptors (*OPRM1, OPRK1 OPRD1*), which have been associated with variations in pain or opioid response in various settings (Lötsch and Geisslinger, 2005). The most important of this group, the μ-opioid receptor encoded by the *OPRM1* gene, carriers several variants of which the 118 A>G (rs1799971) has been studied most extensively since the early description of its association with a functional phenotype in humans (Lötsch et al., 2002).

Almost half of the present sets of genes were chosen based on a computational functional genomics analysis that attributed persisting pain to GO processes of "learning or memory" and "nervous system development" (Ultsch et al., 2016) as likely to reflect systemic features of persisting pain. This implied a functional bias and therefore, the present set of $d = 77$ genes (**Figure 4**) was analyzed whether this bias prevailed when comparing it with the alternative sets of human genes proposed to modulate persisting pain (Mogil, 2012; Zorina-Lichtenwalter et al., 2016). As applied previously (Lippmann et al., 2018), the biological roles of the set of $d = 77$ genes were queried from the Gene Ontology knowledgebase (GO)[11] (Ashburner et al., 2000) where the knowledge about the biological processes, the molecular functions and the cellular components of genes is formulated using a controlled and clearly defined vocabulary of GO terms. Particular biological roles of the set of $d = 77$ genes, among all human genes, were analyzed by

[11] http://www.geneontology.org/

TABLE 3 | Current targeting of the genes included in the proposed NGS panel of persisting pain by novel drugs that are currently under active clinical development and include analgesia as the main clinical target or at least as one of the intended clinical indication.

Gene	Status	Drug	Action	Company
ABHD12	–	–	–	–
ABHD16A	–	–	–	–
ABHD6	Preclinical	Benzylpiperidin methanone	Acylamino-Acid-Releasing Enzyme	Scripps Research Institute
ADCY1	Under Active Development	NB-001	Adenylate Cyclase Inhibitors	Forever Cheer International
ADRB2	Phase II/III	Gencaro	Signal Transduction Modulators	ARCA
BDNF	Phase I	CXB-909	Nerve Growth Factor (NGF) Enhancers	Krenitsky
CACNG2	Preclinical	Hanfangchin	Calcium Channel Blockers	Millenia Hope Kaken
CDK5	Biological Testing	Litvinolin	CDK5/p25 Inhibitors	Hong Kong University
CHRNB2	Biological Testing	Epiboxidine	Nicotinic alpha4beta2 Receptor Agonists	Pfizer
CNR1	Registered	Epidiolex	Cannabinoid Receptor Agonists	InSys Therapeutics
COMT	Clinical	Nitecapone	Catechol-*O*-Methyl Transferase (COMT) Inhibitors	Orion
CSF1	–	–	–	–
DLG4	Preclinical	AB-125	Protein Inhibitors	Lundbeck University of Copenhagen
DRD1	Phase II/III	Ecopipam	Dopamine D1 Receptor (DRD1) Antagonists	Merck & Co.
DRD2	Phase II/III	Sarizotan hydrochloride	Dopamine D2 Receptor (DRD2) Antagonists	Newron
DRD3	Phase II	Brilaroxazine	D3 Receptor (DRD3) Agonists	Reviva Pharmaceuticals
DRD4	Biological Testing	Mesulergine hydrochloride	Dopamine Receptor Agonists	Novartis
EGR1	Phase II	Brivoligide	EGR1 Expression Inhibitors	Adynxx
ESR1	Phase II	Zindoxifene	Selective Estrogen Receptor Modulators	Evonik
FAAH	Phase I/II	Minerval	Fatty Acid Amide Hydrolase (FAAH) Inhibitors	Scripps Research Institute
FKBP5	Phase II	Barusiban	Oxytocin Receptor Antagonist	Ferring
FOS	Registered	Macrilen	FOS Expression Enhancers	Strongbridge Biopharma
FYN	Phase II	Bafetinib	Fyn Kinase Inhibitors	Nippon Shinyaku
GABRA5	Phase III	Ganaxolone	GABA(A) Receptor Modulators	Marinus Pharmaceuticals
GALR2	Preclinical	NAX-810-2	GAL2 Receptor Ligands	NeuroAdjuvants
GCH1	–	–	–	–
GDNF	Phase II	Edonerpic maleate	Signal Transduction Modulators	Toyama
GFRA1	–	–	–	–
GPR132	–	–	–	–
GRIN1	Phase II	Dimiracetam	Signal Transduction Modulators	Metys Pharmaceuticals
GRIN2A	Phase I	Dexanabinol	NMDA Receptor Antagonists	e-Therapeutics Pharmos
GRIN2B	Phase I	Gacyclidine	NMDA Receptor Antagonists	INSERM
GRM5	Phase II	Mavoglurant	Signal Transduction Modulators	Novartis
HCN2	Clinical	Ivabradine	Adrenoceptor Antagonists	Servier
HLA-DQB1	–	–	–	–
HLA-DRB1	–	–	–	–
HRH3	Phase I	Immethridine	Histalean	Abbott
HTR1A	Phase II	Eltoprazine hydrochloride	5-HT1A Receptor Agonists	Elto Pharma
HTR2A	Phase II	Midomafetamine	5-HT2 Receptor Agonists	Assoc
IL10	Phase II	BT-063	Signal Transduction Modulators Anti-IL-10	Biotest AG
IL1B	Phase III	Resunab	IL-1beta Inhibitors	Corbus
IL1R2	–	–	–	–
IL4	–	–	–	–
IL6	Preclinical	Azintrel	Signal Transduction Modulators Anti-IL-6	Jazz Pharmaceuticals
KCNS1	Preclinical	Crotamine	Voltage-Gated K(V) Channel Blockers	Celtic Biotech
KIT	Phase II	Vatalanib succinate	KIT (C-KIT) Inhibitors	Novartis
LTB4R	Phase II	Coversin	Signal Transduction Modulators	Akari Therapeutics
LTB4R2	Phase II	Coversin	Signal Transduction Modulators	Akari Therapeutics
NF1	–	–	–	–
NGF	Phase III	Tanezumab	Anti-Nerve Growth Factor (NGF)	Pfizer
NTF4	–	–	–	–

(Continued)

TABLE 3 | Continued

Gene	Status	Drug	Action	Company
NTRK1	Phase II	Danusertib	NTRK1 Inhibitors	Pfizer
OPRD1	Preclinical	Metenkephalin	Delta-Opioid Receptor Agonists	TNI Pharmaceuticals
OPRK1	Phase III	Morphine glucuronide	Opioid Receptor Agonists	PAION
OPRM1	Registered	Naltrexone	mu-Opioid Receptor Antagonists	Pfizer
OXT	Phase II	Barusiban	Oxytocin Receptor Antagonist	Ferring
P2RX7	Preclinical	BIL-06v	Anti-P2RX7	Biosceptre International
PLCB1	Biological Testing	Vinaxanthone	Signal Transduction Modulators	Roche
PRKCG	Phase III	Rydapt	Protein Kinase C (PKC) Inhibitors	Yeda
PRNP	–	–	–	–
PTN	–	–	–	–
PTPRZ1	–	–	–	–
RELN	Preclinical	IAIPs	Serine Protease Inhibitors	ProThera Biologics
RET	Phase II	Danusertib	Ret (RET) Inhibitors	Pfizer
RUNX1	–	–	–	–
S100B	–	–	–	–
SCN9A	Phase III	Priralfinamide	Voltage-Gated Sodium Channel Blockers	Newron
SLC6A4	Phase II	Litoxetine	Signal Transduction Modulators	Sanofi
SOD2	Phase II	Avasopasem manganese	Superoxide Dismutase (SOD) Mimetics	MetaPhore
TH	–	–	–	–
TLR4	Phase II	Eritoran tetrasodium	Toll-Like Receptor 4 (TLR4) Antagonists	Eisai
TNF	Phase III	Givinostat hydrochloride	TNF-alpha Release Inhibitors	Italfarmaco
TRPA1	Phase II	Cannabidivarin	TRPA1 Agonists	GW Pharmaceuticals
TRPM8	Phase II	Cannabidivarin	TRPM8 Antagonists	GW Pharmaceuticals
TRPV1	Phase I/II	Resiniferatoxin	TRPV1 (Vanilloid VR1 Receptor) Agonists	Icos
TRPV4	Phase II	GSK-2798745	TRPV4 Antagonists	GlaxoSmithKline
TSPO	Clinical	[11C]CB-184	Translocator Protein (TSPO) Ligands	Tokyo Metrop Geriatr Hosp Inst Gerontol

The information was queried from the Thomson Reuters Integrity database at https://integrity.thomson-pharma.com on July 11, 2018.

means of over-representation analysis (ORA). This compared the occurrence of the particular GO terms associated with the present set of genes with their expected occurrence by chance (Backes et al., 2007). In contrast to enrichment analysis, any quantitative criteria such as gene expression values are disregarded (Backes et al., 2007). The analyses were performed using our R library "dbtORA" (Lippmann et al., 2018)[12] on the R software environment (version 3.4.2 for Linux; R Development Core Team, 2008)[13].

Surprisingly, the results of this analysis indicated that the functional bias of the present gene set toward "learning or memory" (GO:0007611) and "nervous system development" (GO:0007399) was not maintained against the alternative gene sets. Instead, a few more general GO terms such as "behavior" ("single organism behavior," GO:0044708), or "response to organic cyclic compound" (GO:0014070) and response to alkaloid (GO:0043279), which could be identified as morphine and cocaine when repeating the analysis with a less conservative α-correction (further details not shown), were overrepresented, as well as the pain specific term "sensory perception of pain" (GO:0019233). A possible explanation that the selection bias of the present gene set was not maintained when comparing it with alternative proposals is that the two biological processes, "learning or memory" and "nervous system development," reflect indeed an important biological function of persisting pain and even when choosing candidate genes without having these processes in mind as for the alternative gene sets, they are nevertheless included. This may be regarded as support for the present gene set as suitable candidates for future association studies with persisting pain phenotypes.

Although the present gene set has been assembled with a focus of a relevance to pain, many of its members have pharmacological implications. Specifically, 58 of the 77 genes (75%) have been chosen as targets of analgesics, approved or under current clinical development (**Table 3**). Moreover, several of the genes in the present NGS panel have been implicated in pharmacogenetic modulations of drug effects (**Table 4**). Possibly the most widely studied gene in analgesic research is *OPRM1* because coding for the primary target of opioids (Peiro et al., 2016). Several polymorphisms have been described in *OPRM1*, among which the best characterized may be rs1799971 (*OPRM1* 118A>G) that leads to an asparagine to aspartate substitution at the extracellular terminal of the receptor protein (Bond et al., 1998). May studies have addressed this variant (for reviews, see Walter et al., 2013; Somogyi et al., 2015).

[12]https://github.com/IME-TMP-FFM/dbtORA
[13]http://CRAN.R-project.org/

TABLE 4 | Summary of variants in genes included in the proposed NGS panel of persisting pain, that have been implicated in a pharmacogenetic context to modulate the effects of drugs administered for the treatment of pain or as disease modifying therapeutics in painful disease.

Modulated process	Gene	Variant	Affected drug	Findings	Reference
G protein coupled signaling	*COMT*	rs4680 (Val158Met)	Morphine	Carriers of val/val and val/met genotype required higher morphine dose compared to carriers of met/met genotype	Reyes-Gibby et al., 2007
	DRD2	rs6275	Heroine	Polymorphism is associated with decreased likelihood of headache disorders	Cargnin et al., 2014
	DRD4	rs1800955	Heroine	Polymorphism had lower pain threshold versus CC/CT controls	Ho et al., 2008
	OPRM1	rs1799971 (A118G)	Various opioids	Tendency toward increased pain in dose-dependent manner with the μ-opioid receptor variant 118G	Lötsch et al., 2009c
	OPRK1	rs1051660	Morphine	Patients with the polymorphism and cancer-related pain may require a reduced dose escalation of morphine	Chatti et al., 2017
Neurotransmitters	*BDNF*	rs6265	Various opioids	Polymorphism is associated with decreased likelihood of headache disorders	Cargnin et al., 2014
	HTR2A	rs12584920	Various opioids	Increased likelihood of having chronic widespread pain	Nicholl et al., 2011
Ion Channels	*TRPV1*	7 intronic SNPs	Capsaicin	TRPV1 polymorphisms had only 50% of the mRNA and protein expression levels of normally sensing subjects	Park et al., 2007
Proinflammatory Cytokines	*IL6*	rs1800795	Etanercept	Polymorphism is associated with increased response to adalimumab, etanercept or infliximab in people with painful Arthritis	Davila-Fajardo et al., 2014
Other	*ESR1*	rs2234693	Leflunomide	Polymorphism is associated with increased response to leflunomide in women with painful Arthritis	Dziedziejko et al., 2011
	FAAH	rs2295632	Various opioids	Polymorphism is associated with increased risk of Respiratory Insufficiency	Biesiada et al., 2014
	TLR4	rs4986790	Methotrexate	Polymorphism associated with increased risk of adverse drug events when treated with folic acid and methotrexate in people with Arthritis	Kooloos et al., 2010
	TNF	rs361525	Infliximab	Polymorphism is associated with increased response to infliximab in people with painful Arthritis	Maxwell et al., 2008

The information was derived by literature search and by querying the Pharmacogenetics Research Network/Knowledge base at http://www.pharmgkb.org (accessed in July 2018). Only key or example references are given.

Summarizing its effects, the variant is associated with decreased receptor expression and signaling efficiency (Oertel et al., 2012) which leads to reproducibly reduced pharmacodynamic effects in human experimental settings while the effect size seems insufficient to be a major factor of opioid response in clinical settings, despite several reports of modulations of opioid demands or side effects. For example, subjects carrying the 118A>G variant were found to have a reduced response to morphine treatment (Hwang et al., 2014), reduced analgesic response to alfentanil (Oertel et al., 2006) and demanded higher doses of morphine for pain relief (Klepstad et al., 2004; Hwang et al., 2014). However, the importance of this variant seems to be comparatively high in patients with an Asian ethnic background, which might be related to the higher allelic frequency as compared to other ethnicities. *COMT* is a key modulator of dopaminergic neurotransmission and in the signaling response to opioids The Val158Met polymorphism (rs4680) causes an amino acid substitution in the enzyme, which reduced the enzyme active to a forth (Peiro et al., 2016). Carriers of the homozygous Met/Met variant had lower morphine requirements than those with a the wild type *COMT* (Rakvag et al., 2005). Furthermore, a modulation of the effects of *TRPV1* targeting analgesics is supported by observations that intronic *TRPV1* variants were associated with insensitivity to capsaicin (Park et al., 2007) while the coding *TRPV1* variant rs8065080 was associated with altered responses to experimentally induced pain

(Kim et al., 2004). Moreover, gain-of-function mutations in *TRPV1* have been associated with increased pain sensitivity (Boukalova et al., 2014), for which *TRPV1* antagonists would enable a specific pharmacogenetics-based personalized cure.

CONCLUSION

The breakthrough in mapping the whole human genome (Lander et al., 2001; Venter et al., 2001) along with genome wide association studies (GWAS) has led to rapid advances in the knowledge of the genetic bases of human diseases (Wellcome Trust Case Control and Consortium, 2007). Genetic research in pain medicine has directed to the recognition of genes in which variants influence pain behavior, post-operative drug requirements, and the temporal developments of pain toward persistence (James, 2013). While many candidate gene association studies have identified multiple genes relevant for pain phenotypes (Fillingim et al., 2008), pain related genetic studies have so far been owned by investigations of a limited number of genes. Roughly ten genes or gene complexes account for over half of the extant findings and several of these candidate gene associations have held up in replication (Mogil, 2012). The selection of variants has been limited and they have been addressed in most studies repeatedly, leading to the perception that genetic research in pain produces often unsatisfactory results (Mogil, 2009). However, this may soon change with the arise of new technologies. In this manuscript, we present a validated NGS assay for a set of 77 genes supported by empirical evidence and computational functional genomics analyses as relevant factors modulating the risk for persisting pain or its clinical picture.

AUTHOR CONTRIBUTIONS

JL, DK, and EK conceived and designed the experiments. DK performed the experiments. JL and DK analyzed the data and wrote the paper. CL provided methodological expertise and bioinformatical tools. DK and JL interpreted the results. EK and MK provided DNA samples.

FUNDING

This work has been funded by the European Union Seventh Framework Programme (FP7/2007 – 2013) under grant agreement no. 602919 ("GLORIA", EK and JL) and the LandesOffensive zur Entwicklung Wissenschaftlich-ökonomischer Exzellenz (LOEWE), LOEWE-Zentrum für Translationale Medizin und Pharmakologie (JL). These public funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fphar.2018.01008/full#supplementary-material

REFERENCES

Abbadie, C., Besson, J. M., and Calvino, B. (1994). c-Fos expression in the spinal cord and pain-related symptoms induced by chronic arthritis in the rat are prevented by pretreatment with *Freund adjuvant*. *J. Neurosci.* 14, 5865–5871. doi: 10.1523/JNEUROSCI.14-10-05865.1994

Achkar, J.-P., Klei, L., De Bakker, P. I. W., Bellone, G., Rebert, N., Scott, R., et al. (2012). Amino acid position 11 of HLA-DRβ1 is a major determinant of chromosome 6p association with ulcerative colitis. *Genes Immun.* 13, 245–252. doi: 10.1038/gene.2011.79

Al-Eitan, L. N., Jaradat, S. A., Su, S. Y., Tay, G. K., and Hulse, G. K. (2012). Mu opioid receptor (OPRM1) as a predictor of treatment outcome in opiate-dependent individuals of Arab descent. *Pharmgen. Pers. Med.* 5, 99–111. doi: 10.2147/PGPM.S33351

Alvarado, S., Tajerian, M., Millecamps, M., Suderman, M., Stone, L. S., and Szyf, M. (2013). Peripheral nerve injury is accompanied by chronic transcriptome-wide changes in the mouse prefrontal cortex. *Mol. Pain* 9:21. doi: 10.1186/1744-8069-9-21

Andreoli, V., De Marco, E. V., Trecroci, F., Cittadella, R., Di Palma, G., and Gambardella, A. (2014). Potential involvement of GRIN2B encoding the NMDA receptor subunit NR2B in the spectrum of Alzheimer's disease. *J. Neural Transm. (Vienna)* 121, 533–542.

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., et al. (2000). Gene ontology: tool for the unification of biology. *Gene Ontol. Consortium. Nat. Genet.* 25, 25–29. doi: 10.1038/75556

Backes, C., Keller, A., Kuentzer, J., Kneissl, B., Comtesse, N., Elnakady, Y. A., et al. (2007). GeneTrail–advanced gene set enrichment analysis. *Nucleic Acids Res.* 35, W186–W192. doi: 10.1093/nar/gkm323

Biesiada, J., Chidambaran, V., Wagner, M., Zhang, X., Martin, L. J., Meller, J., et al. (2014). Genetic risk signatures of opioid-induced respiratory depression following pediatric tonsillectomy. *Pharmacogenomics* 15, 1749–1762. doi: 10.2217/pgs.14.137

Bond, C., Laforge, K. S., Tian, M., Melia, D., Zhang, S., and Borg, L. (1998). Single-nucleotide polymorphism in the human mu opioid receptor gene alters beta-endorphin binding and activity: possible implications for opiate addiction. *Proc. Natl. Acad. Sci. U.S.A.* 95, 9608–9613. doi: 10.1073/pnas.95.16.9608

Boukalova, S., Touska, F., Marsakova, L., Hynkova, A., Sura, L., Chvojka, S., et al. (2014). Gain-of-function mutations in the transient receptor potential channels TRPV1 and TRPA1: how painful? *Physiol. Res.* 63(Suppl. 1), S205–S213.

Bourinet, E., Altier, C., Hildebrand, M. E., Trang, T., Salter, M. W., and Zamponi, G. W. (2014). Calcium-permeable ion channels in pain signaling. *Physiol. Rev.* 94, 81–140. doi: 10.1152/physrev.00023.2013

Bravo, L., Torres-Sanchez, S., Alba-Delgado, C., Mico, J. A., and Berrocoso, E. (2014). Pain exacerbates chronic mild stress-induced changes in noradrenergic transmission in rats. *Eur. Neuropsychopharmacol.* 24, 996–1003. doi: 10.1016/j.euroneuro.2014.01.011

Bravo-Hernández, M., Corleto, J. A., Barragán-Iglesias, P., González-Ramírez, R., Pineda-Farias, J. B., Felix, R., et al. (2016). The α5 subunit containing GABAA receptors contribute to chronic pain. *Pain* 157, 613–626. doi: 10.1097/j.pain.0000000000000410

Bruera, G., Pepe, F., Malapelle, U., Pisapia, P., Mas, A. D., Di Giacomo, D., et al. (2018). KRAS, NRAS and BRAF mutations detected by next generation sequencing, and differential clinical outcome in metastatic colorectal cancer (MCRC) patients treated with first line FIr-B/FOx adding bevacizumab (BEV) to triplet chemotherapy. *Oncotarget* 9, 26279–26290. doi: 10.18632/oncotarget.25180

Buchheit, T., Van De Ven, T., and Shaw, A. (2012). Epigenetics and the transition from acute to chronic pain. *Pain. Med.* 13, 1474–1490. doi: 10.1111/j.1526-4637.2012.01488.x

Buskila, D., Cohen, H., Neumann, L., and Ebstein, R. P. (2004). An association between fibromyalgia and the dopamine D4 receptor exon III repeat polymorphism and relationship to novelty seeking personality traits. *Mol. Psychiatry* 9, 730–731. doi: 10.1038/sj.mp.4001568

Cajanus, K., Holmström, E. J., Wessman, M., Anttila, V., Kaunisto, M. A., and Kalso, E. (2016). Effect of endocannabinoid degradation on pain: role of FAAH polymorphisms in experimental and postoperative pain in women treated for breast cancer. *Pain* 157, 361–369. doi: 10.1097/j.pain.0000000000000398

Camilleri, M., Klee, E. W., Shin, A., Carlson, P., Li, Y., Grover, M., et al. (2014). Irritable bowel syndrome-diarrhea: characterization of genotype by exome sequencing, and phenotypes of bile acid synthesis and colonic transit. *Am. J. Physiol. Gastrointest. Liver Physiol.* 306, G13–G26. doi: 10.1152/ajpgi.00294.2013

Camon, E., Magrane, M., Barrell, D., Lee, V., Dimmer, E., Maslen, J., et al. (2004). The gene ontology annotation (GOA) database: sharing knowledge in uniprot with Gene Ontology. *Nucleic Acids Res.* 32, D262–D266. doi: 10.1093/nar/gkh021

Cargnin, S., Viana, M., Sances, G., Bianchi, M., Ghiotto, N., Tassorelli, C., et al. (2014). Combined effect of common gene variants on response to drug withdrawal therapy in medication overuse headache. *Eur. J. Clin. Pharmacol.* 70, 1195–1202. doi: 10.1007/s00228-014-1726-6

Cattaruzza, M., Slodowski, W., Stojakovic, M., Krzesz, R., and Hecker, M. (2003). Interleukin-10 induction of nitric-oxide synthase expression attenuates CD40-mediated interleukin-12 synthesis in human endothelial cells. *J. Biol. Chem.* 278, 37874–37880. doi: 10.1074/jbc.M301670200

Ceolin, L., Siqueira, D. R., Romitti, M., Ferreira, C. V., and Maia, A. L. (2012). Molecular basis of medullary thyroid carcinoma: the role of RET polymorphisms. *Int. J. Mol. Sci.* 13, 221–239. doi: 10.3390/ijms13010221

Chatti, I., Woillard, J. B., Mili, A., Creveaux, I., Ben Charfeddine, I., Feki, J., et al. (2017). Genetic analysis of mu and kappa opioid receptor and COMT enzyme in cancer pain tunisian patients under opioid treatment. *Iran J. Public Health* 46, 1704–1711.

Chen, C. L., Broom, D. C., Liu, Y., De Nooij, J. C., Li, Z., Cen, C., et al. (2006). Runx1 determines nociceptive sensory neuron phenotype and is required for thermal and neuropathic pain. *Neuron* 49, 365–377. doi: 10.1016/j.neuron.2005.10.036

Chen, D., Zhang, T.-L., and Wang, L.-M. (2014). The association of CSF-1 gene polymorphism with chronic periodontitis in the Han Chinese population. *J. Periodontol.* 85, e304–e312. doi: 10.1902/jop.2014.130688

Chen, L., and Brosnan, C. F. (2006). Regulation of immune response by P2X7 receptor. *Crit. Rev. Immunol.* 26, 499–513. doi: 10.1615/CritRevImmunol.v26.i6.30

Chen, X., Williamson, V. S., An, S.-S., Hettema, J. M., Aggen, S. H., Neale, M. C., et al. (2008). Cannabinoid receptor 1 gene association with nicotine dependence. *Arch. Gen. Psychiatry* 65, 816–824. doi: 10.1001/archpsyc.65.7.816

Choi, P., and Reiser, H. (1998). IL-4: role in disease and regulation of production. *Clin. Exp. Immunol.* 113, 317–319. doi: 10.1046/j.1365-2249.1998.00690.x

Clapauch, R., Mourao, A. F., Mecenas, A. S., Maranhao, P. A., Rossini, A., and Bouskela, E. (2014). Endothelial function and insulin resistance in early postmenopausal women with cardiovascular risk factors: importance of ESR1 and NOS3 polymorphisms. *PLoS One* 9:e103444. doi: 10.1371/journal.pone.0103444

Clapham, D. E. (2003). TRP channels as cellular sensors. *Nature* 426, 517–524. doi: 10.1038/nature02196

Coccia, E. M., Stellacci, E., Marziali, G., Weiss, G., and Battistini, A. (2000). IFN-gamma and IL-4 differently regulate inducible NO synthase gene expression through IRF-1 modulation. *Int. Immunol.* 12, 977–985. doi: 10.1093/intimm/12.7.977

Corominas, R., Ribases, M., Camiña, M., Cuenca-León, E., Pardo, J., Boronat, S., et al. (2009). Two-stage case-control association study of dopamine-related genes and migraine. *BMC Med. Genet.* 10:95. doi: 10.1186/1471-2350-10-95

Coskun, S., Varol, S., Ozdemir, H. H., Agacayak, E., Aydın, B., Kapan, O., et al. (2016). Association of brain-derived neurotrophic factor and nerve growth factor gene polymorphisms with susceptibility to migraine. *Neuropsychiatr. Dis. Treat.* 12, 1779–1785. doi: 10.2147/NDT.S108814

Costigan, M., Belfer, I., Griffin, R. S., Dai, F., Barrett, L. B., Coppola, G., et al. (2010). Multiple chronic pain states are associated with a common amino acid-changing allele in KCNS1. *Brain* 133, 2519–2527. doi: 10.1093/brain/awq195

Coyle, J. T., and Tsai, G. (2004). NMDA receptor function, neuroplasticity, and the pathophysiology of schizophrenia. *Int. Rev. Neurobiol.* 59, 491–515. doi: 10.1016/S0074-7742(04)59019-0

Crock, L. W., Stemler, K. M., Song, D. G., Abbosh, P., Vogt, S. K., Qiu, C. S., et al. (2012). Metabotropic glutamate receptor 5 (mGluR5) regulates bladder nociception. *Mol. Pain* 8:20. doi: 10.1186/1744-8069-8-20

Davila-Fajardo, C. L., Marquez, A., Pascual-Salcedo, D., Moreno Ramos, M. J., Garcia-Portales, R., Magro, C., et al. (2014). Confirmation of -174G/C interleukin-6 gene promoter polymorphism as a genetic marker predicting antitumor necrosis factor treatment outcome. *Pharmacogenet. Genom.* 24, 1–5. doi: 10.1097/FPC.0000000000000013

De Luca, C., Rappa, A. G., Gragnano, G., Malapelle, U., Troncone, G., and Barberis, M. (2018). Idylla assay and next generation sequencing: an integrated EGFR mutational testing algorithm. *J. Clin. Pathol.* 71, 745–750. doi: 10.1136/jclinpath-2018-205197

de Rooij, A. M., Florencia Gosso, M., Haasnoot, G. W., Marinus, J., Verduijn, W., Claas, F. H., et al. (2009). HLA-B62 and HLA-DQ8 are associated with complex regional pain syndrome with fixed dystonia. *Pain* 145, 82–85. doi: 10.1016/j.pain.2009.05.015

Deakin, A. M., Payne, A. N., Whittle, B. J., and Moncada, S. (1995). The modulation of IL-6 and TNF-alpha release by nitric oxide following stimulation of J774 cells with LPS and IFN-gamma. *Cytokine* 7, 408–416. doi: 10.1006/cyto.1995.0056

Diatchenko, L., Anderson, A. D., Slade, G. D., Fillingim, R. B., Shabalina, S. A., Higgins, T. J., et al. (2006). Three major haplotypes of the beta2 adrenergic receptor define psychological profile, blood pressure, and the risk for development of a common musculoskeletal pain disorder. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* 141B, 449–462. doi: 10.1002/ajmg.b.30324

Diatchenko, L., Slade, G. D., Nackley, A. G., Bhalang, K., Sigurdsson, A., Belfer, I., et al. (2005). Genetic basis for individual variations in pain perception and the development of a chronic pain condition. *Hum. Mol. Genet.* 14, 135–143. doi: 10.1093/hmg/ddi013

Dinarello, C. A. (1994). The biological properties of interleukin-1. *Eur. Cytokine Netw.* 5, 517–531.

Dineley, K. T., Pandya, A. A., and Yakel, J. L. (2015). Nicotinic ACh receptors as therapeutic targets in CNS disorders. *Trends Pharmacol. Sci.* 36, 96–108. doi: 10.1016/j.tips.2014.12.002

D'Mello, R., and Dickenson, A. H. (2008). Spinal cord mechanisms of pain. *Br. J. Anaesth.* 101, 8–16. doi: 10.1093/bja/aen088

Doehring, A., Kirchhof, A., and Lötsch, J. (2009). Genetic diagnostics of functional variants of the human dopamine D2 receptor gene. *Psychiatr. Genet.* 19, 259–268. doi: 10.1097/YPG.0b013e32832d0941

Doehring, A., Küsener, N., Flühr, K., Neddermeyer, T. J., Schneider, G., and Lötsch, J. (2011). Effect sizes in experimental pain produced by gender, genetic variants and sensitization procedures. *PLoS One* 6:e17724. doi: 10.1371/journal.pone.0017724

Dominguez, C. A., Kalliomaki, M., Gunnarsson, U., Moen, A., Sandblom, G., Kockum, I., et al. (2013). The DQB1 *03:02 HLA haplotype is associated with increased risk of chronic pain after inguinal hernia surgery and lumbar disc herniation. *Pain* 154, 427–433. doi: 10.1016/j.pain.2012.12.003

Duan, G., Guo, S., Zhang, Y., Ying, Y., Huang, P., Wang, Q., et al. (2015). The effect of SCN9A variation on basal pain sensitivity in the general population: an experimental study in young women. *J. Pain* 16, 971–980. doi: 10.1016/j.jpain.2015.06.011

Dupont, B., and Ceppellini, R. (1989). *Immunobiology of HLA*. New York, NY: Springer-Verlag.

Dutta, S., Gangopadhyay, P. K., Sinha, S., Chatterjee, A., Ghosh, S., and Rajamma, U. (2011). An association analysis of reelin gene (RELN) polymorphisms with childhood epilepsy in eastern Indian population from West Bengal. *Cell. Mol. Neurobiol.* 31, 45–56. doi: 10.1007/s10571-010-9551-7

Dziedziejko, V., Kurzawski, M., Safranow, K., Chlubek, D., and Pawlik, A. (2011). The effect of ESR1 and ESR2 gene polymorphisms on the outcome of rheumatoid arthritis treatment with leflunomide. *Pharmacogenomics* 12, 41–47. doi: 10.2217/pgs.10.164

Edwards, R. R. (2006). Genetic predictors of acute and chronic pain. *Curr. Rheumatol. Rep.* 8, 411–417. doi: 10.1007/s11926-006-0034-2

Emigh, T. H. (1980). A comparison of tests for hardy-weinberg equilibrium. *Biometrics* 36, 627–642. doi: 10.2307/2556115

Ersig, A. L., Schutte, D. L., Standley, J., Leslie, E., Zimmerman, B., Kleiber, C., et al. (2017). Relationship of genetic variants with procedural pain, anxiety, and distress in children. *Biol. Res. Nurs.* 19, 339–349. doi: 10.1177/1099800417692878

Eskola, P. J., Kjaer, P., Daavittila, I. M., Solovieva, S., Okuloff, A., Sorensen, J. S., et al. (2010). Genetic risk factors of disc degeneration among 12-14-year-old danish children: a population study. *Int. J. Mol. Epidemiol. Genet.* 1, 158–165.

Feng, Y., Egan, B., and Wang, J. (2016). Genetic factors in intervertebral disc degeneration. *Genes Dis.* 3, 178–185. doi: 10.1016/j.gendis.2016.04.005

Feng, Y., Zhao, X., Zhou, C., Yang, L., Liu, Y., Bian, C., et al. (2013). The associations between the Val158Met in the catechol-O-methyltransferase (COMT) gene and the risk of uterine leiomyoma (ULM). *Gene* 529, 296–299. doi: 10.1016/j.gene.2013.07.019

Fillingim, R. B., Wallace, M. R., Herbstman, D. M., Ribeiro-Dasilva, M., and Staud, R. (2008). Genetic contributions to pain: a review of findings in humans. *Oral Dis.* 14, 673–682. doi: 10.1111/j.1601-0825.2008.01458.x

Florio, S. K., Loh, C., Huang, S. M., Iwamaye, A. E., Kitto, K. F., Fowler, K. W., et al. (2009). Disruption of nNOS-PSD95 protein-protein interaction inhibits acute thermal hyperalgesia and chronic mechanical allodynia in rodents. *Br. J. Pharmacol.* 158, 494–506. doi: 10.1111/j.1476-5381.2009.00300.x

Franchimont, D., Martens, H., Hagelstein, M. T., Louis, E., Dewe, W., Chrousos, G. P., et al. (1999). Tumor necrosis factor alpha decreases, and interleukin-10 increases, the sensitivity of human monocytes to dexamethasone: potential regulation of the glucocorticoid receptor. *J. Clin. Endocrinol. Metab.* 84, 2834–2839.

Fridovich, I. (1978). The biology of oxygen radicals. *Science* 201, 875–880. doi: 10.1126/science.210504

Fujii, T., Ota, M., Hori, H., Hattori, K., Teraishi, T., Matsuo, J., et al. (2014). The common functional FKBP5 variant rs1360780 is associated with altered cognitive function in aged individuals. *Sci. Rep.* 4:6696. doi: 10.1038/srep06696

Gadotti, V. M., and Zamponi, G. W. (2011). Cellular prion protein protects from inflammatory and neuropathic pain. *Mol. Pain* 7:59. doi: 10.1186/1744-8069-7-59

Gansner, E. R., and North, S. C. (2000). An open graph visualization system and its applications to software engineering. *Softw. Pract. Exp.* 30, 1203–1233. doi: 10.1002/1097-024X(200009)30:11<1203::AID-SPE338>3.0.CO;2-N

Garassino, M. C., Piva, S., La Verde, N., Spagnoletti, I., Iorno, V., Carbone, C., et al. (2013). Randomised phase II trial (NCT00637975) evaluating activity and toxicity of two different escalating strategies for pregabalin and oxycodone combination therapy for neuropathic pain in cancer patients. *PLoS One* 8:e59981. doi: 10.1371/journal.pone.0059981

Garriock, H. A., Tanowitz, M., Kraft, J. B., Dang, V. C., Peters, E. J., Jenkins, G. D., et al. (2010). Association of mu-opioid receptor variants and response to citalopram treatment in major depressive disorder. *Am. J. Psychiatry* 167, 565–573. doi: 10.1176/appi.ajp.2009.08081167

Gębura, K., Świerkot, J., Wysoczańska, B., Korman, L., Nowak, B., Wiland, P., et al. (2017). Polymorphisms within genes involved in regulation of the NF-κB pathway in patients with rheumatoid arthritis. *Int. J. Mol. Sci.* 18, E1432. doi: 10.3390/ijms18071432

Gendron, F. P., Chalimoniuk, M., Strosznajder, J., Shen, S., Gonzalez, F. A., Weisman, G. A., et al. (2003). P2X7 nucleotide receptor activation enhances IFN gamma-induced type II nitric oxide synthase activity in BV-2 microglial cells. *J. Neurochem.* 87, 344–352. doi: 10.1046/j.1471-4159.2003.01995.x

Georgi, A., Jamra, R. A., Klein, K., Villela, A. W., Schumacher, J., Becker, T., et al. (2007). Possible association between genetic variants at the GRIN1 gene and schizophrenia with lifetime history of depressive symptoms in a German sample. *Psychiatr. Genet.* 17, 308–310. doi: 10.1097/YPG.0b013e3280c1e5fb

Gibbs, J. E., and Ray, D. W. (2013). The role of the circadian clock in rheumatoid arthritis. *Arthritis Res. Ther.* 15:205. doi: 10.1186/ar4146

Gilron, I., and Ghasemlou, N. (2014). Chronobiology of chronic pain: focus on diurnal rhythmicity of neuropathic pain. *Curr. Opin. Support. Palliat. Care* 8, 429–436. doi: 10.1097/SPC.0000000000000085

Goodin, B. R., Ness, T. J., and Robbins, M. T. (2015). Oxytocin – A multifunctional analgesic for chronic deep tissue pain. *Curr. Pharm. Des.* 21, 906–913. doi: 10.2174/1381612820666141027111843

Goodyer, I. M., Croudace, T., Dunn, V., Herbert, J., and Jones, P. B. (2010). Cohort profile: risk patterns and processes for psychopathology emerging during adolescence: the ROOTS project. *Int. J. Epidemiol.* 39, 361–369. doi: 10.1093/ije/dyp173

Gramage, E., and Herradon, G. (2010). Genetic deletion of pleiotrophin leads to disruption of spinal nociceptive transmission: evidence for pleiotrophin modulation of morphine-induced analgesia. *Eur. J. Pharmacol.* 647, 97–102. doi: 10.1016/j.ejphar.2010.08.029

Greenbaum, L., Tegeder, I., Barhum, Y., Melamed, E., Roditi, Y., and Djaldetti, R. (2012). Contribution of genetic variants to pain susceptibility in Parkinson disease. *Eur. J. Pain* 16, 1243–1250. doi: 10.1002/j.1532-2149.2012.00134.x

Guerrero, M., Urbano, M., Brown, S. J., Cayanan, C., Ferguson, J., Cameron, M., et al. (2010). "Optimization and characterization of an opioid kappa receptor (OPRK1) antagonist," in *Probe Reports from the NIH Molecular Libraries Program*. Bethesda (MD): National Center for Biotechnology Information (US).

Hack, L. M., Kalsi, G., Aliev, F., Kuo, P.-H., Prescott, C. A., Patterson, D. G., et al. (2011). Limited associations of dopamine system genes with alcohol dependence and related traits in the Irish Affected Sib Pair Study of Alcohol Dependence (IASPSAD). *Alcohol. Clin. Exp. Res.* 35, 376–385. doi: 10.1111/j.1530-0277.2010.01353.x

Hagelberg, N., Forssell, H., Aalto, S., Rinne, J. O., Scheinin, H., Taiminen, T., et al. (2003). Altered dopamine D2 receptor binding in atypical facial pain. *Pain* 106, 43–48. doi: 10.1016/S0304-3959(03)00275-6

Hagelberg, N., Jaaskelainen, S. K., Martikainen, I. K., Mansikka, H., Forssell, H., Scheinin, H., et al. (2004). Striatal dopamine D2 receptors in modulation of pain in humans: a review. *Eur. J. Pharmacol.* 500, 187–192. doi: 10.1016/j.ejphar.2004.07.024

Hahn, W.-H., Suh, J.-S., and Cho, B.-S. (2011). Linkage and association study of neurotrophins and their receptors as novel susceptibility genes for childhood IgA nephropathy. *Pediatr. Res.* 69, 299–305. doi: 10.1203/PDR.0b013e31820b9365

Haus, E., Sackett-Lundeen, L., and Smolensky, M. H. (2012). Rheumatoid arthritis: circadian rhythms in disease activity, signs and symptoms, and rationale for chronotherapy with corticosteroids and other medications. *Bull. NYU Hosp. Jt. Dis.* 70(Suppl. 1), 3–10.

He, G.-H., Cai, W.-K., Zhang, J.-B., Ma, C.-Y., Yan, F., Lu, J., et al. (2016). Associations of polymorphisms in HRH2, HRH3, DAO, and HNMT genes with risk of chronic heart failure. *Biomed. Res. Int.* 2016:1208476. doi: 10.1155/2016/1208476

Herman, A. I., and Balogh, K. N. (2012). Polymorphisms of the serotonin transporter and receptor genes: susceptibility to substance abuse. *Subst. Abuse Rehabil.* 3, 49–57. doi: 10.2147/SAR.S25864

Hillard, C. J., Weinlander, K. M., and Stuhr, K. L. (2012). Contributions of endocannabinoid signaling to psychiatric disorders in humans: genetic and biochemical evidence. *Neuroscience* 204, 207–229. doi: 10.1016/j.neuroscience.2011.11.020

Ho, A. M., Tang, N. L., Cheung, B. K., and Stadlin, A. (2008). Dopamine receptor D4 gene -521C/T polymorphism is associated with opioid dependence through cold-pain responses. *Ann. N. Y. Acad. Sci.* 1139, 20–26. doi: 10.1196/annals.1432.054

Hocking, L. J., Smith, B. H., Jones, G. T., Reid, D. M., Strachan, D. P., and Macfarlane, G. J. (2010). Genetic variation in the beta2-adrenergic receptor but not catecholamine-O-methyltransferase predisposes to chronic pain: results from the 1958 British Birth Cohort study. *Pain* 149, 143–151. doi: 10.1016/j.pain.2010.01.023

Hogart, A., Nagarajan, R. P., Patzel, K. A., Yasui, D. H., and Lasalle, J. M. (2007). 15q11-13 GABAA receptor genes are normally biallelically expressed in brain yet are subject to epigenetic dysregulation in autism-spectrum disorders. *Hum. Mol. Genet.* 16, 691–703. doi: 10.1093/hmg/ddm014

Hohmann, S. W., Angioni, C., Tunaru, S., Lee, S., Woolf, C. J., Offermanns, S., et al. (2017). The G2A receptor (GPR132) contributes to oxaliplatin-induced mechanical pain hypersensitivity. *Sci. Rep.* 7:446. doi: 10.1038/s41598-017-00591-0

Hough, L. B., and Rice, F. L. (2011). H3 receptors and pain modulation: peripheral, spinal, and brain interactions. *J. Pharmacol. Exp. Ther.* 336, 30–37. doi: 10.1124/jpet.110.171264

Hu, Y., Tang, W., Liu, R., Dong, Z., Chen, X., Pan, M., et al. (2014). Higher prevalence of migraine in essential tremor: a case-control study. *Cephalalgia* 34, 1142–1149. doi: 10.1177/0333102414531153

Huang, L., Adachi, N., Nagaro, T., Liu, K., and Arai, T. (2007). Histaminergic involvement in neuropathic pain produced by partial ligation of the sciatic nerve in rats. *Reg. Anesth Pain. Med.* 32, 124–129. doi: 10.1097/00115550-200703000-00006

Hulse, R. P., Donaldson, L. F., and Wynick, D. (2012). Peripheral galanin receptor 2 as a target for the modulation of pain. *Pain Res. Treat.* 2012, 545386. doi: 10.1155/2012/545386

Hutchinson, M. R., Zhang, Y., Shridhar, M., Evans, J. H., Buchanan, M. M., Zhao, T. X., et al. (2010). Evidence that opioids may have toll-like receptor 4 and MD-2 effects. *Brain Behav. Immun.* 24, 83–95. doi: 10.1016/j.bbi.2009.08.004

Hwang, I. C., Park, J. Y., Myung, S. K., Ahn, H. Y., Fukuda, K., and Liao, Q. (2014). OPRM1 A118G gene variant and postoperative opioid requirement: a systematic review and meta-analysis. *Anesthesiology* 121, 825–834. doi: 10.1097/ALN.0000000000000405

Ide, S., Nishizawa, D., Fukuda, K.-I., Kasai, S., Hasegawa, J., Hayashida, M., et al. (2014). Haplotypes of P2RX7 gene polymorphisms are associated with both cold pain sensitivity and analgesic effect of fentanyl. *Mol. Pain* 10:75. doi: 10.1186/1744-8069-10-75

Jaaskelainen, S. K., Lindholm, P., Valmunen, T., Pesonen, U., Taiminen, T., Virtanen, A., et al. (2014). Variation in the dopamine D2 receptor gene plays a key role in human pain and its modulation by transcranial magnetic stimulation. *Pain* 155, 2180–2187. doi: 10.1016/j.pain.2014.08.029

James, S. (2013). Human pain and genetics: some basics. *Br. J. Pain* 7, 171–178. doi: 10.1177/2049463713506408

Jayamanne, A., Greenwood, R., Mitchell, V. A., Aslan, S., Piomelli, D., and Vaughan, C. W. (2006). Actions of the FAAH inhibitor URB597 in neuropathic and inflammatory chronic pain models. *Br. J. Pharmacol.* 147, 281–288. doi: 10.1038/sj.bjp.0706510

Jhun, E., He, Y., Yao, Y., Molokie, R. E., Wilkie, D. J., and Wang, Z. J. (2014). Dopamine D3 receptor Ser9Gly and catechol-o-methyltransferase Val158Met polymorphisms and acute pain in sickle cell disease. *Anesth. Analg.* 119, 1201–1207. doi: 10.1213/ANE.0000000000000382

Jhun, E. H., Yao, Y., He, Y., Mack, A. K., Wilkie, D. J., Molokie, R. E., et al. (2015). Prevalence of pain-related single nucleotide polymorphisms in patients of African origin with sickle cell disease. *Pharmacogenomics* 16, 1795–1806. doi: 10.2217/pgs.15.126

Jurečeková, J., Babušíková, E., Kmeťová, M., Kliment, J., and Dobrota, D. (2015). Estrogen receptor alpha polymorphisms and the risk of prostate cancer development. *J. Cancer Res. Clin. Oncol.* 141, 1963–1971. doi: 10.1007/s00432-015-1966-6

Katusic, Z. S., Stelter, A., and Milstien, S. (1998). Cytokines stimulate GTP cyclohydrolase I gene expression in cultured human umbilical vein endothelial cells. *Arterioscler. Thromb. Vasc. Biol.* 18, 27–32. doi: 10.1161/01.ATV.18.1.27

Kaunisto, M. A., Jokela, R., Tallgren, M., Kambur, O., Tikkanen, E., Tasmuth, T., et al. (2013). Pain in 1000 women treated for breast cancer: a prospective study of pain sensitivity and postoperative pain. *Anesthesiology* 119, 1410–1421. doi: 10.1097/ALN.0000000000000012

Khalil, H., Sereika, S. M., Dai, F., Alexander, S., Conley, Y., Gruen, G., et al. (2017). OPRM1 and COMT gene-gene interaction is associated with postoperative pain and opioid consumption after orthopedic trauma. *Biol. Res. Nurs.* 19, 170–179. doi: 10.1177/1099800416680474

Kim, H., Lee, H., Rowan, J., Brahim, J., and Dionne, R. A. (2006a). Genetic polymorphisms in monoamine neurotransmitter systems show only weak association with acute post-surgical pain in humans. *Mol. Pain* 2:24. doi: 10.1186/1744-8069-2-24

Kim, H., Mittal, D. P., Iadarola, M. J., and Dionne, R. A. (2006b). Genetic predictors for acute experimental cold and heat pain sensitivity in humans. *J. Med. Genet.* 43:e40. doi: 10.1136/jmg.2005.036079

Kim, H., Neubert, J. K., San Miguel, A., Xu, K., Krishnaraju, R. K., Iadarola, M. J., et al. (2004). Genetic influence on variability in human acute experimental pain sensitivity associated with gender, ethnicity and psychological temperament. *Pain* 109, 488–496. doi: 10.1016/j.pain.2004.02.027

Kim, H. Y. (2015). Phospholipids: a neuroinflammation emerging target. *Nat. Chem. Biol.* 11, 99–100. doi: 10.1038/nchembio.1740

Kim, J. I., Kim, J.-W., Park, J.-E., Park, S., Hong, S.-B., Han, D. H., et al. (2017). Association of the GRIN2B rs2284411 polymorphism with methylphenidate response in attention-deficit/hyperactivity disorder. *J. Psychopharmacol.* 31, 1070–1077. doi: 10.1177/0269881116667707

Klein, C. J., Wu, Y., Kilfoyle, D. H., Sandroni, P., Davis, M. D., Gavrilova, R. H., et al. (2013). Infrequent SCN9A mutations in congenital insensitivity to pain and erythromelalgia. *J. Neurol. Neurosurg. Psychiatry* 84, 386–391. doi: 10.1136/jnnp-2012-303719

Klepstad, P., Rakvag, T. T., Kaasa, S., Holthe, M., Dale, O., Borchgrevink, P. C., et al. (2004). The 118 A > G polymorphism in the human micro-opioid receptor gene may increase morphine requirements in patients with pain caused by malignant disease. *Acta Anaesthesiol. Scand.* 48, 1232–1239. doi: 10.1111/j.1399-6576.2004.00517.x

Ko, S. W., Vadakkan, K. I., Ao, H., Gallitano-Mendel, A., Wei, F., Milbrandt, J., et al. (2005). Selective contribution of Egr1 (zif/268) to persistent inflammatory pain. *J. Pain* 6, 12–20. doi: 10.1016/j.jpain.2004.10.001

Kooloos, W. M., Wessels, J. A., Van Der Straaten, T., Allaart, C. F., Huizinga, T. W., and Guchelaar, H. J. (2010). Functional polymorphisms and methotrexate treatment outcome in recent-onset rheumatoid arthritis. *Pharmacogenomics* 11, 163–175. doi: 10.2217/pgs.09.139

Kringel, D., and Lötsch, J. (2015). Pain research funding by the european union seventh framework programme. *Eur. J. Pain* 19, 595–600. doi: 10.1002/ejp.690

Kringel, D., Sisignano, M., Zinn, S., and Lötsch, J. (2017). Next-generation sequencing of the human *TRPV1* gene and the regulating co-players *LTB4R* and *LTB4R2* based on a custom AmpliSeq panel. *PLoS One* 12:e0180116. doi: 10.1371/journal.pone.0180116

Kumar, D., Chakraborty, J., and Das, S. (2012). Epistatic effects between variants of kappa-opioid receptor gene and A118G of mu-opioid receptor gene increase susceptibility to addiction in Indian population. *Prog. Neuropsychopharmacol. Biol. Psychiatry* 36, 225–230. doi: 10.1016/j.pnpbp.2011.10.018

Kumar, V., and Mahal, B. A. (2012). NGF – The TrkA to successful pain treatment. *J. Pain Res.* 5, 279–287. doi: 10.2147/JPR.S33408

Kurita, G. P., Ekholm, O., Kaasa, S., Klepstad, P., Skorpen, F., and Sjøgren, P. (2016). Genetic variation and cognitive dysfunction in opioid-treated patients with cancer. *Brain Behav.* 6:e00471. doi: 10.1002/brb3.471

Kwon, H. J., Jang, W.-C., and Lim, M. H. (2016). Association between RELN gene polymorphisms and attention deficit hyperactivity disorder in korean children. *Psychiatry Investig.* 13, 210–216. doi: 10.4306/pi.2016.13.2.210

Lacroix-Fralish, M. L., Ledoux, J. B., and Mogil, J. S. (2007). The pain genes database: an interactive web browser of pain-related transgenic knockout studies. *Pain* 131, e1–e4. doi: 10.1016/j.pain.2007.04.041

Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., et al. (2001). Initial sequencing and analysis of the human genome. *Nature* 409, 860–921. doi: 10.1038/35057062

Law, P. Y., Reggio, P. H., and Loh, H. H. (2013). Opioid receptors: toward separation of analgesic from undesirable effects. *Trends Biochem. Sci.* 38, 275–282. doi: 10.1016/j.tibs.2013.03.003

Ledergerber, C., and Dessimoz, C. (2011). Base-calling for next-generation sequencing platforms. *Brief. Bioinform.* 12, 489–497. doi: 10.1093/bib/bbq077

Lee, M., and Tracey, I. (2013). Neuro-genetics of persistent pain. *Curr. Opin. Neurobiol.* 23, 127–132. doi: 10.1016/j.conb.2012.11.007

Leung, L., and Cahill, C. M. (2010). TNF-alpha and neuropathic pain–a review. *J. Neuroinflamm.* 7:27. doi: 10.1186/1742-2094-7-27

Li, J. Y., Tao, F., Wu, X. X., Tan, Y. Z., He, L., and Lu, H. (2015). Polymorphic variations in manganese superoxide dismutase (MnSOD) and endothelial nitric oxide synthase (eNOS) genes contribute to the development of type 2 diabetes mellitus in the Chinese Han population. *Genet. Mol. Res.* 14, 12993–13002. doi: 10.4238/2015.October.21.20

Li, T., Zhao, J., Yang, J., Ma, X., Dai, Q., Huang, H., et al. (2016). A Meta-Analysis of the association between ESR1 genetic variants and the risk of breast cancer. *PLoS One* 11:e0153314. doi: 10.1371/journal.pone.0153314

Lin, C.-H., Chaudhuri, K. R., Fan, J.-Y., Ko, C.-I., Rizos, A., Chang, C.-W., et al. (2017). Depression and Catechol-O-methyltransferase (COMT) genetic variants are associated with pain in Parkinson's disease. *Sci. Rep.* 7:6306. doi: 10.1038/s41598-017-06782-z

Lindstedt, F., Karshikoff, B., Schalling, M., Olgart Hoglund, C., Ingvar, M., Lekander, M., et al. (2012). Serotonin-1A receptor polymorphism (rs6295) associated with thermal pain perception. *PLoS One* 7:e43221. doi: 10.1371/journal.pone.0043221

Lippmann, C., Kringel, D., Ultsch, A., and Lötsch, J. (2018). Computational functional genomics-based approaches in analgesic drug discovery and repurposing. *Pharmacogenomics* 19, 783–797. doi: 10.2217/pgs-2018-0036

Liu, J., Shi, Y., Tang, J., Guo, T., Li, X., Yang, Y., et al. (2005). SNPs and haplotypes in the S100B gene reveal association with schizophrenia. *Biochem. Biophys. Res. Commun.* 328, 335–341. doi: 10.1016/j.bbrc.2004.12.175

Liu, Y. N., Yang, X., Suo, Z. W., Xu, Y. M., and Hu, X. D. (2014). Fyn kinase-regulated NMDA receptor- and AMPA receptor-dependent pain sensitization in spinal dorsal horn of mice. *Eur. J. Pain* 18, 1120–1128. doi: 10.1002/j.1532-2149.2014.00455.x

Loggia, M. L., Chonde, D. B., Akeju, O., Arabasz, G., Catana, C., Edwards, R. R., et al. (2015). Evidence for brain glial activation in chronic pain patients. *Brain* 138, 604–615. doi: 10.1093/brain/awu377

Lohmann, K., and Klein, C. (2014). Next generation sequencing and the future of genetic diagnosis. *Neurotherapeutics* 11, 699–707. doi: 10.1007/s13311-014-0288-8

Loncar, Z., Curic, G., Mestrovic, A. H., Mickovic, V., and Bilic, M. (2013). Do IL-1B and IL-1RN modulate chronic low back pain in patients with post-traumatic stress disorder? *Collegium Antropol.* 37, 1237–1244.

Lötsch, J., Doehring, A., Mogil, J. S., Arndt, T., Geisslinger, G., and Ultsch, A. (2013). Functional genomics of pain in analgesic drug development and therapy. *Pharmacol. Ther.* 139, 60–70. doi: 10.1016/j.pharmthera.2013.04.004

Lötsch, J., Fluhr, K., Neddermayer, T., Doehring, A., and Geisslinger, G. (2009a). The consequence of concomitantly present functional genetic variants for the identification of functional genotype-phenotype associations in pain. *Clin. Pharmacol. Ther.* 85, 25–30. doi: 10.1038/clpt.2008.103

Lötsch, J., Geisslinger, G., and Tegeder, I. (2009b). Genetic modulation of the pharmacological treatment of pain. *Pharmacol. Ther.* 124, 168–184. doi: 10.1016/j.pharmthera.2009.06.010

Lötsch, J., Von Hentig, N., Freynhagen, R., Griessinger, N., Zimmermann, M., Doehring, A., et al. (2009c). Cross-sectional analysis of the influence of currently known pharmacogenetic modulators on opioid therapy in outpatient pain centers. *Pharmacogenet. Genom.* 19, 429–436. doi: 10.1097/FPC.0b013e32832b89da

Lötsch, J., and Geisslinger, G. (2005). Are mu-opioid receptor polymorphisms important for clinical opioid therapy? *Trends Mol. Med.* 11, 82–89. doi: 10.1016/j.molmed.2004.12.006

Lötsch, J., and Geisslinger, G. (2010). A critical appraisal of human genotyping for pain therapy. *Trends Pharmacol. Sci.* 31, 312–317. doi: 10.1016/j.tips.2010.04.002

Lötsch, J., Klepstad, P., Doehring, A., and Dale, O. (2010). A GTP cyclohydrolase 1 genetic variant delays cancer pain. *Pain* 148, 103–106. doi: 10.1016/j.pain.2009.10.021

Lötsch, J., Sipilä, R., Tasmuth, T., Kringel, D., Estlander, A. M., Meretoja, T., et al. (2018). Machine-learning-derived classifier predicts absence of persistent pain after breast cancer surgery with high accuracy. *Breast Cancer Res./Treatment*. [Epub ahead of print]. doi: 10.1007/s10549-018-4841-8

Lötsch, J., Skarke, C., Grosch, S., Darimont, J., Schmidt, H., and Geisslinger, G. (2002). The polymorphism A118G of the human mu-opioid receptor gene decreases the pupil constrictory effect of morphine-6-glucuronide but not that of morphine. *Pharmacogenetics* 12, 3–9. doi: 10.1097/00008571-200201000-00002

Lötsch, J., Stuck, B., and Hummel, T. (2006). The human mu-opioid receptor gene polymorphism 118A > G decreases cortical activation in response to specific nociceptive stimulation. *Behav. Neurosci.* 120, 1218–1224. doi: 10.1037/0735-7044.120.6.1218

Lu, H., Zhu, L., Lian, L., Chen, M., Shi, D., and Wang, K. (2015). Genetic variations in the PRKCG gene and osteosarcoma risk in a Chinese population: a case-control study. *Tumour Biol.* 36, 5241–5247. doi: 10.1007/s13277-015-3182-z

Maertens, O., De Schepper, S., Vandesompele, J., Brems, H., Heyns, I., Janssens, S., et al. (2007). Molecular dissection of isolated disease features in mosaic neurofibromatosis type 1. *Am. J. Hum. Genet.* 81, 243–251. doi: 10.1086/519562

Mansour, A. R., Farmer, M. A., Baliki, M. N., and Apkarian, A. V. (2014). Chronic pain: the role of learning and brain plasticity. *Restor. Neurol. Neurosci.* 32, 129–139.

Marcos, M., Pastor, I., De La Calle, C., Barrio-Real, L., Laso, F.-J., and González-Sarmiento, R. (2012). Cannabinoid receptor 1 gene is associated with alcohol dependence. *Alcohol. Clin. Exp. Res.* 36, 267–271. doi: 10.1111/j.1530-0277.2011.01623.x

Martikainen, I. K., Nuechterlein, E. B., Pecina, M., Love, T. M., Cummiford, C. M., Green, C. R., et al. (2015). Chronic back pain is associated with alterations in dopamine neurotransmission in the ventral striatum. *J. Neurosci.* 35, 9957–9965. doi: 10.1523/JNEUROSCI.4605-14.2015

Maxwell, J. R., Potter, C., Hyrich, K. L., Barton, A., Worthington, J., Isaacs, J. D., et al. (2008). Association of the tumour necrosis factor-308 variant with differential response to anti-TNF agents in the treatment of rheumatoid arthritis. *Hum. Mol. Genet.* 17, 3532–3538. doi: 10.1093/hmg/ddn245

McIntosh, A. M., Simen, A. A., Evans, K. L., Hall, J., Macintyre, D. J., Blackwood, D., et al. (2012). Genetic variation in hyperpolarization-activated cyclic nucleotide-gated channels and its relationship with neuroticism, cognition and risk of depression. *Front. Genet.* 3:116. doi: 10.3389/fgene.2012.00116

Mead, S., Poulter, M., Uphill, J., Beck, J., Whitfield, J., Webb, T. E. F., et al. (2009). Genetic risk factors for variant Creutzfeldt-Jakob disease: a genome-wide association study. *Lancet Neurol.* 8, 57–66. doi: 10.1016/S1474-4422(08)70265-5

Metzker, M. L. (2010). Sequencing technologies – The next generation. *Nat. Rev. Genet.* 11, 31–46. doi: 10.1038/nrg2626

Mill, J., Kiss, E., Baji, I., Kapornai, K., Daróczy, G., Vetró, A., et al. (2008). Association study of the estrogen receptor alpha gene (ESR1) and childhood-onset mood disorders. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* 147B, 1323–1326. doi: 10.1002/ajmg.b.30751

Milne, R. L., Burwinkel, B., Michailidou, K., Arias-Perez, J.-I., Zamora, M. P., Menéndez-Rodríguez, P., et al. (2014). Common non-synonymous SNPs associated with breast cancer susceptibility: findings from the Breast Cancer association consortium. *Hum. Mol. Genet.* 23, 6096–6111. doi: 10.1093/hmg/ddu311

Mitra, A. K., Crews, K., Pounds, S., Cao, X., Downing, J. R., Raimondi, S., et al. (2011). Impact of genetic variation in FKBP5 on clinical response in pediatric acute myeloid leukemia patients: a pilot study. *Leukemia* 25, 1354–1356. doi: 10.1038/leu.2011.74

Mocellin, S., Marincola, F., Rossi, C. R., Nitti, D., and Lise, M. (2004). The multifaceted relationship between IL-10 and adaptive immunity: putting together the pieces of a puzzle. *Cytokine Growth. Factor. Rev.* 15, 61–76. doi: 10.1016/j.cytogfr.2003.11.001

Mogil, J. S. (2009). Are we getting anywhere in human pain genetics? *Pain* 146, 231–232. doi: 10.1016/j.pain.2009.07.023

Mogil, J. S. (2012). Pain genetics: past, present and future. *Trends Genet.* 28, 258–266. doi: 10.1016/j.tig.2012.02.004

Mustafa, A. E., Faquih, T., Baz, B., Kattan, R., Al-Issa, A., Tahir, A. I., et al. (2018). Validation of ion torrent(TM) inherited disease panel with the PGM(TM) sequencing platform for rapid and comprehensive mutation detection. *Genes (Basel.)* 9, 267. doi: 10.3390/genes9050267

Nackley, A. G., Shabalina, S. A., Tchivileva, I. E., Satterfield, K., Korchynskyi, O., Makarov, S. S., et al. (2006). Human catechol-O-methyltransferase haplotypes modulate protein expression by altering mRNA secondary structure. *Science* 314, 1930–1933. doi: 10.1126/science.1131262

Naumov, D. E., Perelman, J. M., Kolosov, V. P., Potapova, T. A., Maksimov, V. N., and Zhou, X. (2015). Transient receptor potential melastatin 8 gene polymorphism is associated with cold-induced airway hyperresponsiveness in bronchial asthma. *Respirology* 20, 1192–1197. doi: 10.1111/resp.12605

Nemeth, E., Rivera, S., Gabayan, V., Keller, C., Taudorf, S., Pedersen, B. K., et al. (2004). IL-6 mediates hypoferremia of inflammation by inducing the synthesis of the iron regulatory hormone hepcidin. *J. Clin. Invest.* 113, 1271–1276. doi: 10.1172/JCI200420945

Nicholl, B. I., Holliday, K. L., Macfarlane, G. J., Thomson, W., Davies, K. A., O'neill, T. W., et al. (2011). Association of HTR2A polymorphisms with chronic widespread pain and the extent of musculoskeletal pain: results from two population-based cohorts. *Arthritis Rheum.* 63, 810–818. doi: 10.1002/art.30185

Nissenbaum, J., Devor, M., Seltzer, Z., Gebauer, M., Michaelis, M., Tal, M., et al. (2010). Susceptibility to chronic pain following nerve injury is genetically affected by CACNG2. *Genome Res.* 20, 1180–1190. doi: 10.1101/gr.104976.110

Nöthen, M. M., Cichon, S., Hemmer, S., Hebebrand, J., Remschmidt, H., Lehmkuhl, G., et al. (1994). Human dopamine D4 receptor gene: frequent occurrence of a null allele and observation of homozygosity. *Hum. Mol. Genet.* 3, 2207–2212. doi: 10.1093/hmg/3.12.2207

Obata, K., and Noguchi, K. (2006). BDNF in sensory neurons and chronic pain. *Neurosci. Res.* 55, 1–10. doi: 10.1016/j.neures.2006.01.005

Oertel, B. G., Doehring, A., Roskam, B., Kettner, M., Hackmann, N., Ferreiros, N., et al. (2012). Genetic-epigenetic interaction modulates mu-opioid receptor regulation. *Hum. Mol. Genet.* 21, 4751–4760. doi: 10.1093/hmg/dds314

Oertel, B. G., Schmidt, R., Schneider, A., Geisslinger, G., and Lötsch, J. (2006). The mu-opioid receptor gene polymorphism 118A>G depletes alfentanil-induced analgesia and protects against respiratory depression in homozygous carriers. *Pharmacogenet. Genom.* 16, 625–636. doi: 10.1097/01.fpc.0000220566.90466.a2

Offenbaecher, M., Bondy, B., De Jonge, S., Glatzeder, K., Kruger, M., Schoeps, P., et al. (1999). Possible association of fibromyalgia with a polymorphism in the serotonin transporter gene regulatory region. *Arthritis Rheum.* 42, 2482–2488. doi: 10.1002/1529-0131(199911)42:11<2482::AID-ANR27>3.0.CO;2-B

Onaya, T., Ishii, M., Katoh, H., Shimizu, S., Kasai, H., Kawamura, M., et al. (2013). Predictive index for the onset of medication overuse headache in migraine patients. *Neurol. Sci.* 34, 85–92. doi: 10.1007/s10072-012-0955-7

Onojjighofia, T., Meshkin, B., Nguyen, S. V., Schwartz, D., and Akindele, B. (2014). Perception of analgesia in narcotic users with chronic pain: a multi-center cross-sectional study comparing genotype to pain VAS (P.A.*I.N.* Study). *Neurology* 82:E39.

Oyanguren-Desez, O., Rodríguez-Antigüedad, A., Villoslada, P., Domercq, M., Alberdi, E., and Matute, C. (2011). Gain-of-function of P2X7 receptor gene variants in multiple sclerosis. *Cell Calcium* 50, 468–472. doi: 10.1016/j.ceca.2011.08.002

Palmirotta, R., Barbanti, P., De Marchis, M. L., Egeo, G., Aurilia, C., Fofi, L., et al. (2015). Is SOD2 Ala16Val polymorphism associated with migraine with aura phenotype? *Antioxid. Redox Signal.* 22, 275–279. doi: 10.1089/ars.2014.6069

Park, J. J., Lee, J., Kim, M. A., Back, S. K., Hong, S. K., and Na, H. S. (2007). Induction of total insensitivity to capsaicin and hypersensitivity to garlic extract in human by decreased expression of TRPV1. *Neurosci. Lett.* 411, 87–91. doi: 10.1016/j.neulet.2006.10.046

Peiro, A. M., Planelles, B., Juhasz, G., Bagdy, G., Libert, F., Eschalier, A., et al. (2016). Pharmacogenomics in pain treatment. *Drug Metab. Pers Ther.* 31, 131–142. doi: 10.1515/dmpt-2016-0005

Pelletier, C., and Weidhaas, J. B. (2010). MicroRNA binding site polymorphisms as biomarkers of cancer risk. *Expert. Rev. Mol. Diagn.* 10, 817–829. doi: 10.1586/erm.10.59

Petrenko, A. B., Yamakura, T., Baba, H., and Shimoji, K. (2003). The role of N-methyl-D-aspartate (NMDA) receptors in pain: a review. *Anesth. Analg.* 97, 1108–1116. doi: 10.1213/01.ANE.0000081061.12235.55

Potapova, T. A., Babenko, V. N., Kobzev, V. F., Romashchenko, A. G., Maksimov, V. N., and Voevoda, M. I. (2014). Associations of cold receptor TRPM8 gene single nucleotide polymorphism with blood lipids and anthropometric parameters in Russian population. *Bull. Exp. Biol. Med.* 157, 757–761. doi: 10.1007/s10517-014-2660-4

Potvin, S., Larouche, A., Normand, E., De Souza, J. B., Gaumond, I., Grignon, S., et al. (2009). DRD3 Ser9Gly polymorphism is related to thermal pain perception and modulation in chronic widespread pain patients and healthy controls. *J. Pain* 10, 969–975. doi: 10.1016/j.jpain.2009.03.013

Prasad, H. C., Steiner, J. A., Sutcliffe, J. S., and Blakely, R. D. (2009). Enhanced activity of human serotonin transporter variants associated with autism. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 364, 163–173. doi: 10.1098/rstb.2008.0143

R Development Core Team (2008). *R: A Language and Environment for Statistical Computing*. Vienna: R Development Core Team.

Rakvag, T. T., Klepstad, P., Baar, C., Kvam, T. M., Dale, O., Kaasa, S., et al. (2005). The Val158Met polymorphism of the human catechol-O-methyltransferase (COMT) gene may influence morphine requirements in cancer pain patients. *Pain* 116, 73–78. doi: 10.1016/j.pain.2005.03.032

Reimann, F., Cox, J. J., Belfer, I., Diatchenko, L., Zaykin, D. V., Mchale, D. P., et al. (2010). Pain perception is altered by a nucleotide polymorphism in SCN9A. *Proc. Natl. Acad. Sci. U.S.A.* 107, 5148–5153. doi: 10.1073/pnas.0913181107

Reyes-Gibby, C. C., Shete, S., Rakvag, T., Bhat, S. V., Skorpen, F., Bruera, E., et al. (2007). Exploring joint effects of genes and the clinical efficacy of morphine for cancer pain: OPRM1 and COMT gene. *Pain* 130, 25–30. doi: 10.1016/j.pain.2006.10.023

Ribeiro-Dasilva, M. C., Peres Line, S. R., Leme Godoy, Dos Santos, M. C., Arthuri, M. T., Hou, W., et al. (2009). Estrogen receptor-alpha polymorphisms and predisposition to TMJ disorder. *J. Pain* 10, 527–533. doi: 10.1016/j.jpain.2008.11.012

Rodrigues, P., De Marco, G., Furriol, J., Mansego, M. L., Pineda-Alonso, M., Gonzalez-Neira, A., et al. (2014). Oxidative stress in susceptibility to breast cancer: study in Spanish population. *BMC Cancer* 14:861. doi: 10.1186/1471-2407-14-861

Root, T. L., Szatkiewicz, J. P., Jonassaint, C. R., Thornton, L. M., Pinheiro, A. P., Strober, M., et al. (2011). Association of candidate genes with phenotypic traits relevant to anorexia nervosa. *Eur. Eat Disord. Rev.* 19, 487–493. doi: 10.1002/erv.1138

Sah, D. W., Ossipov, M. H., Rossomando, A., Silvian, L., and Porreca, F. (2005). New approaches for the treatment of pain: the GDNF family of neurotrophic growth factors. *Curr. Top. Med. Chem.* 5, 577–583. doi: 10.2174/1568026054367593

Sanger, F., and Coulson, A. R. (1975). A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *J. Mol. Biol.* 94, 441–448. doi: 10.1016/0022-2836(75)90213-2

Sanger, F., Nicklen, S., and Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. U.S.A.* 74, 5463–5467. doi: 10.1073/pnas.74.12.5463

Sato, M., Ohashi, J., Tsuchiya, N., Kashiwase, K., Ishikawa, Y., Arita, H., et al. (2002). Association of HLA-A*3303-B*4403-DRB1*1302 haplotype, but not of TNFA promoter and NKp30 polymorphism, with postherpetic neuralgia (PHN) in the Japanese population. *Genes Immun.* 3, 477–481. doi: 10.1038/sj.gene.6363890

Savas, S., Schmidt, S., Jarjanazi, H., and Ozcelik, H. (2006). Functional nsSNPs from carcinogenesis-related genes expressed in breast tissue: potential breast cancer risk alleles and their distribution across human populations. *Hum. Genomics* 2, 287–296. doi: 10.1186/1479-7364-2-5-287

Schleinitz, D., Carmienke, S., Böttcher, Y., Tönjes, A., Berndt, J., Klöting, N., et al. (2010). Role of genetic variation in the cannabinoid type 1 receptor gene (CNR1) in the pathophysiology of human obesity. *Pharmacogenomics* 11, 693–702. doi: 10.2217/pgs.10.42

Schmidt, S., Hauser, M. A., Scott, W. K., Postel, E. A., Agarwal, A., Gallins, P., et al. (2006). Cigarette smoking strongly modifies the association of LOC387715 and age-related macular degeneration. *Am. J. Hum. Genet.* 78, 852–864. doi: 10.1086/503822

Schott, K., Gutlich, M., and Ziegler, I. (1993). Induction of GTP-cyclohydrolase I mRNA expression by lectin activation and interferon-gamma treatment in human cells associated with the immune response. *J. Cell. Physiol.* 156, 12–16. doi: 10.1002/jcp.1041560103

Schurks, M., Kurth, T., Buring, J. E., and Zee, R. Y. (2009). A candidate gene association study of 77 polymorphisms in migraine. *J. Pain* 10, 759–766. doi: 10.1016/j.jpain.2009.01.326

Schürks, M., Rist, P. M., and Kurth, T. (2010). Sex hormone receptor gene polymorphisms and migraine: a systematic review and meta-analysis. *Cephalalgia* 30, 1306–1328. doi: 10.1177/0333102410364155

Schütz, M., Oertel, B. G., Heimann, D., Doehring, A., Walter, C., Dimova, V., et al. (2014). Consequences of a human TRPA1 genetic variant on the perception of nociceptive and olfactory stimuli. *PLoS One* 9:e95592. doi: 10.1371/journal.pone.0095592

Schwartz, E. S., Kim, H. Y., Wang, J., Lee, I., Klann, E., Chung, J. M., et al. (2009). Persistent pain is dependent on spinal mitochondrial antioxidant levels. *J. Neurosci.* 29, 159–168. doi: 10.1523/JNEUROSCI.3792-08.2009

Schweinhardt, P., Sauro, K. M., and Bushnell, M. C. (2008). Fibromyalgia: a disorder of the brain? *Neuroscientist* 14, 415–421. doi: 10.1177/1073858407312521

Serý, O., Prikryl, R., Castulík, L., and Sťastný, F. (2010). A118G polymorphism of OPRM1 gene is associated with schizophrenia. *J. Mol. Neurosci.* 41, 219–222. doi: 10.1007/s12031-010-9327-z

Shah, N. D., Shah, P. S., Panchal, Y. Y., Katudia, K. H., Khatri, N. B., Ray, H. S. P., et al. (2018). Mutation analysis of BRCA1/2 mutations with special reference to polymorphic SNPs in Indian breast cancer patients. *Appl. Clin. Genet.* 11, 59–67. doi: 10.2147/TACG.S155955

Sherry, S. T., Ward, M. H., Kholodov, M., Baker, J., Phan, L., Smigielski, E. M., et al. (2001). dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.* 29, 308–311. doi: 10.1093/nar/29.1.308

Shi, J., Badner, J. A., Hattori, E., Potash, J. B., Willour, V. L., Mcmahon, F. J., et al. (2008). Neurotransmission and bipolar disorder: a systematic family-based association study. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* 147B, 1270–1277. doi: 10.1002/ajmg.b.30769

Shi, T.-J. S., Liu, S.-X. L., Hammarberg, H., Watanabe, M., Xu, Z.-Q. D., and Hökfelt, T. (2008). Phospholipase C{beta}3 in mouse and human dorsal root ganglia and spinal cord is a possible target for treatment of neuropathic pain. *Proc. Natl. Acad. Sci. U.S.A.* 105, 20004–20008. doi: 10.1073/pnas.0810899105

Shoskes, D. A., Albakri, Q., Thomas, K., and Cook, D. (2002). Cytokine polymorphisms in men with chronic prostatitis/chronic pelvic pain syndrome: association with diagnosis and treatment response. *J. Urol.* 168, 331–335. doi: 10.1016/S0022-5347(05)64916-6

Siqueira, E. C. D., De Sousa, S. F., França, J. A., Diniz, M. G., Pereira, T. D. S. F., Moreira, R. G., et al. (2017). Targeted next-generation sequencing of glandular odontogenic cyst: a preliminary study. *Oral. Surg. Oral Med. Oral. Pathol. Oral Radiol.* 124, 490–494. doi: 10.1016/j.oooo.2017.07.001

Skarke, C., Kirchhof, A., Geisslinger, G., and Lötsch, J. (2004). Comprehensive mu-opioid-receptor genotyping by pyrosequencing. *Clin. Chem.* 50, 640–644. doi: 10.1373/clinchem.2003.027607

Skarke, C., Kirchhof, A., Geisslinger, G., and Lötsch, J. (2005). Rapid genotyping for relevant CYP1A2 alleles by pyrosequencing. *Eur. J. Clin. Pharmacol.* 61, 887–892. doi: 10.1007/s00228-005-0029-3

Sluka, K. A., and Audette, K. M. (2006). Activation of protein kinase C in the spinal cord produces mechanical hyperalgesia by activating glutamate receptors, but does not mediate chronic muscle-induced hyperalgesia. *Mol. Pain* 2:13. doi: 10.1186/1744-8069-2-13

Smith, F. L., Fujimori, K., Lowe, J., and Welch, S. P. (1998). Characterization of delta9-tetrahydrocannabinol and anandamide antinociception in nonarthritic and arthritic rats. *Pharmacol. Biochem. Behav.* 60, 183–191. doi: 10.1016/S0091-3057(97)00583-2

Snider, W. D., and McMahon, S. B. (1998). Tackling pain at the source: new ideas about nociceptors. *Neuron* 20, 629–632. doi: 10.1016/S0896-6273(00)81003-X

Solus, J. F., Chung, C. P., Oeser, A., Li, C., Rho, Y. H., Bradley, K. M., et al. (2015). Genetics of serum concentration of IL-6 and TNFα in systemic lupus erythematosus and rheumatoid arthritis: a candidate gene analysis. *Clin. Rheumatol.* 34, 1375–1382. doi: 10.1007/s10067-015-2881-6

Somogyi, A. A., Coller, J. K., and Barratt, D. T. (2015). Pharmacogenetics of opioid response. *Clin. Pharmacol. Ther.* 97, 125–127. doi: 10.1002/cpt.23

Somogyi, A. A., Sia, A. T., Tan, E.-C., Coller, J. K., Hutchinson, M. R., and Barratt, D. T. (2016). Ethnicity-dependent influence of innate immune genetic markers on morphine PCA requirements and adverse effects in postoperative pain. *Pain* 157, 2458–2466. doi: 10.1097/j.pain.0000000000000661

Sonoda, T., Takada, J., Iba, K., Asakura, S., Yamashita, T., and Mori, M. (2012). Interaction between ESRα polymorphisms and environmental factors in osteoporosis. *J. Orthop. Res.* 30, 1529–1534. doi: 10.1002/jor.22083

Sorge, R. E., Trang, T., Dorfman, R., Smith, S. B., Beggs, S., Ritchie, J., et al. (2012). Genetically determined P2X7 receptor pore formation regulates variability in chronic pain sensitivity. *Nat. Med.* 18, 595–599. doi: 10.1038/nm.2710

Stephens, K., Cooper, B. A., West, C., Paul, S. M., Baggott, C. R., Merriman, J. D., et al. (2014). Associations between cytokine gene variations and severe persistent breast pain in women following breast cancer surgery. *J. Pain* 15, 169–180. doi: 10.1016/j.jpain.2013.09.015

Sugaya, K., Nishijima, S., Yamada, T., Miyazato, M., Hatano, T., and Ogawa, Y. (2002). Molecular analysis of adrenergic receptor genes and interleukin-4/interleukin-4 receptor genes in patients with interstitial cystitis. *J. Urol.* 168, 2668–2671. doi: 10.1016/S0022-5347(05)64241-3

Sun, Y.-G., Gracias, N. G., Drobish, J. K., Vasko, M. R., Gereau, R. W., and Chen, Z.-F. (2009). The c-kit signaling pathway is involved in the development of persistent pain. *Pain* 144, 178–186. doi: 10.1016/j.pain.2009.04.011

Tegeder, I., Costigan, M., Griffin, R. S., Abele, A., Belfer, I., Schmidt, H., et al. (2006). GTP cyclohydrolase and tetrahydrobiopterin regulate pain sensitivity and persistence. *Nat. Med.* 12, 1269–1277. doi: 10.1038/nm1490

Thuault, S. (2016). A peripheral messenger for chronic pain. *Nat. Neurosci.* 19:9. doi: 10.1038/nn.4217

Thulasiraman, K., and Swamy, M. N. S. (1992). *Graphs : Theory and Algorithms*. New York, NY: Wiley. doi: 10.1002/9781118033104

Tian, W., Fu, Y., Garcia-Elias, A., Fernández-Fernández, J. M., Vicente, R., Kramer, P. L., et al. (2009). A loss-of-function nonsynonymous polymorphism in the osmoregulatory TRPV4 gene is associated with human hyponatremia. *Proc. Natl. Acad. Sci. U.S.A.* 106, 14034–14039. doi: 10.1073/pnas.0904084106

Tsai, S.-J., Hong, C.-J., Cheng, C.-Y., Liao, D.-L., and Liou, Y.-J. (2007). Association study of polymorphisms in post-synaptic density protein 95 (PSD-95) with schizophrenia. *J. Neural Transm. (Vienna)* 114, 423–426. doi: 10.1007/s00702-006-0587-2

Tsantoulas, C., Mooney, E. R., and Mcnaughton, P. A. (2016). HCN2 ion channels: basic science opens up possibilities for therapeutic intervention in neuropathic pain. *Biochem. J.* 473, 2717–2736. doi: 10.1042/BCJ20160287

Tulah, A. S., Beghe, B., Barton, S. J., Holloway, J. W., and Sayers, I. (2012). Leukotriene B4 receptor locus gene characterisation and association studies in asthma. *BMC Med. Genet.* 13:110. doi: 10.1186/1471-2350-13-110

Ultsch, A., Kringel, D., Kalso, E., Mogil, J. S., and Lötsch, J. (2016). A data science approach to candidate gene selection of pain regarded as a process of learning and neural plasticity. *Pain* 157, 2747–2757. doi: 10.1097/j.pain.0000000000000694

Ursu, D., Ebert, P., Langron, E., Ruble, C., Munsie, L., Zou, W., et al. (2014). Gain and loss of function of P2X7 receptors: mechanisms, pharmacology and relevance to diabetic neuropathic pain. *Mol Pain* 10:37. doi: 10.1186/1744-8069-10-37

Vaclavikova, E., Dvorakova, S., Skaba, R., Pos, L., Sykorova, V., Halkova, T., et al. (2014). RET variants and haplotype analysis in a cohort of Czech patients with Hirschsprung disease. *PLoS One* 9:e98957. doi: 10.1371/journal.pone.0098957

Vadakkan, K. I., Wang, H., Ko, S. W., Zastepa, E., Petrovic, M. J., Sluka, K. A., et al. (2006). Genetic reduction of chronic muscle pain in mice lacking calcium/calmodulin-stimulated adenylyl cyclases. *Mol. Pain* 2:7. doi: 10.1186/1744-8069-2-7

Vassalli, P. (1992). The pathophysiology of tumor necrosis factors. *Annu. Rev. Immunol.* 10, 411–452. doi: 10.1146/annurev.iy.10.040192.002211

Venn, J. (1880). On the diagrammatic and mechanical representation of propositions and reasonings. *Dublin Philos. Mag. J. Sci.* 9, 1–18. doi: 10.1080/14786448008626877

Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., et al. (2001). The sequence of the human genome. *Science* 291, 1304–1351. doi: 10.1126/science.1058040

Walker, K., Bowes, M., Panesar, M., Davis, A., Gentry, C., Kesingland, A., et al. (2001). Metabotropic glutamate receptor subtype 5 (mGlu5) and nociceptive function. I. Selective blockade of mGlu5 receptors in models of acute, persistent and chronic pain. *Neuropharmacology* 40, 1–9. doi: 10.1016/S0028-3908(00)00113-1

Walter, C., Doehring, A., Oertel, B. G., and Lötsch, J. (2013). μ-opioid receptor gene variant OPRM1 118 A>G: a summary of its molecular and clinical consequences for pain. *Pharmacogenomics* 14, 1915–1925. doi: 10.2217/pgs.13.187

Wang, C., Wang, H., Pang, J., Li, L., Zhang, S., Song, G., et al. (2014). Glial cell-derived neurotrophic factor attenuates neuropathic pain in a mouse model of chronic constriction injury: possible involvement of E-cadherin/p120ctn signaling. *J. Mol. Neurosci.* 54, 156–163. doi: 10.1007/s12031-014-0266-y

Wang, X.-S., Song, H.-B., Chen, S., Zhang, W., Liu, J.-Q., Huang, C., et al. (2015). Association of single nucleotide polymorphisms of ABCB1, OPRM1 and COMT with pain perception in cancer patients. *J. Huazhong Univ. Sci. Technolog. Med. Sci.* 35, 752–758. doi: 10.1007/s11596-015-1502-6

Wedrén, S., Lovmar, L., Humphreys, K., Magnusson, C., Melhus, H., Syvänen, A.-C., et al. (2008). Estrogen receptor alpha gene polymorphism and endometrial cancer risk–a case-control study. *BMC Cancer* 8:322. doi: 10.1186/1471-2407-8-322

Wellcome Trust, Case Control, and Consortium. (2007). Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 447, 661–678. doi: 10.1038/nature05911

Wise, B. L., Demissie, S., Cupples, L. A., Felson, D. T., Yang, M., Shearman, A. M., et al. (2009). The relationship of estrogen receptor-alpha and -beta genes with osteoarthritis of the hand. *J. Rheumatol.* 36, 2772–2779. doi: 10.3899/jrheum.081208

Wolters, P. L., Burns, K. M., Martin, S., Baldwin, A., Dombi, E., Toledo-Tamula, M. A., et al. (2015). Pain interference in youth with neurofibromatosis type 1 and plexiform neurofibromas and relation to disease severity, social-emotional functioning, and quality of life. *Am. J. Med. Genet. A* 167A, 2103–2113. doi: 10.1002/ajmg.a.37123

Woo, J. H., Kim, H., Kim, J. H., and Kim, J. G. (2015). Cannabinoid receptor gene polymorphisms and bone mineral density in Korean postmenopausal women. *Menopause* 22, 512–519. doi: 10.1097/GME.0000000000000339

Yamamoto, M., Ito, Y., Mitsuma, N., Hattori, N., and Sobue, G. (2003). Pain-related differential expression of NGF, GDNF, IL-6, and their receptors in human vasculitic neuropathies. *Intern. Med.* 42, 1100–1103. doi: 10.2169/internalmedicine.42.1100

Yang, L., Gu, X., Zhang, W., Zhang, J., and Ma, Z. (2014). Cdk5 inhibitor roscovitine alleviates neuropathic pain in the dorsal root ganglia by downregulating N-methyl-D-aspartate receptor subunit 2A. *Neurol. Sci.* 35, 1365–1371. doi: 10.1007/s10072-014-1713-9

Yücel, Y., Coşkun, S., Cengiz, B., Özdemir, H. H., Uzar, E., Çim, A., et al. (2016). Association of polymorphisms within the serotonin receptor genes 5-HTR1A, 5-HTR1B, 5-HTR2A and 5-HTR2C and migraine susceptibility in a turkish population. *Clin. Psychopharmacol. Neurosci.* 14, 250–255. doi: 10.9758/cpn.2016.14.3.250

Yucesoy, B., Johnson, V. J., Lummus, Z. L., Kissling, G. E., Fluharty, K., Gautrin, D., et al. (2012). Genetic variants in antioxidant genes are associated with diisocyanate-induced asthma. *Toxicol. Sci.* 129, 166–173. doi: 10.1093/toxsci/kfs183

Zanette, S. A., Dussan-Sarria, J. A., Souza, A., Deitos, A., Torres, I. L. S., and Caumo, W. (2014). Higher serum S100B and BDNF levels are correlated with a lower pressure-pain threshold in fibromyalgia. *Mol. Pain* 10:46. doi: 10.1186/1744-8069-10-46

Zhang, L., Rao, F., Zhang, K., Khandrika, S., Das, M., Vaingankar, S. M., et al. (2007). Discovery of common human genetic variants of GTP cyclohydrolase 1 (GCH1) governing nitric oxide, autonomic activity, and cardiovascular risk. *J. Clin. Invest.* 117, 2658–2671. doi: 10.1172/JCI31093

Zinn S., Sisignano M., Kern K., Pierre S., Tunaru S., Jordan H., et al. (2017). The leukotriene B4 receptors BLT1 and BLT2 form an antagonistic sensitizing system in peripheral sensory neurons. *J. Biol. Chem.* 292, 6123–6134. doi: 10.1074/jbc.M116.769125

Zorina-Lichtenwalter, K., Meloto, C. B., Khoury, S., and Diatchenko, L. (2016). Genetic predictors of human chronic pain conditions. *Neuroscience* 338, 36–62. doi: 10.1016/j.neuroscience.2016.04.041

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Erklärung zu den Autorenanteilen an der Publikation:

Machine-learned analysis of the association of next-generation sequencing–based genotypes with persistent pain after breast cancer surgery (*printed*)

Name der Zeitschrift: PAIN

Beteiligte Autoren: D Kringel, M A Kaunisto, E Kalso und J Lötsch

Was hat der Promovierende bzw. was haben die Koautoren beigetragen?

(1) zu Entwicklung und Planung

Promovierender: 30%
Autor MAK:20%
Autor EK: 20%
Autor JL: 30%

(2) zur Durchführung der einzelnen Untersuchungen und Experimente

Promovierender: 100% (DNA Extraktion, NGS-Panel Design, Sequenzierung, Varianten Annotation, Methoden Validierung)

(3) zur Erstellung der Datensammlung und Abbildungen

Promovierender: 30% (Datenerhebung, Run-Statistik)
Autor JL: 70% (Erstellung der Abbildungen)

(4) zur Analyse und Interpretation der Daten

Promovierender: 30% (Datenvorverarbeitung, Charakterisierung und Assoziation der Varianten, Dateninterpretation)
Autor JL: 70% (Datenvisualisierung, Programmierung der KI, Anwendung der Klassifikatoren, Datenanalyse, Dateninterpretation)

(5) zum Verfassen des Manuskripts

Promovierender: 40%
Autor MAK: 10%
Autor EK: 10%
Autor JL: 40%

Zustimmende Bestätigungen der oben genannten Angaben:

_____ _____

Datum/Ort Unterschrift Promovend

_____ _____

Datum/Ort Unterschrift Betreuer

Research Paper

PAIN®

Machine-learned analysis of the association of next-generation sequencing–based genotypes with persistent pain after breast cancer surgery

Dario Kringel[a,b], Mari A. Kaunisto[c], Eija Kalso[d], Jörn Lötsch[a,e,*]

Abstract

Cancer and its surgical treatment are among the most important triggering events for persistent pain, but additional factors need to be present for the clinical manifestation, such as variants in pain-relevant genes. In a cohort of 140 women undergoing breast cancer surgery, assigned based on a 3-year follow-up to either a persistent or nonpersistent pain phenotype, next-generation sequencing was performed for 77 genes selected for known functional involvement in persistent pain. Applying machine-learning and item categorization techniques, 21 variants in 13 different genes were found to be relevant to the assignment of a patient to either the persistent pain or the nonpersistent pain phenotype group. In descending order of importance for correct group assignment, the relevant genes comprised *DRD1*, *FAAH*, *GCH1*, *GPR132*, *OPRM1*, *DRD3*, *RELN*, *GABRA5*, *NF1*, *COMT*, *TRPA1*, *ABHD6*, and *DRD4*, of which one in the *DRD4* gene was a novel discovery. Particularly relevant variants were found in the *DRD1* and *GPR132* genes, or in a cis-eCTL position of the *OPRM1* gene. Supervised machine-learning–based classifiers, trained with 2/3 of the data, identified the correct pain phenotype group in the remaining 1/3 of the patients at accuracies and areas under the receiver operator characteristic curves of 65% to 72%. When using conservative classical statistical approaches, none of the variants passed α-corrected testing. The present data analysis approach, using machine learning and training artificial intelligences, provided biologically plausible results and outperformed classical approaches to genotype–phenotype association.

Keywords: Data science, Machine learning, Next-generation sequencing, Cancer pain, Persistent pain, Breast cancer surgery, Patients, Women

1. Introduction

Persistent pain is a major health care issue affecting about a fifth of the European population.[9,28] Cancer and its surgical treatment figure among the most important triggering events. However, persistent pain does not develop in every patient[72]; indeed, its prevalence among breast cancer survivors has been reported to vary between 25% and 60%.[123] In a more recent analysis, 13.5% of women operated for breast cancer reported at least moderate pain 1 year after surgery.[99] This variance may be partly due to the way in which persistent pain was defined in the respective studies. Nevertheless, it has been shown that many different individual patient factors make a major contribution to its development after breast cancer surgery.[85,103,129]

A genetic component contributing to the individual risk of chronic pain has been the subject of research for more than half a century.[36] Current evidence indicates that many gene variants modulate the individual perception of pain and its progression to persistence.[24,25,81,102,159] The association of these gene variants with persistent pain after breast cancer surgery is currently an active research topic (**Table 1**). Selected functional genetic variants clearly have been a research focus.[32] However, it is only recently that the broad available next-generation sequencing (NGS)[100] has provided unrestricted access to all genetic information.[32]

In the present analysis, we assessed whether NGS-derived genotypes were associated with persistent pain in a subgroup of patients who had been treated with breast cancer surgery. Next-generation sequencing allowed us to rank the importance of candidate genes in this clinical context based on any specific variant, thereby exceeding most previous approaches in which preselected variants were used. DNA samples and pain data were available from a cohort of 1000 women[58] among whom 70 had been diagnosed with persistent pain, based on ratings acquired up to 36 months after surgery.[85] A recent data-driven analytical approach was used[63] in which mainly machine-learning algorithms for classifier building were used to uncover the genetic background of persistent pain after breast cancer surgery, although without explicitly aiming to identify a genetic biomarker for clinical use. Techniques of feature selection[41,119] were used to rank the genes in terms of the relative importance of genetic variants, thereby training machine-learned algorithms to perform genotype vs phenotype associations.

Sponsorships or competing interests that may be relevant to content are disclosed at the end of this article.

[a] Institute of Clinical Pharmacology, Goethe-University, Frankfurt am Main, Germany, [b] Faculty of Biological Sciences (FB15), Goethe-University, Frankfurt am Main, Germany, [c] Institute for Molecular Medicine Finland (FIMM), HiLIFE, University of Helsinki, Helsinki, Finland, [d] Division of Pain Medicine, Department of Anaesthesiology, Intensive Care and Pain Medicine, Helsinki University Hospital, University of Helsinki, Helsinki, Finland, [e] Project Group Translational Medicine and Pharmacology (IME-TMP), Fraunhofer Institute of Molecular Biology and Applied Ecology, Frankfurt am Main, Germany

**Corresponding author. Address: Goethe-University, Theodor-Stern-Kai 7, 60590 Frankfurt am Main, Germany. Tel.: +49-69-6301-4589; fax: +49-69-6301-4354. E-mail address: j.loetsch@em.uni-frankfurt.de (J. Lötsch).*

Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's Web site (www.painjournalonline.com).

PAIN 160 (2019) 2263–2277

http://dx.doi.org/10.1097/j.pain.0000000000001616

Table 1

Reports of a genetic modulation of persistent pain in the clinical context of breast cancer.

Gene	NCBI	Variant	Clinical association	Ref.
COMT	1312	rs165774	Increased heat pain sensitivity	55
COMT	1312	rs887200	Increased cold pain sensitivity	55
FAAH	2166	rs324420	Decreased cold pain sensitivity	12
OPRM1	4988	rs1799971	Increased analgesic dosing demands	13
P2RX7	5027	rs208294	Development of persistent postsurgical pain	56
IFNG	3458	rs2069718	Development of persistent postsurgical pain	134
IL1R1	3554	rs2110726	Decreased odds of reporting breast pain	92
IL1R2	7850	rs11674595	Development of severe persistent postsurgical pain	134
IL4	3565	rs2243248	Development of persistent postsurgical pain	134
IL10	3586	rs3024498	Development of persistent postsurgical pain	134
IL13	3596	rs1295686	Increased odds of reporting breast pain	92
NFKB1	4790	rs4648141	Development of persistent postsurgical pain	134
KCNS1	3787	rs4499491	Increased odds of reporting breast pain	14,68
KCNJ3	3760	rs7574878	Decreased odds of reporting breast pain	68
KCNJ6	3763	rs2835914	Decreased odds of reporting breast pain	68
KCNK9	51305	rs3780039	Increased odds of reporting breast pain	68
ADRB2	154	rs2400707	Development of mild persistent postsurgical pain	61
HTR3A	3359	rs10160548	Development of mild persistent postsurgical pain	61
SLC6A2	6530	rs1566652	Development of mild persistent postsurgical pain	61
TPH2	121278	rs11179000	Development of mild persistent postsurgical pain	61
HTR2A	3356	rs2296972	Development of moderate persistent postsurgical pain	61
SLC6A2	6530	rs17841327	Development of moderate persistent postsurgical pain	61
SLC6A3	6531	rs464049	Development of severe persistent postsurgical pain	61
TNF	7124	rs1800610	Development of mild persistent postsurgical pain	135
IL6	3569	rs2069840	Development of mild persistent postsurgical pain	135
CACNG2	10369	rs4820242	Development of persistent postsurgical pain	105

The list was created based on a PubMed search for (Pain OR hyperalgesia OR allodynia) AND (cancer OR tumor) AND breast AND (genetic OR gene) AND (polymorphism OR SNP OR modulation), conducted on June 9, 2018. This search provided 47 hits. After elimination of 31, 16 reports were kept.

2. Methods

2.1. Patients and pain phenotype

The study followed the Declaration of Helsinki, and both the Coordinating Ethics Committee (journal number 136/E6/2006) and the Ethics Committee of the Department of Surgery (148/E6/05) of the Hospital District of Helsinki and Uusimaa approved the study protocol. Informed written consent was obtained from each patient. The cohort has been described in detail previously.[58,99] In brief, 1000 women aged 28 to 75 years, suffering from unilateral nonmetastasized breast cancer, were enrolled during the preoperative visit. Exclusion criteria were neoadjuvant therapy[144] and immediate breast reconstruction surgery. The patients were treated with breast-conserving surgery or mastectomy, sentinel node biopsy, and/or axillary clearance. Postoperative analgesia was standardized, consisting of oral acetaminophen and intravenous oxycodone; no regional anesthesia was used. Adjuvant treatments (radiation therapy, chemotherapy, and hormonal therapy) were given according to international guidelines.[99]

Postsurgical pain intensity was assessed with posted questionnaires at months 1, 6, 12, 24, and 36 after surgery using the numerical rating scale (NRS) ranging from 0 (no pain) to 10 (the most severe pain that can be imagined).[35] For the diagnosis of persistent pain, NRS data acquired 12 to 36 months after the surgery were used. As discussed previously,[129] this more adequately reflects the clinical setting of breast cancer surgery than the original definition of persistent postsurgical pain, which proposes a cutoff at 2 months.[90] This seems premature, as adjuvant therapies after breast cancer surgery continue longer. Persistent pain was defined on the basis of NRS ratings, as described in detail previously.[85] In brief, patients were assigned to the "persistent pain" subgroup if the following conditions applied: $NRS_{month36} > 3$ and $NRS_{month12...month36} > 0$ and $(NRS_{month36} - NRS_{month24}) \geq 0$, whereas patients were assigned to the "non-persistent pain" group if $NRS_{month36} \leq 3$ and $NRS_{month12...month36} \leq 3$. Applying these selection criteria to the cohort of 1000 women, n = 70 patients were identified with persistent pain.[85]

2.2. Next-generation sequencing

The association of the patients' genotype with the pain-related phenotype was explored in a set of candidate genes, using the

complete information from coding and regulatory parts accessible by means of NGS. Next-generation sequencing no longer restricts genotype vs phenotype association assessments to a few single-nucleotide polymorphisms as had been the standard a decade ago.[80] Specifically, NGS DNA sequencing now refers to a number of high-throughput technologies that are capable of parallel analyzes of large numbers of different DNA sequences in a single reaction.[117] Two different commercial NGS platforms are currently widely used for diagnostic purposes: the MiSeq/HiSeq/NextSeq (Illumina, Hayward, CA) and the IonTorrent PGM (Life Technologies, Carlsbad, CA). Although both platforms differ in the technical implementation of NGS,[63,64] multiple studies have shown that both NGS platforms provide reliable sequencing results in routine clinical diagnostics.[22,94,128]

2.3. Gene selection

A computational functional genomics–based approach for candidate gene selection was used, as the present genetic panel had been assembled solely based on evidence of the involvement of the genes in persistent pain or on functional genomic analyses. The Thomson Reuters Integrity database at https://integrity.thomson-pharma.com on July 11, 2018, indicated that 75% of these genes are targets of approved or novel analgesics under current clinical development.[64] The list of genes partially overlapped with 2 other proposed gene sets involved in the modulation of pain[102,159] and was in agreement with the genes listed in the "PainGenes" database (http://www.jbldesign.com/jmogil/enter.html).[67] However, combining all proposals of pain-relevant genes into a large panel was not an option because of the technical limitations of the IonTorrent, restricting the panel size to 500 kb (pipeline version 5.6.2).

2.4. Gene sequencing

Next-generation sequencing was performed for the n = 70 patients assigned to the "persistent pain" group, according to the above-mentioned NRS-based criteria. In addition, a similarly sized random sample of age- and body mass index–matched patients, assigned to the "nonpersistent pain" subgroup based on the above-mentioned criteria, was included in the analysis (age: persistent pain group: 57.9 ± 7.9 years, nonpersistent pain group: 58.1 ± 8 years, the Wilcoxon test: W = 2476.5, P = 0.9136; body mass index: persistent pain group: 26.6 ± 4.8 kg/m^2, nonpersistent pain group: 25.6 ± 3.8 kg/m^2, the Wilcoxon test: W = 2224.5, P = 0.3483). The matching criteria reflected the demographic parameters found to be relevant for the prediction of persistent pain in the complete cohort of 1000 women.[85] Additional relevant parameters included psychological parameters and acute postoperative pain-related parameters; however, matching for those made creation of a 70-case-sized control sample impossible due to too many restrictions. Furthermore, an earlier analysis of the same cohort performed at 1 year after the surgery identified, in addition to above-mentioned demographic or pain-related parameters, smoking as a risk factor.[129] With respect to smoking, the present persistent and nonpersistent pain subsamples were similar (the χ^2 test of smoking behavior: χ^2 = 0.552, df = 3, P = 0.9073).

Sequencing of coding and regulatory parts of the selected genes in the n = 140 DNA samples was performed on an IonTorrent PGM. A multiplex PCR amplification strategy for the coding gene sequences was established online (Ion Ampliseq Designer; http://www.ampliseq.com) to amplify the target regions specified above. All the amplicons were designed with 25 base pair exon padding. After a comparison of several primer design options, the design providing the maximum target sequence coverage was chosen. The ordered 1953 amplicons covered approximately 97.5% of the target sequence. The Enriched Ion Sphere Particles that carried many copies of the same DNA fragment were subjected to sequencing on an Ion 318 Chip to sequence pooled libraries with 7 to 8 samples. The high-capacity 318 chip was chosen (instead of the low-capacity 314 or the medium-capacity 316 chip) to obtain a high-sequencing depth of coverage for a genomic DNA library with >95% of bases at 30 times base calls (30×, ie, 30 unique reads that include a given nucleotide). Finally, sequencing was performed using the sequencing kit (Ion PGM Hi-Q Sequencing Kit; Life Technologies, Darmstadt, Germany), with the 200 base pair single-end run configuration, according to the manufacturer's instructions. The samples were split into 16 equal batches of 8 to 10 samples. The determination of nucleotide variants from the NGS data was performed using the software plugin "The Torrent Variant Caller" as provided by the manufacturer of the IonTorrent device. Specifically, a called variant is defined as a nucleotide that disagrees with the nucleotide found in the reference sequence GRCh37/hg19 (https://www.ncbi.nlm.nih.gov/assembly/GCF_000001405.13/).

2.5. Data analysis

Data analysis was performed using the R software package (version 3.4.4 for Linux; http://CRAN.R-project.org/)[114] on an Intel Core i9 computer running on Ubuntu Linux 18.04.1 64-bit. The analytical protocol for NGS-based genotype vs phenotype association has recently been established in similarly sized cohorts carrying or not carrying a pain-related phenotype.[63] An overview of the data analysis is shown in **Fig. 1**. The analysis was performed in 2 main steps comprising (1) feature selection of most informative genetic variants followed by (2) supervised machine-learning–based genotype × phenotype association using the selected variants.

2.6. Selection of informative genetic variants

As NGS produced many candidate genetic variables, a dimension reduction procedure was implemented to narrow the focus to the most relevant variables. Therefore, the data analysis started with feature selection.[41,119] This was performed in 5 steps as described previously in all detail.[63] In brief, first, variants found in the whole patient cohort at allelic frequencies below 10% were omitted. Second, variants for which the distribution of homozygous and heterozygous carriers differed from expectation, according to the Hardy–Weinberg equilibrium,[45] were omitted. Third, noninformative variants, such as those carried by almost all or only a few subjects, were detected based on the Shannon information,[125] and only the most profitable was retained based on computed ABC analysis.[146] The latter is a categorization technique for the selection of the most important subset among a larger set of positive numerical items, dividing on a mathematical basis the set into 3 disjointed subsets "A," "B," and "C,"[154] of which set "A"contains "the important few."[54] This was calculated using our R package "ABCanalysis" (http://cran.r-project.org/package=ABCanalysis).[146] Fourth, further variants unlikely to provide a suitable basis for phenotype class assignment were excluded based on effect size estimates quantified using χ^2 statistics[110] and again followed by ABC analysis of the χ^2 values. Fifth, random forests combined with computed ABC analysis[88] were used for feature selection. Specifically, a random

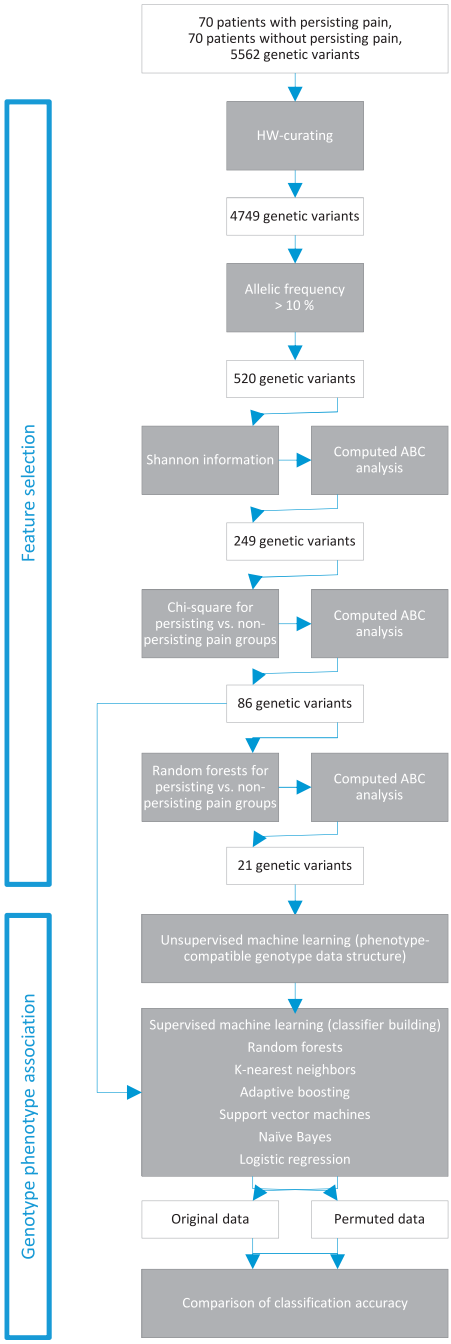


Figure 1. Flowchart of the data analysis, including a complex feature selection step and the analysis of whether several machine-learned algorithms, when trained with the original genetic information, associated the pain phenotype group better than when trained with nonsense information consisting of permuted genotypes. Feature selection started with the full NGS information. Subsequently, variants not in the Hardy–Weinberg equilibrium[45] and those rarer than 10% allelic frequency were eliminated. The next steps involved analysis of information content using the Shannon[125] information criterion and variants most unequally distributed among phenotype groups, based on the χ^2 statistic, selected as the "A" subset of a computed ABC analysis,[146] ie, the "most profitable" items.[109] Finally, the genetic variants were ranked for their importance in a random forest classifier, and only the most relevant items in ABC set "A" were maintained. This selection was used for genotype vs phenotype associations, once as original genotype per patient, and gain with permuted genetic information. To verify that the random forest step in feature selection had not led to a feature set biased toward this particular algorithm, the classification was also performed using the feature set obtained before the last step of feature selection. HW, Hardy–Weinberg; NGS, next-generation sequencing.

forest–based classifier[8,46] was created to assign a patients based on the NGS information to a pain phenotype. Genetic variants were iteratively left out from the classification task, and the decrease in classification accuracy associated with the omitted variant was retained. The calculations were done using the R library "randomForest" (https://cran.r-project.org/package=randomForest).[74] Subsequently, computed ABC analysis was applied to select only variants that, when omitted, produced an important drop in accuracy, ie, belonged to ABC set "A" in this calculation. The analyses were performed following the concept of a nested cross-validation analysis.[148] Specifically, using 1000 times Monte-Carlo[37] resampling, the original data set was class proportionally split into disjoint training (2/3 of the data) and test (1/3 of the data) data subsets. Further measures against overfitting have been described with previous analogous analyses.[63,82]

2.7. Genotype vs phenotype association

After feature selection, the genetic variants selected for further analysis were included in the genotype vs phenotype association assessments. The analysis followed the same concept as applied in a previous analogous assessment previously described in detail.[63] In brief, the relationship between the genotypic and the phenotypic data structure was first explored using unsupervised machine learning to identify a genetic marker pattern that is compatible with the class structure of the phenotypes. Following establishment that the genetic data contained a structure onto which the phenotype class structure can be superimposed to a statistically significant degree, mapping of the genetic parameters to the phenotype classes was performed. The main idea was to train an artificial intelligence, implemented as different types of supervised machine learning, to learn how the genetic information was associated with the pain-related phenotype, and to subsequently use the trained intelligence to predict a phenotype from new genetic data. Should this application perform better than guessing, the phenotype can be regarded as being related to the genotype.

The use of the artificial intelligence algorithms for these 2 main steps of the genotype vs phenotype association analysis has been described in detail previously.[63] In brief, for pattern identification in the prepossessed NGS genotypes, possibly reflected in the phenotype class structure, unsupervised machine learning was implemented as a swarm of intelligent agents called DataBots.[141] A parameter-free focusing projection method of a polar swarm, *Pswarm*, was used that exploits concepts of self-organization and swarm intelligence.[63] Agents of *Pswarm* operate on a toroid polar grid. Following learning, DataBots carrying items with similar features, ie, DataBots associated with similar data points, are placed in groups on the projection grid. The identification of emergent cluster structures was enhanced by displaying the distances between data points as a third dimension implemented as a so-called U-matrix.[83,87] The corresponding visualization technique is a topographical map colored like a geographical map facilitating the recognition of data structures as "valleys" separated by "mountain ridges." The calculations were performed using the R library "DatabionicSwarm" (M. Thrun, https://cran.r-project.org/package=DatabionicSwarm).[140] Finally, projected data cluster was detected using the Ward method.[152] The association of the genotype clusters with the phenotype classes was tested using the Fisher exact statistics.[30]

Supervised algorithms were used to map the genetic input space to the output space, defined by the phenotype class

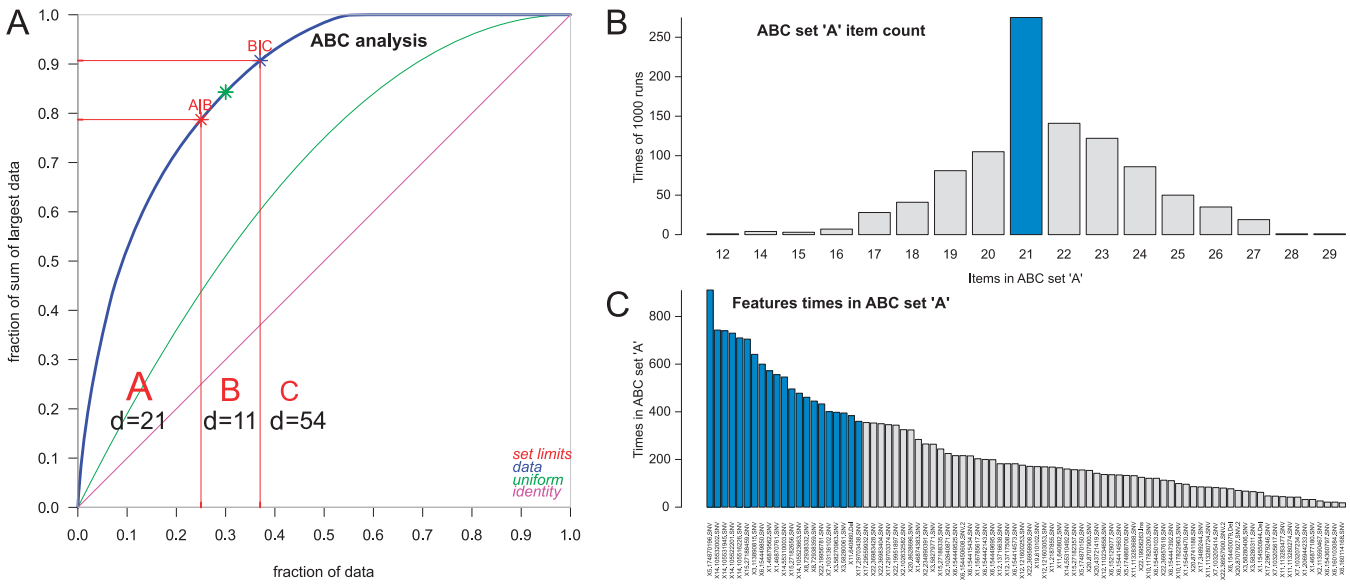


Figure 2. Selection of informative variants using random forest analysis followed by computed ABC analysis: (A) Computed ABC analysis of the forest-based feature ranking (Supplementary Fig. 1, http://links.lww.com/PAIN/A795) of the importance of the genetic variants for class association of patients to the "persistent pain" or the "nonpersistent pain" subgroups. The ABC plot (blue line) shows the cumulative distribution function of the mean decreases in accuracy, along with the identity distribution, x_i = constant (magenta line), ie, each feature contributes similarly to the classification accuracy (for further details about computed ABC analysis, see Ref. 146). The red lines indicate the borders between ABC sets "A," "B," and "C." Only set "A" containing the most profitable items was selected as most genes for the identification of patients with persistent pain. (B) Bar plot of the number of genetic variants found in ABC set A during the 1000 runs. (C) Bar plot of the features' importance in descending order of their appearance in ABC set "A" during the 1000 runs. As set "A" had most frequently a size of d = 21, the 21 variants (blue bars) found most frequently in set "A" were selected as the most relevant lipid mediators for the association with persistent pain. The figure was created using the R software package (version 3.4.4 for Linux; http://CRAN.R-project.org/).[114] In particular, the computed ABC analysis was performed and plotted using our R package "ABCanalysis" (http://cran.r-project.org/package=ABCanalysis).[146]

information. These methods included: (1) random forests,[8,46] (2) adaptive boosting,[121] (3) k-nearest neighbors (kNN[20]), (4) logistic regression,[151] (5) support vector machines,[18] and (6) naive Bayes classification.[2] The methods were chosen heuristically, with the intention to cover a variety of machine-learned classifiers previously used in pain research,[86] such as prototype-based learning (eg, kNN) or ensemble-learning–based classifications (eg, boosting of simple decision trees or random forests), with the addition of classical methods such as logistic regression included for its vicinity to statistical approaches, or naive Bayes.

Briefly, random forests were introduced during the description of the feature selection. Boosting[121] is another variety of ensemble learning that approaches classification through a set of weak learners that are tuned to provide a single strong learner.[59] In the present analysis, adaptive boosting was used as a successful algorithm for binary classification.[122] These calculations were made using the R package "ada" (http://cran.r-project.org/package=ada with the partitioning and classification package "rpart" https://cran.r-project.org/package=rpart).[21] K-nearest neighbor classification[20] was implemented as a 5NN classifier, as the default of the R package "KernelKnn" (Mouselimis L, https://cran.r-project.org/package=KernelKnn). Standard Bayesian classifiers were obtained using the R package "e1071" (Meyer D, https://cran.r-project.org/package=e1071). Support vector machines, which classify data mainly based on geometrical and statistical approaches used for finding an optimum decision surface (hyperplane), were created using the R library "kernlab" (https://cran.r-project.org/package=kernlab).[57] Finally, logistic regression[151] was performed using the "glm" command and the "family = binomial" switch, as implemented in the R "stats" package[114] provided with the basic installation of the software core package (http://www.R-project.org/).

The analyses were performed on the original and the negative control data sets in 1000 cross-validation runs using Monte-Carlo[37] resampling and random splits of the original training data set into disjoint new training (2/3 of the data) and test (1/3 of the data) data subsets. The performance of all classifiers was assessed on the test data subsets created during cross-validation and is reported as the median of the resampling runs. The mapping of the feature space (NGS data) to the output space (pain phenotype groups) was addressed by training the machine-learning methods on the original data set and repeated on a negative control data set created by random permutation of the genetic data in the actual training subset. The expectation was that the prediction would be consistently better when using the original genotypes than when using the permuted genotypes. When trained with permuted data, a classification better than chance, or more precisely, better than a zeroR classifier that simply assigns every case to the majority group resulting in 50% in the present sample, would hint at possible overfitting.

Measures to protect against overfitting were implemented as described elsewhere.[82] These included: (1) tuning of available hyperparameters, (2) the above-mentioned 1000 cross-validation runs using random splits of the original data set into disjoint training (2/3 of the data) and test (1/3 of the data) data subsets, (3) the above-mentioned negative control condition, and (4) the use of 6 different classifiers to avoid reliance of the analysis on a single method in which occasionally overfitting may have occurred. The addition of further classifiers also considered a possible bias in the features selection that relied on random forests. In addition, the genotype vs phenotype associations were repeated using the feature set that had been obtained before random forests followed by ABC analysis had been performed during feature selection.

Table 2

Genetic variants (d = 21) that were included in the genotype–phenotype associations following the feature selection step of the data analysis (Fig. 1).

Gene	Variant	DNA change	Molecular consequence	dbSNP ID
DRD1	X5.174870196.SNV	C>T	UTR5 variant	rs5326
FAAH	X1.46879562.SNV	T>G	UTR3 variant	rs2295632
FAAH	X1.46870761.SNV	C>A	Missense variant	rs324420
GCH1	X14.55310003.SNV	T>C	UTR3 variant	rs45454691
GPR132	X14.105516226.SNV	C>G	UTR3 variant	rs12890396
OPRM1	X6.154449850.SNV	C>A	UTR3 variant	rs613355
DRD3	X3.113890815.SNV	C>T	Missense variant	rs6280
RELN	X7.103136102.SNV	C>A	Intronic variant	rs2528873
GABRA5	X15.27188459.SNV	C>T	Synonymous coding variant	rs140685
GPR132	X14.105532002.SNV	T>C	UTR5 variant	rs3809469
GPR132	X14.105531945.SNV	A>G	UTR5 variant	rs3809470
NF1	X17.29703438.SNV	C>G	UTR3 variant	rs1800845
GPR132	X14.105522201.SNV	C>T	UTR5 variant	rs7157567
COMT	X22.19956781.SNV	G>A	UTR3 variant	rs165599
TRPA1	X8.72938359.SNV	T>C	Intronic variant	rs2305019
TRPA1	X8.72938332.SNV	G>A	Intronic variant	rs2305017
GABRA5	X15.27182656.SNV	A>G	Intronic variant	rs61999613
GPR132	X14.105523663.SNV	G>A	Intronic variant	rs7147439
ABHD6	X3.58270963.SNV	T>C	Intronic variant	rs6774235
ABHD6	X3.58280061.SNV	T>C	UTR3 variant	rs6924
DRD4	X11.640860.Del	delG	UTR3 variant	—

Their potential biological functions were queried from several publicly available databases (NCBI gene index database at http://www.ncbi.nlm.nih.gov/gene; GeneCards at http://www.genecards.org, Short Genetic Variations database (dbSNP) at https://www.ncbi.nlm.nih.gov/snp, and the "1000 Genomes Browser" at https://www.ncbi.nlm.nih.gov/variation/tools/1000genomes; all accessed on June 2018). The order of the variants corresponds to their importance for phenotype association shown in Supplementary Figure 2, http://links.lww.com/PAIN/A795.
Del, nucleotide deletion; Ins, nucleotide insertion; SNV, single-nucleotide variation.

In addition to the machine-learned approach to genotype–phenotype association, a classical χ^2 test analysis was performed on the variants found at allelic frequencies >10%, without any further feature selection.

3. Results

Next-generation sequencing data were obtained from n = 140 patients equally distributed across phenotype classes of "persistent pain" or "nonpersistent pain." The genetic information initially comprised d = 4748 loci for which in at least one subject, an allele was observed that differed from the hg19 reference genomic sequence and for which the allelic distribution corresponded to the expectation from the Hardy–Weinberg equilibrium (the Fisher exact tests: $P > 0.05$). Most of the variants were rare, selecting only those found at allelic frequencies ≥10% provided d = 520 variants. Following feature selection based on the Shannon information criterion, d = 249 variants passed, and following application of the criterion based on χ^2 statistics for phenotype group differences, d = 86 genetic features remained in the data set. The final step of feature selection reduced the set further to d = 21 variants (**Fig. 2**). The selected variants covered 13 different genes and had different putative molecular functional consequences (**Table 2**). The frequencies of the minor alleles ranged between 19.3% and 40%, with a median of 31%. Specifically, variants in the ABC set "A" belonged, in descending order of importance (Supplementary Fig. 2, available at http://links.lww.com/PAIN/A795), to the genes *DRD1*, *FAAH*, *GCH1*, *GPR132*, *OPRM1*, *DRD3*, *RELN*, *GABRA5*, *NF1*, *COMT*, *TRPA1*, *ABHD6*, and *DRD4* (for gene names, see **Table 3**).

Unsupervised machine learning, aiming at data structure detection, was applied to analyze the 140 × 21-sized matrix comprising d = 21 genetic variants acquired in n = 140 patients. Training of a swarm of intelligent data bots provided a structure-preserving projection of the high-dimensional data space $D = \{x_{i,d}, i = 1, \ldots, n\} \subset \mathbb{R}^d$ onto a 2-dimensional toroid projection grid (**Fig. 3**). Following addition of the U-matrix, a cluster structure emerged from the separation of the data bots carrying the genetic information into 2 distinct groups as visually indicated by a "mountain range" on the topographic map analogy (**Fig. 3** top). This was verified by Ward clustering that indicated 2 clusters differing with respect to the pattern of genetic variants (**Fig. 3**). Finally, the cluster membership was found to be unequally distributed among the phenotypes (the Fisher exact test: $P < 0.05$). This supported further exploration of the genetic information for relevance for the phenotypic classification.

Supervised machine learning, applied in cross-validation experiments using 1000 Monte-Carlo random resamplings of 2/3 vs (new training subset) 1/3 (new test) of the data (Supplementary Fig. 1, available at http://links.lww.com/PAIN/A795), provided the consistent observation that when using the true NGS genotypes, the phenotype class assignment was better than guessing (**Table 4**). This is indicated by the fact that values for accuracy exceeded 50%, which would be zeroR expectation

Table 3

Genes included in the NGS panel of persistent pain.

Gene symbol	NCBI	Gene description	Ref.
ABHD12	26090	Abhydrolase domain containing 12	60
ABHD16A	7920	Abhydrolase domain containing 16A	60
ABHD6	57406	Abhydrolase domain containing 6	60
ADCY1	107	Adenylate cyclase 1	147
ADRB2	154	Adrenoceptor beta 2	47
BDNF	627	Brain-derived neurotrophic factor	106
CACNG2	10369	Calcium voltage-gated channel auxiliary subunit gamma 2	105
CDK5	1020	Cyclin-dependent kinase 5	157
CHRNB2	1141	Cholinergic receptor, nicotinic, beta 2	26
CNR1	1268	Cannabinoid receptor 1 (brain)	131
COMT	1312	Catechol-O-methyltransferase	29
CSF1	1435	Colony-stimulating factor 1	142
DLG4	1742	Discs, large homolog 4 (Drosophila)	31
DRD1	1812	Dopamine receptor D1	108
DRD2	1813	Dopamine receptor D2	108
DRD3	1814	Dopamine receptor D3	112
DRD4	1815	Dopamine receptor D4	11
EGR1	1958	Early growth response 1	62
ESR1	2099	Extrogen receptor 1	116
FAAH	2166	Fatty acid amide hydrolase	53
FKBP5	2289	Fk506-binding protein 5	33
FOS	2353	Cellular oncogene FOS	1
FYN	2534	Src family tyrosine kinase	76
GABRA5	2558	GABA A receptor, alpha 5	6
GALR2	8811	Galanin receptor 2	50
GCH1	2643	GTP cyclohydrolase 1	139
GDNF	2668	Glial cell–derived neurotrophic factor	120
GFRA1	2674	GDNF family receptor alpha 1	156
GPR132	29933	G-protein-coupled receptor 132	48
GRIN1	2902	Glutamate receptor, NMDA 1	111
GRIN2A	2903	Glutamate receptor, NMDA 2A	111
GRIN2B	2904	Glutamate receptor, NMDA 2B	111
GRM5	2915	Glutamate receptor, metabotropic 5	150
HCN2	610	Hyperpolarization-activated cyclic nucleotide-gated	145
HLA-DQB1	3119	Major histocompatibility complex, class II, DQ beta 1	27
HLA-DRB1	3123	Major histocompatibility complex, class II, DR beta 1	27
HRH3	11255	Histamine receptor H3	49
HTR1A	3350	5-hydroxytryptamine (serotonin) receptor 1A	75
HTR2A	3356	5-hydroxytryptamine (serotonin) receptor 2A	104
IL10	3586	Interleukin 10	134
IL1B	3553	Interleukin 1B	78
IL1R2	7850	Interleukin 1 receptor type 2	134
IL4	3565	Interleukin 4	136
IL6	3569	Interleukin 6	127
KCNS1	3787	Potassium voltage-gated channel, modifier subfamily S, member 1	19
KIT	3815	Tyrosine kinase KIT	137
LTB4R	1241	Leukotriene b4 receptor	65
LTB4R2	56413	Leukotriene b4 receptor 2	65
NF1	4763	Neurofibromin 1	155
NGF	4803	Nerve growth factor	66
NTF4	4909	Neurotrophin 4	66
NTRK1	4914	Neurotrophic tyrosine kinase 1	66
OPRD1	4985	Opioid receptor delta 1	71
OPRK1	4986	Opioid receptor kappa 1	40
OPRM1	4988	Opioid receptor mu 1	79
OXT	5020	Oxytocin prepropeptide	38
P2RX7	5027	Purinergic receptor P2X7	133
PLCB1	23236	Phospholipase C, beta 1	126
PRKCG	5582	Protein kinase C, gamma	130
PRNP	5621	Prion protein	34
PTN	5764	Pleiotrophin	39
PTPRZ1	5803	Protein tyrosine phosphatase Z 1	
RELN	5649	Reelin	10
RET	5979	RET receptor tyrosine kinase	132
RUNX1	861	Runt-related transcription factor 1	16
S100B	6285	S100 calcium-binding protein B	158
SCN9A	6335	Sodium voltage-gated alpha subunit 9	115
SLC6A4	6532	Serotonin transporter	107
SOD2	6648	Superoxide dismutase 2	124
TH	7054	Tyrosine hydroxylase	7
TLR4	7099	Toll-like receptor 4	51
TNF	7124	Tumor necrosis factor	73
TRPA1	8989	Transient receptor potential cation channel, subfamily A, member 1	5
TRPM8	79054	Transient receptor potential cation channel, subfamily M, member 8	5
TRPV1	7442	Transient receptor potential cation channel, subfamily V, member 1	5
TRPV4	59341	Transient receptor potential cation channel, subfamily V, member 4	5
TSPO	706	Translocator protein	77

The selection of the genes has been described previously in more detail,[64] including evidence for the involvement of the respective gene in persistent pain.
NGS, next-generation sequencing; NMDA, *N*-methyl-D-aspartate.

(Supplementary Fig. 3, available at http://links.lww.com/PAIN/A795). Specifically, accuracies obtained with random forests, adaptive boosting, kNN, naive Bayes, support vector machine, and regression-based classifiers were 71.7, 71.7, 65.2, 67.4, 71.7, and 65.2, respectively; the values of the areas under the receiver operator characteristic were similar as those of the accuracies (**Table 4**). The above-mentioned obtained balanced accuracies exceeded the balanced accuracies that were obtained when training the algorithms with d = 86 variants that

had been identified before the last step of feature selection (67.4%, 67.4%, 64.1%, 65.2%, 67.4%, and 56.5%, respectively), emphasizing the utility of a dimension reduction before training machine-learning algorithms. By contrast, class assignment using the permuted, and therefore meaningless, genotype information used as the negative control was 50%, thus, corresponding to chance. The best median classification accuracy with the true genotypes was obtained with random forests, adaptive boosting, and support vector machines (**Table 4**). The use of random forests already in feature selection did not produce a bias toward this particular algorithm, as indicated by the similar classification performance of all algorithms. This was true either with the set of 21 variants included in the final result or with the set of 86 variants that had been obtained during feature selection before random forests had been included.

Finally, a classical χ^2 test–based genotype vs phenotype association was performed for comparison. Using the d = 520 variants observed at allelic frequencies ≥10% resulted in 24 *P* values <0.05 but, only at the uncorrected α-level (**Fig. 4**). However, when corrected according to Bonferroni,[4] the resulting α level of 9.61×10^{-5} for the d = 520 genetic variants was exceeded for all gene loci.

4. Discussion

The analyses showed that selected genotypes, obtained by NGS of coding and regulatory parts of genes reported to be involved in the modulation of pain, are able to correctly assign a patient phenotype to either pain persistence after breast cancer surgery or not. This can be concluded from the observations that (1) machine-learned algorithms trained with the true genetic information were able to assign an individual to the correct pain phenotype group better than just guessing, and that (2) algorithms trained with the random genetic information were unable to perform this task while not differing from performance obtained by guessing. Thus, the main result of the preset analysis is the unequivocal establishment of a genetic background of pain persistence after breast cancer surgery. Moreover, the present analysis narrowed the focus from initially 77 candidate genes to finally 13 genes in which variants were associated with group differences in the development of pain over the 3 years after breast cancer surgery.

An interesting result of the present analysis is that the highest-ranked variants are located in genes involved in dopaminergic pathways, including the receptor genes *DRD1*, *DRD3*, *DRD4*, and the tyrosine hydroxylase gene *TH*, which codes for a rate-limiting enzyme in dopamine synthesis. So far, variants in these genes have been implicated in promoting chronic back pain.[42,43,52,91] Their implication in pain after breast cancer surgery is novel. Variations in genes operating in the catecholaminergic and serotonergic pathways (*COMT*, *HTR2A*, *HTR3A*, *SLC6A2*, and *SCL6A3*) are logical candidates that may contribute to the development and severity of persistent pain after breast surgery. Catecholamine and serotonin serve as principal neurotransmitters and play important roles in both the peripheral and central mechanisms of pain.[61] These neurotransmitters activate peripheral nociceptors during tissue injury and contribute to the development of central sensitization.[96] Dysfunction in both catecholamine and serotonin neurotransmission is implicated in the development of persistent pain syndromes.[3]

Several of the highest-ranking variants (Supplementary Fig. 2, available at http://links.lww.com/PAIN/A795) are located in the G-protein-coupled receptor 132 (G2A) gene (*GPR132*). The G2A receptor is an antiproliferative cell cycle regulator that can be induced by several different stimuli, including different classes of DNA-damaging agents such as hydroxyurea, 5′-fluorouracil, cytosine arabinoside, etoposide, taxol, or doxorubicin.[153] Of note, most of the present patients had received docetaxel in addition to 5-fluorouracil, epirubicin, and cyclophosphamide.[129] The G2A receptor is implicated in chemotherapy-induced peripheral neuropathic pain, as its activation sensitized the ligand-gated ion-channel TRPV1 in sensory neurons through activation of protein kinase C.[48] The highest-ranked *GPR132* SNP in the present association analysis was rs12890396, which is a micro-RNA target site that has been implicated among risk factors for endometriosis.

GABAergic mechanisms play a role in the mediation and perception of pain, and at least some types of persistent pain have been associated with a decline in GABAergic tone.[93] At the cellular level, opioids and cannabinoids are thought to activate descending analgesic pathways indirectly by suppressing inhibitory GABAergic inputs into output neurons of these descending pathways.[70] In addition to these roles of GABRA5 in pain, the receptor has been implicated in the progression of breast cancer where *GABRA5* mutations have been found in the primary tumor and first metastases.[97] Within the candidate gene approach, the appearance of *COMT* among the genes with the most relevant variants for the association to the present persistent pain phenotype groups provides support to the suggestion[23] that it is a key genetic modulator of pain in the average population. Interestingly, other *COMT* polymorphisms increasing heat or cold pain sensitivity before breast cancer surgery have already been detected in the same cohort of women as previously reported.[55] Similarly, *GCH1* variants have been implicated in protecting against pain. A pain-protective haplotype has been reported, which is composed of 15 single-nucleotide polymorphisms in the *GCH1* gene.[139] Presently, the X14.55310003.SNV corresponding to the database listed variant rs45454691 was among the highest-ranked modulators of the risk of persistent pain. This variant is located in the 3′ downstream region of the gene. It was more frequent in the "persistent pain" group (Supplementary Fig. 1, available at http://links.lww.com/PAIN/A795). However, it is not included in the "pain-protective" *GCH1* haplotype, mentioned above, and therefore, its higher frequency among patients with pain does not contradict prior knowledge.

Reelin is a glycoprotein that modulates *N*-methyl-D-aspartate receptor function[17] and plays a role in sensory processing.[101] RELN mutations have been associated with mechanical and thermal hypersensitivity.[149] A direct link to breast cancer may be based on the role of reelin in the control of cell migration and tumor invasiveness.[15] In addition, a role of neuroinflammation in persistent pain is increasingly recognized. For example, one of the sites of interaction of the immune system with chronic pain has been identified as neuroimmune crosstalk at the glial–opioid interface.[143] In addition, *FAAH*[95] and *GCH1*[69] also play a role in pain-related inflammatory processes. *FAAH* variants have been identified in an independent analysis of 9 different pain phenotypes in the same cohort,[12] as well as *P2RX7*.[56] Moreover, a link between FAAH and breast cancer is supported by reports that cannabinoid receptor agonists inhibited tumor growth and metastasis of breast cancer.[113]

Considering the role of the μ-opioid receptor gene *OPRM1* in pain, through both its effect on pain sensitivity and opioid analgesia, it was not surprising that it was among the highest-ranking variants. The 118A>G single-nucleotide polymorphism (*OPRM1* rs1799971) is of great scientific interest. It modulates both the expression and signaling efficiency of the μ-opioid

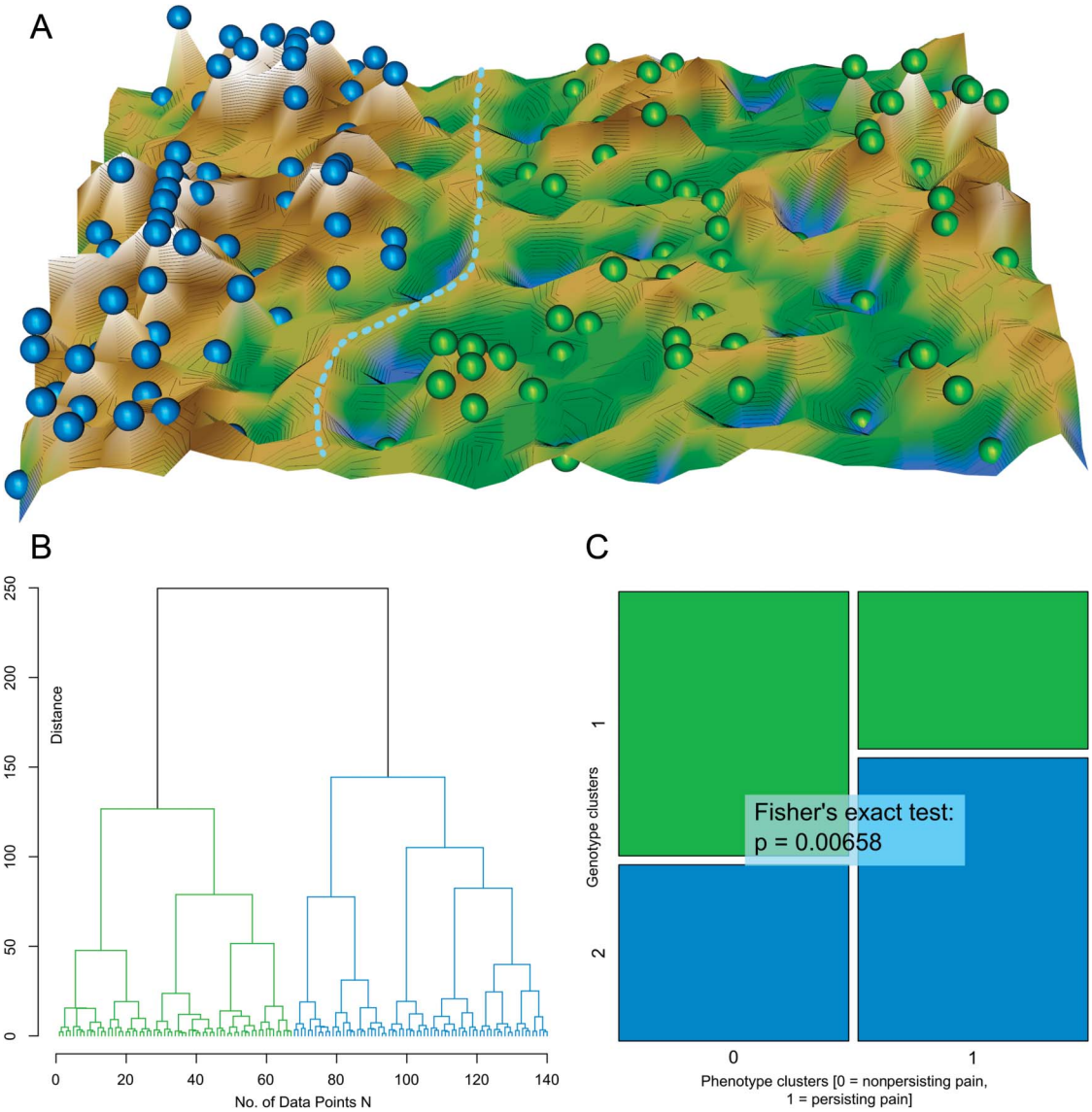


Figure 3. Data structure found in the NGS genotypes and its relation with the phenotypes: (A) Visualization of high-dimensional data consisting of d = 21 variants in 13 different genes analyzed in n = 140 subjects. The data were projected onto a 2-dimensional grid using a parameter-free projection polar swarm, *Pswarm*.[141] During the learning phase, the DataBots were allowed adaptively adjusting their location on the grid close to DataBots carrying data with similar features, with successively decreasing search radius. When the algorithm ends, the DataBots become projected points. To enhance the emergence of data structures on this projection, a generalized U-matrix displaying the distance in the high-dimensional space was added as a third dimension to this visualization. The U-matrix was colored to appear as a geographical map with brown (up to snow-covered) heights and green valleys with blue lakes. Watersheds indicate borderlines between different groups of subjects according to the pattern of repeated cold pain measurements. The data points are colored according to the emerging 2-cluster structure. (B) Ward clustering of the projected data clearly indicated 2 clusters. (C) Subjects belonging to the different genotype clusters were unevenly distributed across the pain phenotype groups, ie, assignment to either the nonpersisting pain or the persisting pain group, at a statistical significance level of $P < 0.05$ (the Fisher exact test). The mosaic plots represent the contingency table of the genotype vs phenotype class structure (membership size). The figure has been created using the R software package (version 3.4.4 for Linux; http://CRAN.R-project.org/),[114] in particular the libraries "DatabionicSwarm" (https://cran.r-project.org/package=DatabionicSwarm)[141] and "gplots" (Warnes et al., https://cran.r-project.org/package=gplots). NGS, next-generation sequencing.

receptor. However, in the present cohort, rs1799971 was not associated with persistent pain, although an earlier analysis[13] had associated rs1799971 with postoperative oxycodone demands and several other studies with the demands of other opioids. On the other hand, *OPRM1* rs613355 was among the most relevant genetic variants. This variant seems to be an expression quantitative trait locus (eQTL), defined as genomic loci that explain all or a fraction of variation in expression levels of mRNAs.[118] Specifically, rs613355 is listed as located in a cis-eQTL locus for *OPRM1*,[44] a locus that maps to the approximate location of their gene-of-origin. Its association with *OPRM1* expression was statistically significant at $P = 0.01$.

In the light of the limitations of the present data, the accuracy of 71% for the prediction of persistent pain seems remarkable. First, the cohort was small, owing to the <10% of patients who developed persistent pain among the originally enrolled 1000 patients,[98] and replication of these findings is needed in an independent and larger sample. Second, this study included replications using data splits and resampling techniques in cross-validation experiments. However, a completely independent control cohort was not available. Nevertheless, some of the candidate genes were taken from positive reports of an association of variants with pain-related phenotypes, making the present analysis, in part, a replication study of earlier

Table 4

Performance of classifiers obtained using different machine-learned methods (random forests, adaptive boosting, k-nearest neighbors [kNN], naive Bayes, and support vector machines [SVM]) that were trained with (1) the original genetic information and (2) using training data subsets constructed to provide a negative control by permuting the original genotypes.

Parameter	Random forests			Boosting			kNN			Naive Bayes			SVM			Regression		
(%)	Original data (d = 21 variants)	Original data (d = 86 variants)	Permuted training data	Original data (d = 21 variants)	Original data (d = 86 variants)	Permuted training data	Original data (d = 21 variants)	Original data (d = 86 variants)	Permuted training data	Original data (d = 21 variants)	Original data (d = 86 variants)	Permuted training data	Original data (d = 21 variants)	Original data (d = 86 variants)	Permuted training data	Original data (d = 21 variants)	Original data (d = 86 variants)	Permuted training data
Sensitivity, recall	73.9	65.2	47.8	69.6	69.6	47.8	60.9	69.6	52.2	69.6	60.9	47.8	69.6	69.6	47.8	56.5	56.5	50
Specificity	73.9	69.6	47.8	73.9	69.6	47.8	69.6	56.5	52.2	69.6	65.2	52.2	73.9	65.2	52.2	73.9	56.5	52.2
Positive predictive value, precision	72	66.7	50	71.4	68.4	50	66.7	62.5	50	68.2	65.2	50	71.4	66.7	50	66.7	55.2	50
Negative predictive value	72.7	66.7	50	70.8	68.2	50	64.3	66.7	50	68.2	64	50	70.8	66.7	50	62.3	55	50
F1	72.3	66.7	50	70.8	68.1	50	63.9	66.7	51.1	68.2	63.6	50	71.1	66.7	50	60.9	54.5	50
Accuracy	71.7	67.4	50	71.7	67.4	50	65.2	64.1	50	67.4	65.2	50	71.7	67.4	50	65.2	54.3	50
Area under the ROC curve	71.7	67.4	50	71.7	67.4	50	65.2	64.1	50	67.4	65.2	50	71.7	67.4	50	65.2	56.5	50

Results represent the medians of the test performance measures from 1000 model runs using Monte-Carlo resampling with splits into 2/3 of the data (training data subset) and 1/3 (test data subset).
ROC, receiver operating characteristic.

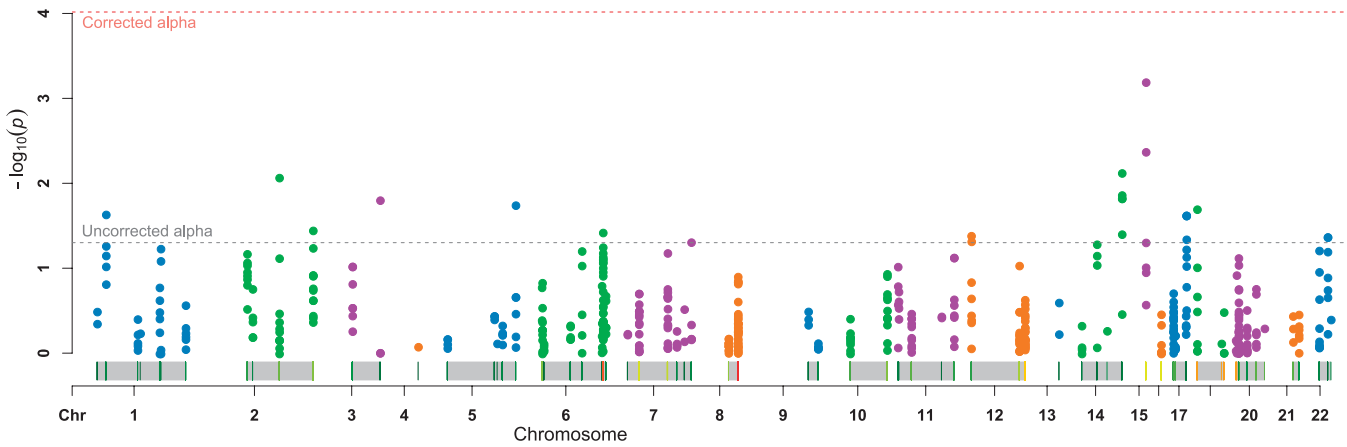


Figure 4. Dotplot (Manhattan plot) of the results of the χ^2 test–based genotype vs phenotype association tests for d = 520 gene loci in 77 genes (Table 2) found at allelic frequencies ≥ 10% in the n = 140 patients after breast cancer surgery. The α levels, without (red) and with (gray/green) correction for multiple testing according to Bonferroni,[4] are indicated as dashed lines. The figure was created using the R software package (version 3.4.4 for Linux; http://CRAN.R-project.org/)[114] and the package "CMplot" (Lilin Yin; https://github.com/YinLiLin/R-CMplot).

evidence. Moreover, the fact that 75% of these genes are targets of approved or novel analgesics under current clinical development supports the relevance of the findings. Of the 13 genes in the final set identified in the present analysis, 10 (77%) figure among the targets of analgesics, namely *ABHD6*, *COMT*, *DRD1*, *DRD3*, *DRD4*, *FAAH*, *GABRA5*, *OPRM1*, *RELN*, and *TRPA1*. This suggests that the present set of gene variants may pave the way for future efforts to establish personalized therapy in persistent pain after breast cancer surgery, when some of the drugs under development become clinically available. Third, although this study used a candidate gene approach limited to 77 genes due to technical limitations of the AmpliSeq panel, it exceeded classical candidate gene approaches by including a genome-wide selection of genes and, importantly, by including the complete sequence of exomic and regulatory parts of the selected genes.

Using a case–control approach in which we compared a subgroup of patients with persistent pain to a similarly sized sample of patients who did not have persistent pain has some limitations. Specifically, it is possible that certain identified variants would also have been seen in the unsampled members of the nonpersistent pain subgroup. That the present results are an effect of this sampling may be contradicted by the reproduction of variants that have been assessed previously using the full cohort while not using NGS.[12,55] With NGS, the financial effort had to be limited, and therefore, the present sample was taken to compare the genotypes of the subjects with the extreme pain phenotypes of interest, while intermediate phenotypes were omitted. This is a standard design that has several variants such as using matched pairs or similar designs. Therefore, the numerical values of the pain phenotype group association accuracy may require revision when applied to a nonselected cohort. Hence, the reported classification performances of the algorithms are not presented as diagnostic tools but have been used in a knowledge-discovery manner aimed at identifying genetic variants distinctive between extreme pain phenotypes after breast cancer surgery. A further limitation of this study is the omission of rare variants found at allelic frequencies <10%. Indeed, it is possible that the development of persistent pain is regulated by some of the omitted variants and moreover, in each patient by another variant. This would require another analysis such as proposed elsewhere.[138] However, the present already complex analysis focused on the group-level differences and was therefore not yet further extended.

The intensity of pain is the traditional target of research and therapy concepts. However, it becomes clearer that this may underestimate the complexity of pain and reflects its clinical facets only partly. The present group definition may therefore reflect the clinically relevant phenotype only partially. Furthermore, the present analysis focused on genetic variants while ignoring other modulators of persistent pain. This partly owes to the conceptualization as a genetic association study and partly to the intention to observe nonredundancy to previous analyses of the same cohort. We know from the previously analyses of this cohort that demographic, psychological, and early pain-related parameters are predictive for the development of persistent pain or the lack of such development.[84,85,89] Rather than adding these parameters again, it would be interesting in the future to combine several different classifiers to study whether this increases the overall predictive value.

Finally, this study still used a candidate approach because of technical and project limitations, and the sample size was also limited. While providing a hierarchy of candidate genes and variants in the present clinical context, replication in larger studies and in studies with unrestricted inclusion of all pain-relevant genes or the whole genome could be based on the present positive results.

5. Conclusions

Using NGS-derived genetic information on the coding and regulatory sequences of 77 candidate genes, machine-learned analysis indicated that the genotypes provide useful information for the allocation of the patients to either a "persistent pain" or "nonpersistent pain" phenotype group in a 3-year follow-up after breast cancer surgery. Although candidate genes had been selected on the basis of previous reports in persistent pain, the present results provided a ranking of genes with respect to their importance in the clinical context. A particular association with breast cancer could be observed for most of the highest-ranking genes. These findings strengthen the specificity of the present results for the clinical setting of breast cancer surgery. The involvement of the genes in breast cancer applies to the whole cohort, while the present analysis identified differences with

respect to the development of pain. Thus, in addition to modulation of the clinical course of breast cancer, specific modulation of pain is provided by certain variants. For example, an important role of the *DRD1* and *GPR132* genes, or of a variant located in a cis-eCTL position of the *OPRM1* gene, in the clinical context of pain after breast cancer surgery, was highlighted as a novel finding.

Conflict of interest statement

The authors have no conflicts of interest to declare.

Acknowledgement

The authors thank Professor Mike Parnham for native English proofreading the manuscript.
This work was supported by the European Union Seventh Framework Programme (FP7/2007—2013) under grant agreement no. 602919 (GLORIA, E.K., J.L.), by the Academy of Finland (E.K.) and by the Helsinki University Hospital Governmental Research funds (TYH2008225, TYH2010210, E.K.). Additional support, in particular for the NGS laboratory equipment, was received from the Landesoffensive zur Entwicklung wissenschaftlich-ökonomischer Exzellenz (LOEWE), Centre for Translational Medicine and Pharmacology (J.L.). The funders had no role in method design, data selection and analysis, decision to publish, or preparation of the manuscript.
Data from this study were analyzed in a nonredundant manner as published in Breast Cancer Res Treat. 2018. doi: 10.1007/s10549-018-4841-8, Br J Anaesth. 2018 Nov;121(5):1123-1132, Br J Anaesth. 2017 Oct 1;119(4):821-829, and J Clin Oncol 2017; 35: 1660-1667.

Appendix A. Supplemental digital content

Supplemental digital content associated with this article can be found online at http://links.lww.com/PAIN/A795.

Article history:
Received 19 December 2018
Received in revised form 29 April 2019
Accepted 3 May 2019
Available online 15 May 2019

References

[1] Abbadie C, Besson JM, Calvino B. c-Fos expression in the spinal cord and pain-related symptoms induced by chronic arthritis in the rat are prevented by pretreatment with Freund adjuvant. J Neurosci 1994;14:5865–71.
[2] Bayes M, Price M. An essay towards solving a problem in the doctrine of chances. By the late Rev. Mr. Bayes, F. R. S. Communicated by Mr. Price, in a Letter to John Canton, A. M. F. R. S. Phil Trans 1763;53:370–418.
[3] Becker S, Schweinhardt P. Dysfunctional neurotransmitter systems in fibromyalgia, their role in central stress circuitry and pharmacological actions on these systems. Pain Res Treat 2012;2012:741746.
[4] Bonferroni CE. Teoria statistica delle classi e calcolo delle probabilita. Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commerciali di Firenze 1936;8:3–62.
[5] Bourinet E, Altier C, Hildebrand ME, Trang T, Salter MW, Zamponi GW. Calcium-permeable ion channels in pain signaling. Physiol Rev 2014;94:81–140.
[6] Bravo-Hernández M, Corleto JA, Barragán-Iglesias P, González-Ramírez R, Pineda-Farias JB, Felix R, Calcutt NA, Delgado-Lezama R, Marsala M, Granados-Soto V. The α5 subunit containing GABAA receptors contribute to chronic pain. PAIN 2016;157:613–26.
[7] Bravo L, Torres-Sanchez S, Alba-Delgado C, Mico JA, Berrocoso E. Pain exacerbates chronic mild stress-induced changes in noradrenergic transmission in rats. Eur Neuropsychopharmacol 2014;24:996–1003.
[8] Breiman L. Random forests. Mach Learn 2001;45:5–32.
[9] Breivik H, Collett B, Ventafridda V, Cohen R, Gallacher D. Survey of chronic pain in Europe: prevalence, impact on daily life, and treatment. Eur J Pain 2006;10:287–333.
[10] Buchheit T, Van de Ven T, Shaw A. Epigenetics and the transition from acute to chronic pain. Pain Med 2012;13:1474–90.
[11] Buskila D, Cohen H, Neumann L, Ebstein RP. An association between fibromyalgia and the dopamine D4 receptor exon III repeat polymorphism and relationship to novelty seeking personality traits. Mol Psychiatry 2004;9:730–1.
[12] Cajanus K, Holmström EJ, Wessman M, Anttila V, Kaunisto MA, Kalso E. Effect of endocannabinoid degradation on pain: role of FAAH polymorphisms in experimental and postoperative pain in women treated for breast cancer. PAIN 2016;157:361–9.
[13] Cajanus K, Kaunisto MA, Tallgren M, Jokela R, Kalso E. How much oxycodone is needed for adequate analgesia after breast cancer surgery: effect of the OPRM1 118A>G polymorphism. J Pain 2014;15:1248–56.
[14] Carpenter FC Jr. ECG telemetry within a small, closed chamber. Aerosp Med 1970;41:402–6.
[15] Castellano E, Molina-Arcas M, Krygowska AA, East P, Warne P, Nicol A, Downward J. RAS signalling through PI3-Kinase controls cell migration via modulation of Reelin expression. Nat Commun 2016;7:11245.
[16] Chen CL, Broom DC, Liu Y, de Nooij JC, Li Z, Cen C, Samad OA, Jessell TM, Woolf CJ, Ma Q. Runx1 determines nociceptive sensory neuron phenotype and is required for thermal and neuropathic pain. Neuron 2006;49:365–77.
[17] Chen Y, Beffert U, Ertunc M, Tang TS, Kavalali ET, Bezprozvanny I, Herz J. Reelin modulates NMDA receptor activity in cortical neurons. J Neurosci 2005;25:8209–16.
[18] Cortes C, Vapnik V. Support-vector networks. Mach Learn 1995;20:273–97.
[19] Costigan M, Belfer I, Griffin RS, Dai F, Barrett LB, Coppola G, Wu T, Kiselycznyk C, Poddar M, Lu Y, Diatchenko L, Smith S, Cobos EJ, Zaykin D, Allchorne A, Gershon E, Livneh J, Shen PH, Nikolajsen L, Karppinen J, Mannikko M, Kelempisioti A, Goldman D, Maixner W, Geschwind DH, Max MB, Seltzer Z, Woolf CJ. Multiple chronic pain states are associated with a common amino acid-changing allele in KCNS1. Brain 2010;133:2519–27.
[20] Cover T, Hart P. Nearest neighbor pattern classification. IEEE Trans Inf Theor 1967;13:21–7.
[21] Culp M, Johnson K, Michailides G. ada: an R package for stochastic boosting. J Stat Softw 2006;17:27.
[22] de Leng WW, Gadellaa-van Hooijdonk CG, Barendregt-Smouter FA, Koudijs MJ, Nijman I, Hinrichs JW, Cuppen E, van Lieshout S, Loberg RD, de Jonge M, Voest EE, de Weger RA, Steeghs N, Langenberg MH, Sleijfer S, Willems SM, Lolkema MP. Targeted next generation sequencing as a reliable diagnostic assay for the detection of somatic mutations in tumours using minimal DNA amounts from formalin fixed paraffin embedded material. PLoS One 2016;11:e0149405.
[23] Diatchenko L, Nackley AG, Slade GD, Bhalang K, Belfer I, Max MB, Goldman D, Maixner W. Catechol-O-methyltransferase gene polymorphisms are associated with multiple pain-evoking stimuli. PAIN 2006;125:216–24.
[24] Diatchenko L, Nackley AG, Tchivileva IE, Shabalina SA, Maixner W. Genetic architecture of human pain perception. Trends Genet 2007;23:605–13.
[25] Diatchenko L, Slade GD, Nackley AG, Bhalang K, Sigurdsson A, Belfer I, Goldman D, Xu K, Shabalina SA, Shagin D, Max MB, Makarov SS, Maixner W. Genetic basis for individual variations in pain perception and the development of a chronic pain condition. Hum Mol Genet 2005;14:135–43.
[26] Dineley KT, Pandya AA, Yakel JL. Nicotinic ACh receptors as therapeutic targets in CNS disorders. Trends Pharmacol Sci 2015;36:96–108.
[27] Dominguez CA, Kalliomaki M, Gunnarsson U, Moen A, Sandblom G, Kockum I, Lavant E, Olsson T, Nyberg F, Rygh LJ, Roe C, Gjerstad J, Gordh T, Piehl F. The DQB1 *03:02 HLA haplotype is associated with increased risk of chronic pain after inguinal hernia surgery and lumbar disc herniation. PAIN 2013;154:427–33.
[28] Elliott AM, Smith BH, Penny KI, Smith WC, Chambers WA. The epidemiology of chronic pain in the community. Lancet 1999;354:1248–52.
[29] Feng Y, Zhao X, Zhou C, Yang L, Liu Y, Bian C, Gou J, Lin X, Wang Z, Zhao X. The associations between the Val158Met in the catechol-O-

methyltransferase (COMT) gene and the risk of uterine leiomyoma (ULM). Gene 2013;529:296–9.
[30] Fisher RA. On the interpretation of chi square from contingency tables, and the calculation of P. J R Stat Soc 1922;85:87–94.
[31] Florio SK, Loh C, Huang SM, Iwamaye AE, Kitto KF, Fowler KW, Treiberg JA, Hayflick JS, Walker JM, Fairbanks CA, Lai Y. Disruption of nNOS-PSD95 protein-protein interaction inhibits acute thermal hyperalgesia and chronic mechanical allodynia in rodents. Br J Pharmacol 2009;158: 494–506.
[32] Frank M, Prenzler A, Eils R, Graf von der Schulenburg JM. Genome sequencing: a systematic review of health economic evidence. Health Econ Rev 2013;3:29.
[33] Fujii T, Ota M, Hori H, Hattori K, Teraishi T, Matsuo J, Kinoshita Y, Ishida I, Nagashima A, Kunugi H. The common functional FKBP5 variant rs1360780 is associated with altered cognitive function in aged individuals. Sci Rep 2014;4:6696.
[34] Gadotti VM, Zamponi GW. Cellular prion protein protects from inflammatory and neuropathic pain. Mol Pain 2011;7:59.
[35] Gagliese L, Weizblit N, Ellis W, Chan VW. The measurement of postoperative pain: a comparison of intensity scales in younger and older surgical patients. PAIN 2005;117:412–20.
[36] Godinova AM. Genetic analysis of migraine [in Russian]. Zh Nevropatol Psikhiatr Im S S Korsakova 1965;65:1132–8.
[37] Good PI. Resampling methods: a practical guide to data analysis. Boston: Birkhäuser, 2006.
[38] Goodin BR, Ness TJ, Robbins MT. Oxytocin—a multifunctional analgesic for chronic deep tissue pain. Curr Pharm Des 2015;21: 906–13.
[39] Gramage E, Herradon G. Genetic deletion of pleiotrophin leads to disruption of spinal nociceptive transmission: evidence for pleiotrophin modulation of morphine-induced analgesia. Eur J Pharmacol 2010;647: 97–102.
[40] Guerrero M, Urbano M, Brown SJ, Cayanan C, Ferguson J, Cameron M, Devi LA, Roberts E, Rosen H. Optimization and characterization of an opioid kappa receptor (OPRK1) antagonist. Probe reports from the NIH molecular libraries program. Bethesda: National Center for Biotechnology Information (US), 2010.
[41] Guyon I, Elisseeff A. An introduction to variable and feature selection. J Mach Learn Res 2003;3:1157–82.
[42] Hagelberg N, Forssell H, Aalto S, Rinne JO, Scheinin H, Taiminen T, Nagren K, Eskola O, Jaaskelainen SK. Altered dopamine D2 receptor binding in atypical facial pain. PAIN 2003;106:43–8.
[43] Hagelberg N, Jaaskelainen SK, Martikainen IK, Mansikka H, Forssell H, Scheinin H, Hietala J, Pertovaara A. Striatal dopamine D2 receptors in modulation of pain in humans: a review. Eur J Pharmacol 2004;500: 187–92.
[44] Hancock DB, Levy JL, Gaddis NC, Glasheen C, Saccone NL, Page GP, Hulse G, Wildenauer D, Kelty E, Schwab S, Degenhardt L, Martin NG, Montgomery GW, Attia J, Holliday EG, McEvoy M, Scott RJ, Bierut LJ, Nelson EC, Kral A, Johnson EO. Cis-expression quantitative trait loci mapping reveals replicable associations with heroin addiction in OPRM1. Biol Psychiatry 2015;78:474–84.
[45] Hardy GH. Mendelian proportions in a mixed population. Science 1908; 28:49–50.
[46] Ho TK. Random decision forests. Proceedings of the Third International Conference on Document Analysis and Recognition—Volume 1: IEEE Computer Society, 1995. p. 278.
[47] Hocking LJ, Smith BH, Jones GT, Reid DM, Strachan DP, Macfarlane GJ. Genetic variation in the beta2-adrenergic receptor but not catecholamine-O-methyltransferase predisposes to chronic pain: results from the 1958 British Birth Cohort Study. PAIN 2010;149: 143–51.
[48] Hohmann SW, Angioni C, Tunaru S, Lee S, Woolf CJ, Offermanns S, Geisslinger G, Scholich K, Sisignano M. The G2A receptor (GPR132) contributes to oxaliplatin-induced mechanical pain hypersensitivity. Sci Rep 2017;7:446.
[49] Huang L, Adachi N, Nagaro T, Liu K, Arai T. Histaminergic involvement in neuropathic pain produced by partial ligation of the sciatic nerve in rats. Reg Anesth Pain Med 2007;32:124–9.
[50] Hulse RP, Donaldson LF, Wynick D. Peripheral galanin receptor 2 as a target for the modulation of pain. Pain Res Treat 2012;2012:545386.
[51] Hutchinson MR, Zhang Y, Shridhar M, Evans JH, Buchanan MM, Zhao TX, Slivka PF, Coats BD, Rezvani N, Wieseler J, Hughes TS, Landgraf KE, Chan S, Fong S, Phipps S, Falke JJ, Leinwand LA, Maier SF, Yin H, Rice KC, Watkins LR. Evidence that opioids may have toll-like receptor 4 and MD-2 effects. Brain Behav Immun 2010;24:83–95.
[52] Jaaskelainen SK, Lindholm P, Valmunen T, Pesonen U, Taiminen T, Virtanen A, Lamusuo S, Forssell H, Hagelberg N, Hietala J, Pertovaara A. Variation in the dopamine D2 receptor gene plays a key role in human pain and its modulation by transcranial magnetic stimulation. PAIN 2014;155:2180–7.
[53] Jayamanne A, Greenwood R, Mitchell VA, Aslan S, Piomelli D, Vaughan CW. Actions of the FAAH inhibitor URB597 in neuropathic and inflammatory chronic pain models. Br J Pharmacol 2006;147:281–8.
[54] Juran JM. The non-Pareto principle; Mea culpa. Qual Prog 1975;8:8–9.
[55] Kambur O, Kaunisto MA, Tikkanen E, Leal SM, Ripatti S, Kalso EA. Effect of catechol-o-methyltransferase-gene (COMT) variants on experimental and acute postoperative pain in 1,000 women undergoing surgery for breast cancer. Anesthesiology 2013;119: 1422–33.
[56] Kambur O, Kaunisto MA, Winsvold BS, Wilsgaard T, Stubhaug A, Zwart JA, Kalso E, Nielsen CS. Genetic variation in P2RX7 and pain tolerance. PAIN 2018;159:1064–73.
[57] Karatzoglou A, Smola A, Hornik K, Zeileis A. Kernlab—an S4 package for kernel methods in R. J Stat Softw 2004;11:1–20.
[58] Kaunisto MA, Jokela R, Tallgren M, Kambur O, Tikkanen E, Tasmuth T, Sipilä R, Palotie A, Estlander AM, Leidenius M, Ripatti S, Kalso EA. Pain in 1,000 women treated for breast cancer: a prospective study of pain sensitivity and postoperative pain. Anesthesiology 2013;119:1410–21.
[59] Kearns M, Valiant LG. Crytographic limitations on learning Boolean formulae and finite automata. Proceedings of the twenty-first annual ACM symposium on Theory of computing. Seattle, WA: ACM, 1989. p. 433–444.
[60] Kim HY. Phospholipids: a neuroinflammation emerging target. Nat Chem Biol 2015;11:99–100.
[61] Knisely MR, Conley YP, Kober KM, Smoot B, Paul SM, Levine JD, Miaskowski C. Associations between catecholaminergic and serotonergic genes and persistent breast pain phenotypes following breast cancer surgery. J Pain 2018;19:1130–46.
[62] Ko SW, Vadakkan KI, Ao H, Gallitano-Mendel A, Wei F, Milbrandt J, Zhuo M. Selective contribution of Egr1 (zif/268) to persistent inflammatory pain. J Pain 2005;6:12–20.
[63] Kringel D, Geisslinger G, Resch E, Oertel BG, Thrun MC, Heinemann S, Lotsch J. Machine-learned analysis of the association of next-generation sequencing-based human TRPV1 and TRPA1 genotypes with the sensitivity to heat stimuli and topically applied capsaicin. PAIN 2018;159:1366–81.
[64] Kringel D, Kaunisto MA, Lippmann C, Kalso E, Lötsch J. Development of an AmpliSeq panel for next-generation sequencing of a set of genetic predictors of persisting pain. Front Pharmacol 2018;9:1008.
[65] Kringel D, Sisignano M, Zinn S, Lötsch J. Next-generation sequencing of the human TRPV1 gene and the regulating co-players LTB4R and LTB4R2 based on a custom AmpliSeq panel. PLoS One 2017;12: e0180116.
[66] Kumar V, Mahal BA. NGF—the TrkA to successful pain treatment. J Pain Res 2012;5:279–87.
[67] Lacroix-Fralish ML, Ledoux JB, Mogil JS. The Pain Genes Database: an interactive web browser of pain-related transgenic knockout studies. PAIN 2007;131:3 e1–4.
[68] Langford DJ, West C, Elboim C, Cooper BA, Abrams G, Paul SM, Schmidt BL, Levine JD, Merriman JD, Dhruva A, Neuhaus J, Leutwyler H, Baggott C, Sullivan CW, Aouizerat BE, Miaskowski C. Variations in potassium channel genes are associated with breast pain in women prior to breast cancer surgery. J Neurogenet 2014;28:122–35.
[69] Latremoliere A, Latini A, Andrews N, Cronin SJ, Fujita M, Gorska K, Hovius R, Romero C, Chuaiphichai S, Painter M, Miracca G, Babaniyi O, Remor AP, Duong K, Riva P, Barrett LB, Ferreiros N, Naylor A, Penninger JM, Tegeder I, Zhong J, Blagg J, Channon KM, Johnsson K, Costigan M, Woolf CJ. Reduction of neuropathic and inflammatory pain through inhibition of the tetrahydrobiopterin pathway. Neuron 2015;86: 1393–406.
[70] Lau BK, Vaughan CW. Descending modulation of pain: the GABA disinhibition hypothesis of analgesia. Curr Opin Neurobiol 2014;29: 159–64.
[71] Law PY, Reggio PH, Loh HH. Opioid receptors: toward separation of analgesic from undesirable effects. Trends Biochem Sci 2013;38: 275–82.
[72] Lee M, Tracey I. Neuro-genetics of persistent pain. Curr Opin Neurobiol 2013;23:127–32.
[73] Leung L, Cahill CM. TNF-alpha and neuropathic pain—a review. J Neuroinflammation 2010;7:27.
[74] Liaw A, Wiener M. Classification and regression by randomForest. R News 2002;2:18–22.
[75] Lindstedt F, Karshikoff B, Schalling M, Olgart Hoglund C, Ingvar M, Lekander M, Kosek E. Serotonin-1A receptor polymorphism (rs6295) associated with thermal pain perception. PLoS One 2012;7:e43221.

[76] Liu YN, Yang X, Suo ZW, Xu YM, Hu XD. Fyn kinase-regulated NMDA receptor- and AMPA receptor-dependent pain sensitization in spinal dorsal horn of mice. Eur J Pain 2014;18:1120–8.
[77] Loggia ML, Chonde DB, Akeju O, Arabasz G, Catana C, Edwards RR, Hill E, Hsu S, Izquierdo-Garcia D, Ji RR, Riley M, Wasan AD, Zurcher NR, Albrecht DS, Vangel MG, Rosen BR, Napadow V, Hooker JM. Evidence for brain glial activation in chronic pain patients. Brain 2015;138:604–15.
[78] Loncar Z, Curic G, Mestrovic AH, Mickovic V, Bilic M. Do IL-1B and IL-1RN modulate chronic low back pain in patients with post-traumatic stress disorder? Coll Antropol 2013;37:1237–44.
[79] Lötsch J, Geisslinger G. Are mu-opioid receptor polymorphisms important for clinical opioid therapy? Trends Mol Med 2005;11:82–9.
[80] Lötsch J, Geisslinger G. Current evidence for a modulation of nociception by human genetic polymorphisms. PAIN 2007;132:18–22.
[81] Lötsch J, Geisslinger G, Tegeder I. Genetic modulation of the pharmacological treatment of pain. Pharmacol Ther 2009;124:168–84.
[82] Lötsch J, Hummel T. A machine-learned analysis suggests non-redundant diagnostic information in olfactory subtests. IBRO Rep 2019;6:64–73.
[83] Lötsch J, Lerch F, Djaldetti R, Tegeder I, Ultsch A. Identification of disease-distinct complex biomarker patterns by means of unsupervised machine-learning using an interactive R toolbox (Umatrix). Bi Data Anal 2018;3:5.
[84] Lötsch J, Sipilä R, Dimova V, Kalso E. Machine-learned selection of psychological questionnaire items relevant to the development of persistent pain after breast cancer surgery. Br J Anaesth 2018;121: 1123–32.
[85] Lötsch J, Sipilä R, Tasmuth T, Kringel D, Estlander AM, Meretoja T, Kalso E, Ultsch A. Machine-learning-derived classifier predicts absence of persistent pain after breast cancer surgery with high accuracy. Breast Cancer Res Treat 2018;171:399–411.
[86] Lotsch J, Ultsch A. Machine learning in pain research. PAIN 2017;159: 623–30.
[87] Lötsch J, Ultsch A. Exploiting the structures of the U-matrix. In: Villmann T, Schleif FM, Kaden M, Lange M, editors. Advances in intelligent systems and computing, Vol. 295. Heidelberg: Springer, 2014. p. 248–57.
[88] Lötsch J, Ultsch A. Random forests followed by computed ABC analysis as a feature selection method for machine-learning in biomedical data. In: Imaizumi T editor, Proceedings of the Conference of the International Federation of Classification Societies IFCS-2017: Springer, 2017.
[89] Lötsch J, Ultsch A, Kalso E. Prediction of persistent post-surgery pain by preoperative cold pain sensitivity: biomarker development with machine-learning-derived analysis. Br J Anaesth 2017;119:821–29.
[90] Macrae WA. Chronic pain after surgery. Br J Anaesth 2001;87:88–98.
[91] Martikainen IK, Nuechterlein EB, Pecina M, Love TM, Cummiford CM, Green CR, Stohler CS, Zubieta JK. Chronic back pain is associated with alterations in dopamine neurotransmission in the ventral striatum. J Neurosci 2015;35:9957–65.
[92] McCann B, Miaskowski C, Koetters T, Baggott C, West C, Levine JD, Elboim C, Abrams G, Hamolsky D, Dunn L, Rugo H, Dodd M, Paul SM, Neuhaus J, Cooper B, Schmidt B, Langford D, Cataldo J, Aouizerat BE. Associations between pro- and anti-inflammatory cytokine genes and breast pain in women prior to breast cancer surgery. J Pain 2012;13: 425–37.
[93] McCarson KE, Enna SJ. GABA pharmacology: the search for analgesics. Neurochem Res 2014;39:1948–63.
[94] McCourt CM, McArt DG, Mills K, Catherwood MA, Maxwell P, Waugh DJ, Hamilton P, O'Sullivan JM, Salto-Tellez M. Validation of next generation sequencing technologies in comparison to current diagnostic gold standards for BRAF, EGFR and KRAS mutational analysis. PLoS One 2013;8:e69604.
[95] McDougall JJ, Muley MM, Philpott HT, Reid A, Krustev E. Early blockade of joint inflammation with a fatty acid amide hydrolase inhibitor decreases end-stage osteoarthritis pain and peripheral neuropathy in mice. Arthritis Res Ther 2017;19:106.
[96] Meacham K, Shepherd A, Mohapatra DP, Haroutounian S. Neuropathic pain: central vs. peripheral mechanisms. Curr Pain Headache Rep 2017; 21:28.
[97] Meißner T, Mark A, Williams C, Berdel WE, Wiebe S, Kerkhoff A, Wardelmann E, Gaiser T, Müller-Tidow C, Rosenstiel P, Arnold N, Leyland-Jones B, Franke A, Stanulla M, Forster M. Metastatic triple-negative breast cancer patient with TP53 tumor mutation experienced 11 months progression-free survival on bortezomib monotherapy without adverse events after ending standard treatments with grade 3 adverse events. Cold Spring Harb Mol Case Stud 2017;3:a001677.
[98] Meretoja TJ, Andersen KG, Bruce J, Haasio L, Sipila R, Scott NW, Ripatti S, Kehlet H, Kalso E. Clinical prediction model and tool for assessing risk of persistent pain after breast cancer surgery. J Clin Oncol 2017;35: 1660–7.
[99] Meretoja TJ, Leidenius MH, Tasmuth T, Sipila R, Kalso E. Pain at 12 months after surgery for breast cancer. JAMA 2014;311:90–2.
[100] Metzker ML. Sequencing technologies—the next generation. Nat Rev Genet 2010;11:31–46.
[101] Mitchell CP, Chen Y, Kundakovic M, Costa E, Grayson DR. Histone deacetylase inhibitors decrease reelin promoter methylation in vitro. J Neurochem 2005;93:483–92.
[102] Mogil JS. Pain genetics: past, present and future. Trends Genet 2012; 28:258–66.
[103] Mustonen L, Aho T, Harno H, Sipilä R, Meretoja T, Kalso E. What makes surgical nerve injury painful? A 4-9 year follow-up of patients with intercostobrachial nerve resection in women treated for breast cancer. PAIN 2018;160:246–56.
[104] Nicholl BI, Holliday KL, Macfarlane GJ, Thomson W, Davies KA, O'Neill TW, Bartfai G, Boonen S, Casanueva FF, Finn JD, Forti G, Giwercman A, Huhtaniemi IT, Kula K, Punab M, Silman AJ, Vanderschueren D, Wu FC, McBeth J; European Male Ageing Study G. Association of HTR2A polymorphisms with chronic widespread pain and the extent of musculoskeletal pain: results from two population-based cohorts. Arthritis Rheum 2011;63:810–18.
[105] Nissenbaum J, Devor M, Seltzer Z, Gebauer M, Michaelis M, Tal M, Dorfman R, Abitbul-Yarkoni M, Lu Y, Elahipanah T, delCanho S, Minert A, Fried K, Persson AK, Shpigler H, Shabo E, Yakir B, Pisante A, Darvasi A. Susceptibility to chronic pain following nerve injury is genetically affected by CACNG2. Genome Res 2010;20:1180–90.
[106] Obata K, Noguchi K. BDNF in sensory neurons and chronic pain. Neurosci Res 2006;55:1–10.
[107] Offenbaecher M, Bondy B, de Jonge S, Glatzeder K, Kruger M, Schoeps P, Ackenheil M. Possible association of fibromyalgia with a polymorphism in the serotonin transporter gene regulatory region. Arthritis Rheum 1999;42:2482–8.
[108] Onojjighofia T, Meshkin B, Nguyen SV, Schwartz D, Akindele B. Perception of analgesia in narcotic users with chronic pain: a multi-center cross-sectional study comparing genotype to pain VAS (P.A.I.N. Study). Neurology 2014;82:P4.349.
[109] Pareto V. Manuale di economia politica, Milan: Società editrice libraria, 1909. revised and translated into French as Manuel d'économie politique.
[110] Pearson K. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. Philosophical Mag 1900;50:157–75.
[111] Petrenko AB, Yamakura T, Baba H, Shimoji K. The role of N-methyl-D-aspartate (NMDA) receptors in pain: a review. Anesth Analg 2003;97: 1108–16.
[112] Potvin S, Larouche A, Normand E, de Souza JB, Gaumond I, Grignon S, Marchand S. DRD3 Ser9Gly polymorphism is related to thermal pain perception and modulation in chronic widespread pain patients and healthy controls. J Pain 2009;10:969–75.
[113] Qamri Z, Preet A, Nasser MW, Bass CE, Leone G, Barsky SH, Ganju RK. Synthetic cannabinoid receptor agonists inhibit tumor growth and metastasis of breast cancer. Mol Cancer Ther 2009;8: 3117–29.
[114] R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, 2008. Available online at: https://www.R-project.org/.
[115] Reimann F, Cox JJ, Belfer I, Diatchenko L, Zaykin DV, McHale DP, Drenth JP, Dai F, Wheeler J, Sanders F, Wood L, Wu TX, Karppinen J, Nikolajsen L, Mannikko M, Max MB, Kiselycznyk C, Poddar M, Te Morsche RH, Smith S, Gibson D, Kelempisioti A, Maixner W, Gribble FM, Woods CG. Pain perception is altered by a nucleotide polymorphism in SCN9A. Proc Natl Acad Sci U S A 2010;107:5148–53.
[116] Ribeiro-Dasilva MC, Peres Line SR, Leme Godoy dos Santos MC, Arthuri MT, Hou W, Fillingim RB, Rizzatti Barbosa CM. Estrogen receptor-alpha polymorphisms and predisposition to TMJ disorder. J Pain 2009;10:527–33.
[117] Rizzo JM, Buck MJ. Key principles and clinical applications of "next-generation" DNA sequencing. Cancer Prev Res (Phila) 2012;5:887–900.
[118] Rockman MV, Kruglyak L. Genetics of global gene expression. Nat Rev Genet 2006;7:862–72.
[119] Saeys Y, Inza I, Larranaga P. A review of feature selection techniques in bioinformatics. Bioinformatics 2007;23:2507–17.
[120] Sah DW, Ossipov MH, Rossomando A, Silvian L, Porreca F. New approaches for the treatment of pain: the GDNF family of neurotrophic growth factors. Curr Top Med Chem 2005;5:577–83.

[121] Schapire RE, Freund Y. A short introduction to boosting. J Jpn Soc Artif Intell 1999;14:771–80.
[122] Schapire RE, Freund Y. Boosting: foundations and algorithms: The MIT Press, 2012.
[123] Schou Bredal I, Smeby NA, Ottesen S, Warncke T, Schlichting E. Chronic pain in breast cancer survivors: comparison of psychosocial, surgical, and medical characteristics between survivors with and without pain. J Pain Symptom Manage 2014;48:852–62.
[124] Schwartz ES, Kim HY, Wang J, Lee I, Klann E, Chung JM, Chung K. Persistent pain is dependent on spinal mitochondrial antioxidant levels. J Neurosci 2009;29:159–68.
[125] Shannon CE. A mathematical theory of communication. Bell Syst Techn J 1951;30:50–64.
[126] Shi TJS, Liu SXL, Hammarberg H, Watanabe M, Xu ZQD, Hökfelt T. Phospholipase C{beta}3 in mouse and human dorsal root ganglia and spinal cord is a possible target for treatment of neuropathic pain. Proc Natl Acad Sci U S A 2008;105:20004–8.
[127] Shoskes DA, Albakri Q, Thomas K, Cook D. Cytokine polymorphisms in men with chronic prostatitis/chronic pelvic pain syndrome: association with diagnosis and treatment response. J Urol 2002;168:331–5.
[128] Sie D, Snijders PJ, Meijer GA, Doeleman MW, van Moorsel MI, van Essen HF, Eijk PP, Grunberg K, van Grieken NC, Thunnissen E, Verheul HM, Smit EF, Ylstra B, Heideman DA. Performance of amplicon-based next generation DNA sequencing for diagnostic gene mutation profiling in oncopathology. Cell Oncol (Dordr) 2014;37:353–61.
[129] Sipilä R, Estlander AM, Tasmuth T, Kataja M, Kalso E. Development of a screening instrument for risk factors of persistent pain after breast cancer surgery. Br J Cancer 2012;107:1459–66.
[130] Sluka KA, Audette KM. Activation of protein kinase C in the spinal cord produces mechanical hyperalgesia by activating glutamate receptors, but does not mediate chronic muscle-induced hyperalgesia. Mol Pain 2006;2:13.
[131] Smith FL, Fujimori K, Lowe J, Welch SP. Characterization of delta9-tetrahydrocannabinol and anandamide antinociception in nonarthritic and arthritic rats. Pharmacol Biochem Behav 1998;60:183–91.
[132] Snider WD, McMahon SB. Tackling pain at the source: new ideas about nociceptors. Neuron 1998;20:629–32.
[133] Sorge RE, Trang T, Dorfman R, Smith SB, Beggs S, Ritchie J, Austin JS, Zaykin DV, Vander Meulen H, Costigan M, Herbert TA, Yarkoni-Abitbul M, Tichauer D, Livneh J, Gershon E, Zheng M, Tan K, John SL, Slade GD, Jordan J, Woolf CJ, Peltz G, Maixner W, Diatchenko L, Seltzer Z, Salter MW, Mogil JS. Genetically determined P2X7 receptor pore formation regulates variability in chronic pain sensitivity. Nat Med 2012;18:595–9.
[134] Stephens K, Cooper BA, West C, Paul SM, Baggott CR, Merriman JD, Dhruva A, Kober KM, Langford DJ, Leutwyler H, Luce JA, Schmidt BL, Abrams GM, Elboim C, Hamolsky D, Levine JD, Miaskowski C, Aouizerat BE. Associations between cytokine gene variations and severe persistent breast pain in women following breast cancer surgery. J Pain 2014;15:169–80.
[135] Stephens KE, Levine JD, Aouizerat BE, Paul SM, Abrams G, Conley YP, Miaskowski C. Associations between genetic and epigenetic variations in cytokine genes and mild persistent breast pain in women following breast cancer surgery. Cytokine 2017;99:203–13.
[136] Sugaya K, Nishijima S, Yamada T, Miyazato M, Hatano T, Ogawa Y. Molecular analysis of adrenergic receptor genes and interleukin-4/interleukin-4 receptor genes in patients with interstitial cystitis. J Urol 2002;168:2668–71.
[137] Sun YG, Gracias NG, Drobish JK, Vasko MR, Gereau RW, Chen ZF. The c-kit signaling pathway is involved in the development of persistent pain. PAIN 2009;144:178–86.
[138] Taudien S, Lausser L, Giamarellos-Bourboulis EJ, Sponholz C, Schoneweck F, Felder M, Schirra LR, Schmid F, Gogos C, Groth S, Petersen BS, Franke A, Lieb W, Huse K, Zipfel PF, Kurzai O, Moepps B, Gierschik P, Bauer M, Scherag A, Kestler HA, Platzer M. Genetic factors of the disease course after sepsis: rare deleterious variants are predictive. EBioMedicine 2016;12:227–38.
[139] Tegeder I, Costigan M, Griffin RS, Abele A, Belfer I, Schmidt H, Ehnert C, Nejim J, Marian C, Scholz J, Wu T, Allchorne A, Diatchenko L, Binshtok AM, Goldman D, Adolph J, Sama S, Atlas SJ, Carlezon WA, Parsegian A, Lötsch J, Fillingim RB, Maixner W, Geisslinger G, Max MB, Woolf CJ. GTP cyclohydrolase and tetrahydrobiopterin regulate pain sensitivity and persistence. Nat Med 2006;12:1269–77.
[140] Thrun MC. Projection-based clustering through self-organization and swarm intelligence. Wiesbaden, Germany: Springer, 2018.
[141] Thrun MC. Projection-based clustering through self-organization and swarm intelligence: combining cluster analysis with the visualization of high-dimensional data: Springer Fachmedien Wiesbaden, 2018.
[142] Thuault S. A peripheral messenger for chronic pain. Nat Neurosci 2016;19:9.
[143] Tian L, Ma L, Kaarela T, Li Z. Neuroimmune crosstalk in the central nervous system and its significance for neurological diseases. J Neuroinflammation 2012;9:155.
[144] Trimble EL, Ungerleider RS, Abrams JA, Kaplan RS, Feigal EG, Smith MA, Carter CL, Friedman MA. Neoadjuvant therapy in cancer treatment. Cancer 1993;72(11 suppl):3515–24.
[145] Tsantoulas C, Mooney ER, McNaughton PA. HCN2 ion channels: basic science opens up possibilities for therapeutic intervention in neuropathic pain. Biochem J 2016;473:2717–36.
[146] Ultsch A, Lötsch J. Computed ABC analysis for rational selection of most informative variables in multivariate data. PLoS One 2015;10:e0129767.
[147] Vadakkan KI, Wang H, Ko SW, Zastepa E, Petrovic MJ, Sluka KA, Zhuo M. Genetic reduction of chronic muscle pain in mice lacking calcium/calmodulin-stimulated adenylyl cyclases. Mol Pain 2006;2:7.
[148] Varma S, Simon R. Bias in error estimation when using cross-validation for model selection. BMC Bioinformatics 2006;7:91.
[149] Villeda SA, Akopians AL, Babayan AH, Basbaum AI, Phelps PE. Absence of Reelin results in altered nociception and aberrant neuronal positioning in the dorsal spinal cord. Neuroscience 2006;139:1385–96.
[150] Walker K, Bowes M, Panesar M, Davis A, Gentry C, Kesingland A, Gasparini F, Spooren W, Stoehr N, Pagano A, Flor PJ, Vranesic I, Lingenhoehl K, Johnson EC, Varney M, Urban L, Kuhn R. Metabotropic glutamate receptor subtype 5 (mGlu5) and nociceptive function. I. Selective blockade of mGlu5 receptors in models of acute, persistent and chronic pain. Neuropharmacology 2001;40:1–9.
[151] Walker SH, Duncan DB. Estimation of the probability of an event as a function of several independent variables. Biometrika 1967;54:167–79.
[152] Ward JH Jr. Hierarchical grouping to optimize an objective function. J Am Stat Assoc 1963;58:236–44.
[153] Weng Z, Fluckiger AC, Nisitani S, Wahl MI, Le LQ, Hunter CA, Fernal AA, Le Beau MM, Witte ON. A DNA damage and stress inducible G protein-coupled receptor blocks cells in G(2)/M. Proc Natl Acad Sci U S A 1998;95:12334–9.
[154] Wild A. Best practice in inventory management. New York: Wiley, 1997.
[155] Wolters PL, Burns KM, Martin S, Baldwin A, Dombi E, Toledo-Tamula MA, Dudley WN, Gillespie A, Widemann BC. Pain interference in youth with neurofibromatosis type 1 and plexiform neurofibromas and relation to disease severity, social-emotional functioning, and quality of life. Am J Med Genet A 2015;167A:2103–13.
[156] Yamamoto M, Ito Y, Mitsuma N, Hattori N, Sobue G. Pain-related differential expression of NGF, GDNF, IL-6, and their receptors in human vasculitic neuropathies. Intern Med 2003;42:1100–3.
[157] Yang L, Gu X, Zhang W, Zhang J, Ma Z. Cdk5 inhibitor roscovitine alleviates neuropathic pain in the dorsal root ganglia by downregulating N-methyl-D-aspartate receptor subunit 2A. Neurol Sci 2014;35:1365–71.
[158] Zanette SA, Dussan-Sarria JA, Souza A, Deitos A, Torres ILS, Caumo W. Higher serum S100B and BDNF levels are correlated with a lower pressure-pain threshold in fibromyalgia. Mol Pain 2014;10:46.
[159] Zorina-Lichtenwalter K, Meloto CB, Khoury S, Diatchenko L. Genetic predictors of human chronic pain conditions. Neuroscience 2016;338:36–62.