Quantitative N-glycoproteome, phosphoproteome and ubiquitinome analyses for studying B-cell receptor signaling in B-cell lymphoma

Dissertation

zur Erlangung des Doktorgrades der Naturwissenschaften

vorgelegt beim Fachbereich Biochemie, Chemie und Pharmazie (FB 14)
der Johann Wolfgang Goethe-Universität
in Frankfurt am Main

von

Yanlong Ji

aus Nanle, China

Frankfurt am Main 2020

(D30)

vom Fachbereich Biochemie, Chemie und Pharmazie (FB 14)
der Johann Wolfgang Goethe-Universität als Dissertation angenommen.

Dekan: Prof. Dr. Clemens Glaubitz

Gutachter: Prof. Dr. Volker Dötsch
Prof. Dr. Thomas Oellerich

Datum der Disputation: 24.03.2021

Affidavit

Except where stated otherwise by reference or acknowledgment, the work presented was generated by myself under the supervision of my advisors during my doctoral studies. All contributions from colleagues are explicitly referenced in the thesis. Whenever a figure, table or text is identical to a previous publication, it is stated explicitly in the thesis that copyright permission and/or co-author agreement has been obtained.

Göttingen, 22.09.2020

Yanlong Ji

Contents

List of Figures

List of Tables

Abstract

Three types of post-translation modifications (PTMs) containing N-glycosylation, phosphorylation, ubiquitylation were characterized in diffuse large B-cell lymphoma (DLBCL) on a global scale using quantitative mass spectrometry based proteomics technology in this study.

DLBCL is the most common type of malignant lymphomas and has a heterogeneous gene expression profiling, phenotype and clinical response to chemotherapy. DLBCL is a good model for the correct classification of cancers into molecularly different subtypes, which benefits for the selection of rational therapeutic strategies. It resulted in two histologically indistinguishable subtypes-activated B-cell-like (ABC) subgroup and germinal center B-cell-like (GCB) subgroup according to gene expression profiling. Signals originating from the B-cell receptor (BCR), the key protein on the surface of B cells, promote growth and survival of DLBCL. Antigen-dependent/independent BCR signaling is found in DLBCL subtypes.

Recent researches reveal that glycosylation plays role in human cells via site-specific regulation. Aberrant N-glycosylation in BCR-related effectors, such as, CD79a, immunoglobulin M or G (IgM or IgG), has been found to be associated with lymphoid malignancies. However, accurate quantification of intact glycopeptides and their individual glycan moieties in a cell-wide manner is still challenging. Here we established a site-specific quantitative N-glycoproteomics platform termed SugarQuant. It included a fast sample preparation workflow using Protein Aggregation Capture (PAC), an optimized multi-notch MS3 acquisition workflow (Glyco-SPS-MS3), a self-developed R-based tool (GlycoBinder). The robustness and accuracy of quantitation in SugarQuant were proved in a study using the different amounts of TMT-labelled IgM N-glycopeptides spiked into a background of TMT-labelled yeast peptides. Next, we used SugarQuant to identify and quantify more than 5000 unique glycoforms in Burkitt's lymphoma cells treated with a series of doses of 2-deoxy-2-fluoro-L-fucose (2FF) and determine the more accurate site-specific glycosylation changes that occurred upon inhibition of fucosylation compared to using MS2 analysis. It revealed that 2FF-sensitive N-glycosylation on key players in BCR-mediated signaling in DG75. Furthermore, 2FF treatment also affects phosphorylation of the key players involving in B cell receptor signaling.

Then we investigated the site-specific quantitative N-glycoproteome in the cell lines of DLBCL subtypes using SugarQuant. More than 7000 unique intact glycopeptides (glycoforms) were quantified in five ABC DLBCL and four GCB DLBCL cell lines. The glycoproteome mapping (intact glycopeptide expressions) in each cell line allows to segregate DLBCL subtypes. The majority of these glycoforms were from the key cell-surface BCR effectors, such as IgM, CD79 and PTPRC. Lastly, we investigated the change of fucosylated glycopeptides in TMD8 cell line upon knockout of the fucosyltransferase FUT8, which is responsible for core-fucose synthesis, and by the treatment with 2FF. The results revealed that FUT8 might also regulate the synthesis of sub/terminal fucose on glycan chain and the inhibition of fucosylation increased the sialyated glycopeptide expression.

Phosphorylation is involved in regulating multiple processes as an important mediator in BCR signaling. Likewise, ubiquitylation plays vital roles in the activation of the nuclear factor-kappaB (NF-κB) pathway in BCR signaling. There are two vital upstream BCR-proximal tyrosine kinases, Bruton's tyrosine kinase (BTK) and spleen tyrosine kinase (SYK), which regulate the auto-phosphorylation and phosphorylation of other proteins in BCR signaling pathway. Here we investigated the dynamics of downstream phosphorylation and ubiquitylation signaling in ABC DLBCL and GCB DLBCL cell lines

upon the inhibitions of BTK and SYK using quantitative proteomics strategy. In the phosphoproteome analysis, a large dataset of quantified phosphorylation sites was obtained in the three ABC and four GCB DLBCL cell lines. BCR signaling in the subtypes of DLBCL cell lines was found to be highly individual in distinct cell lines. These significantly regulated phosphorylation events in each cell line with individual treatment were involved in multiple Reactome pathways, such as, M phase, signaling by Rho GTPases and diseases of signal transduction. Moreover, the gene regulation-related biological processes including chromosome organization and medication, DNA metabolic process, nuclear export, were involved in the DLBCL cell lines. In the ubiquitinome analysis, we identified more than 15,000 ubiquitylation sites in two ABC and one GCB cell lines upon the inhibition of BTK and SYK. The different ubiquitylation events observed in ABC and GCB subtypes revealed distinct BCR signaling pathways in two subtypes. The similar signaling perturbations across each cell line upon BTK and SYK inhibition, which were obtained from the significantly regulated ubiquitylated peptides expression, revealed the cell-type-specific concordance in ubiquitylation regulation upon BTK and SYK inhibition. These ubiquitylation modified proteins who bore the significantly regulated ubi-peptides in the samples were also found to be highly involved in gene regulatory processes. Furtherly, the integration analysis of phosphoproteome and ubiquitinome data found that the majority of regulated proteins were exclusively identified by a single type of PTM. There were also 592 proteins which are significantly modified by both phosphorylation and ubiquitylation. The Reactome analysis showed that significantly differential phosphorylation and ubiquitylation hits cover primarily distinct pathways, with exception such as mRNA processing, cell cycle checkpoint and M phase.

In summary, we characterized the phosphorylation and ubiquitylation changes in DLBCL subtypes upon inhibition of BTK and SYK by stable isotope labeling with amino acids in cell culture (SILAC)-based quantitative phosphoproteome and ubiquitinome analysis. It revealed that some of the gene regulation related proteins involved BCR signaling pathway, such as, DNA (cytosine-5)-methyltransferase 1 (DNMT1). Moreover, we developed a robust and accurate site-specific quantitative N-glycoproteomics approach. It was used to investigate the intact N-glycopeptides expression in subtypes of DLBCL cell lines and the fucosylated glycopeptide expression in the TMD8 cell line upon the inhibition of fucosylation. These provided valuable datasets for the discovery of unknown effectors and further understanding of the BCR signaling network.

1. Introduction

1.1 Hematological malignancies

Hematopoiesis is the formation of blood cellular components and all cellular blood components are derived from hematopoietic stem cells (HSCs). HSCs reside in the bone marrow and have the unique long-lived self-renewal capacity to preserve the multipotency of mother cells and to generate all of the different mature blood cell types throughout life [1]. When they differentiate, at least some of their daughter cells remain as HSCs, so the pool of stem cells is not depleted. This phenomenon is defined as asymmetric division [2]. HSCs differentiate in the bone marrow into myeloid or lymphoid stem cells. Myeloid stem cells give rise to a second level of lineage-specific progenitor cells that go on to produce neutrophils, monocytes, eosinophils, basophils, mast cells, megakaryocytes, and erythrocytes. Monocytes differentiate further into macrophages in peripheral tissue compartments. Lymphoid stem cells give rise to B-cell, T-cell, and NK cell lineages [3].

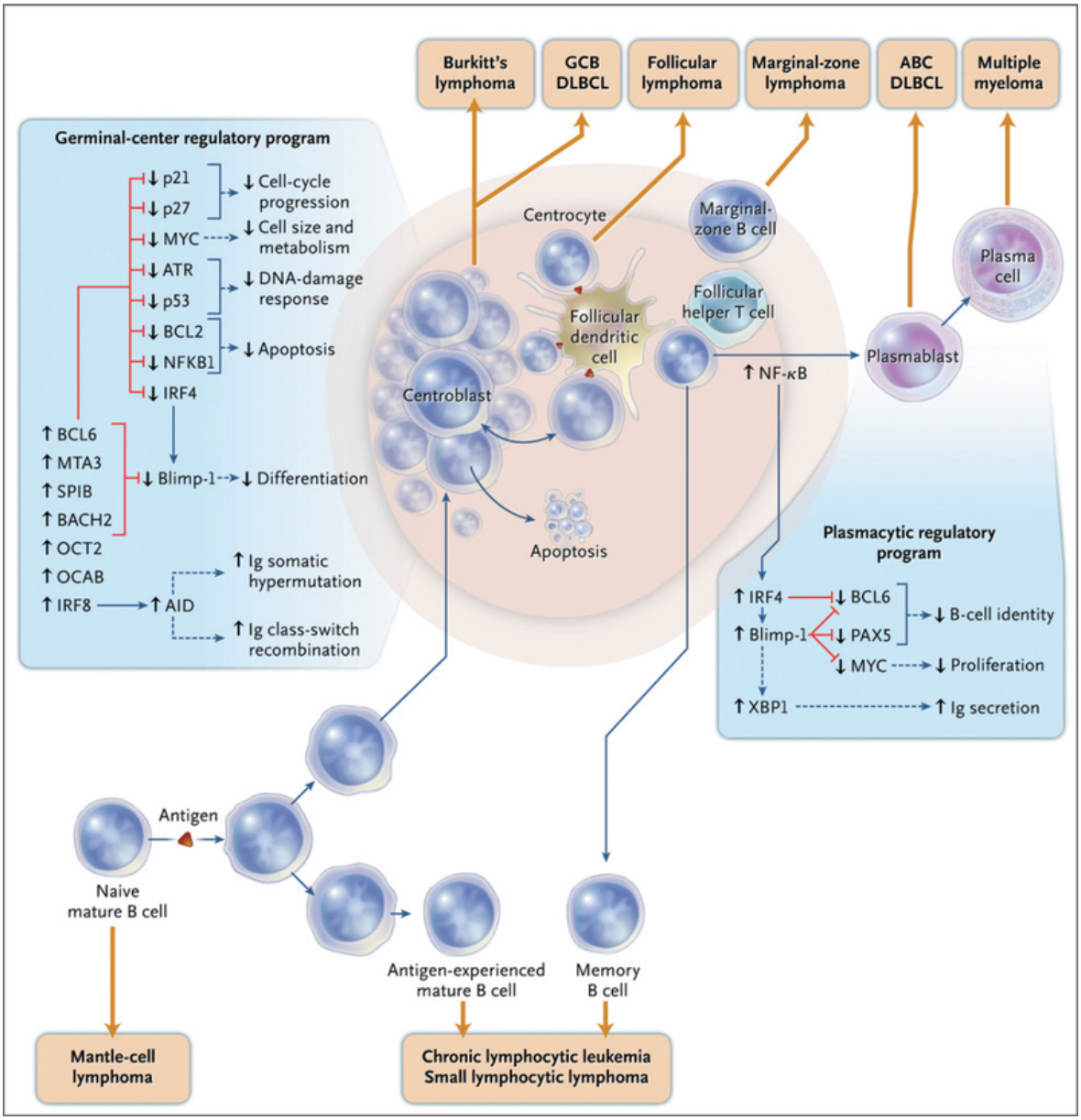


Figure 1.1 B-cell differentiation and lymphomagenesis.

This figure is adapted from Lenz and Staudt, 2010 (New England Journal of Medicine) [4].

Hematological malignancies are malignant neoplasms derived from myeloid and lymphoid blood cell lineages and affect the blood, bone marrow and lymphatic system [5, 6]. There are three main types of hematological malignancies, which are leukemia, lymphoma and myeloma (**Figure 1.1**). Leukemias are a group of life threatening malignant disorders of the blood and bone marrow [7]. They are divided into myeloid and lymphatic origin according to the malignant transformed cell type. Furthermore, there are two different forms as chronic and acute ones among leukemias. Acute leukemia usually develops quickly. The number of leukemia cells increases rapidly and these abnormal cells don't do the work of normal white blood cells. Chronic leukemia always develops slowly. The leukemia cells work almost as well as normal white blood cells. Till now, there are four main types of leukemia including acute lymphocytic leukemia (ALL), acute myeloid leukemia (AML), chronic lymphocytic leukemia (CLL), and chronic myeloid leukemia (CML) [6]. Lymphoma is a type of hematological malignancy that develops from lymphocytes involving the lymphatic system [8]. The two main categories of lymphomas are Hodgkin's lymphoma (HL) and the non-Hodgkin lymphomas (NHL) [9]. HL is marked by the presence of a type of cell called the Hodgkin-Reed-Sternberg (HRS) cells which accounts for about 10% of all lymphomas [10]. HL is further subclassified as either nodular lymphocyte predominant Hodgkin lymphoma or the classical Hodgkin lymphoma with the four specific subtypes: nodular sclerosis, lymphocyte-rich, mixed cellularity, and lymphocyte depletion [11]. NHL includes a diverse spectrum of cancers or tumors involving the immune system. About 85%-90% of NHL are derived from B cells, whereas the others are from T cells or NK cells. Depending on the exhibited features, eg, morphology, immune phenotype, rearrangement of the immunoglobulin genes and recurrent chromosomal aberrations, NHL can be divided into many subtypes [10]. All the subtypes can arise at multiple stages of normal B-cell development [12]. For example, Mantle cell lymphoma (MCL) is derived from naïve mature B-cells, and characterized by expression of CD5, CD20 and Cyclin D1 as well as Cyclin D1 rearrangement as a consequence of the t(11;14) recurrent translocation [13]. Small lymphocytic lymphoma (SLL) is mostly derived from antigen-experienced mature B-cells and memory-B cells which exist with the strong immune-phenotype of CD79a, CD23 and CD5 [13]. Several types of lymphoma including diffuse large-B-cell lymphoma (DLBCL), follicular lymphoma (FL) and Burkitt's lymphoma arise from germinal-center B cells. FL is the most common small cell lymphoma, which is characterized by the t(14;18) recurrent translocation resulting in rearrangement of BCL-2. Myeloma is a type of hematological malignancy of plasma cells, which are white blood cells that normally produce antibodies during immune defense. Multiple myeloma is its most common type [14]. Two types of NHL: Burkitt's lymphoma and diffuse large B-cell lymphoma (DLBCL) were mainly analyzed in this study and will be described in further details.

1.1.1 Burkitt's lymphoma

Burkitt's lymphoma (BL) is a highly aggressive B-cell NHL and the most proliferative of all cancers with a 24-48 h cell doubling time [15]. It was firstly described by the Irish surgeon Denis Burkitt working in Kampala Uganda, Africa in 1958 depending on 38 cases of "sarcomas" involving the jaws of African children seen at Mulago Hospital [16]. BL was found to be of lymphoma lineage in 1960 by the pathologist George O'Connor [17]. Four years later, Epstein-Barr virus (EBV) was identified in the tumor tissue by the three virologists, Michael Anthony Epstein, Yvonne Barr and Bert Achong [18]. Currently, BL was divided into three main clinical variants: endemic BL (eBL), sporadic BL (sBL) and immunodeficiency-related BL. eBL (also called "African variant") remains the most common pediatric malignancy in sub-Saharan Africa. It is associated with malaria infection [19, 20] and EBV is present in almost all cases of eBL. sBL (also known as "non-African") mainly occurs across the rest of the world and is rarely associated with EBV infection. The immunodeficiency-related type is usually

associated with human immunodeficiency virus (HIV) infection. These types are similar to morphology, immunophenotype. At the genetic and transcriptional level, sBL and HIV subtypes have a relatively higher relation between each other comparing to eBL by whole genome and transcriptome sequencing of all three subtypes [21]. The karyotype of BL is the increasing genetic complexity linked to disease progression [22]. The activation-induced deaminase (AID) is one important driver of the GC gene expression program, since it mediates somatic hypermutation occurring in the centroblast-rich dark zone of normal GC cells and immunoglobulin (IG) class-switch recombination (CSR) in the centrocyte-rich light zone of the GC [23]. Immunoglobulin heavy-chain locus with MYC translocation which are mostly introduced by AID is the hallmark of BL. Three variants were found till now including the most common translocation t(8;14) (q24;32) variant which occurs in 70-80% patients, t(2;8)(p12;q24) and t(8;22)(q24;q11) which occur in 10–15% of patients [15]. MYC is a family of regulator genes and oncogenes that code for transcription factors. Three related human genes of c-Myc, I-Myc and n-Myc are belonging to Myc family. C-Myc (also referred to as MYC) was the first discovered gene in this family and was identified at the breakpoint of the t(8;14) translocation in BL. The BL associated translocation brings Ig genes with their enhancer regions in close proximity to the transcription factor MYC leading to its constitutive transcription deregulation [24]. In BL, MYC is subjected to a mutation rate of 70% and mutations always influence its transactivation domain [25]. The inactivating mutations of TP53 are also common in BL occurring in 35% of cases and it is probable that overexpression of MYC in primary cells induces the TP53-dependent apoptotic pathway [26]. Mutations in the CCND3 gene encoding cyclin D3 which regulates the G1-S cell cycle transition promotes the proliferation of BL. Highly recurrent mutations affect TCF-3 which is a basic helix-loop-helix transcription factor [27] and its negative regulator ID3. The TCF-3/ID3 module plays a central role in BL tumorigenesis due to that all BL cell lines rely on TCF-3 for survival and proliferation including those without mutations in TCF-3/ID3 [28, 29]. ID3's silencing mutations were a common feature in all three BL subtypes [30, 31]. ID3 protein also binds to TCF4 besides TCF3, which was proved by mass-spectrometry-based proteomics technology [21]. TCF4 can play the role of TCF3 when TCF3 has lost its function. ID3 loss potentiates the effect of MYC overexpression leading to rapid tumorigenesis in BL. More and more driver genes including coding and non-coding mutations were identified in BL by whole genome and transcriptome analysis [21, 32]. With respect to cellular signaling, BL is characterized by the tonic antigen-independent B cell receptor signaling, which just engages the phosphoinositide 3-kinase (PI3K) pathway [25]. Intensive chemotherapy is currently the common treatment for BL. New treatment regimens are urgently needed for BL due to the toxic side effects of chemotherapy and lower resource in Africa for chemotherapy.

1.1.2 Diffuse large B-cell lymphoma

DLBCL is an aggressive and the most common B-cell NHL across the world, which accounts for about 40% of all NHLs diagnoses [33]. DLBCL are phenotypically and genetically heterogeneous within morphologically indistinguishable tumors and differential therapeutic responses. Three distinct molecular cell-of-origin subgroups of DLBCL have been identified by gene-expression profiling using DNA microarrays, including germinal center B-cell-like (GCB), activated B-cell-like (ABC) and unclassified [30, 34]. The GCB subtype is derived from rapidly proliferating germinal center B cells in the centroblast zone [34, 35]. ABC DLBCLs express genes which are similar with that from antigen-activated B cells and it is significantly regulated by the continuous activation of certain pathways that are normally activated during B cells interacting with antigen, in particular, the NF-κB pathway [36]. ABC DLBCL has a worse survival rate compared to GCB DLBCL (40% versus 74% in three-year

progression-free survival rate) when treated with the most common regimen of R-CHOP (rituximab plus cyclophosphamide, doxorubicin, vincristine and prednisone) for DLBCL [37].

In 2018, Schmitz et al used a multiplatform genomic analysis of gene expression and structural genomic abnormalities including DNA copy-number alterations, translocations and mutations in 574 DLBCL biopsy samples to define DLBCL subtypes [30]. Four main genetic subtypes named MCD (based on *MYD88*^{L265P} and *CD79B* mutations), BN2 (*BCL6* translocations and *NOTCH2* mutations), N1 (based on *NOTCH1* mutations) and EZB (based on *EZH2* mutations and *BCL2* translocations) were identified based on the co-occurrence of genetic alterations in about 46 % of biopsy samples used. Meanwhile, they found that the four subtypes are also different in gene expression signatures and responses to immunochemotherapy [38]. The MCD subtype is mainly comprised of ABC cases and share many same genetic alterations with primary extranodal lymphomas. It relies on the B cell receptor (BCR) dependent NF-κB pathway which was reflected in a favorable clinical outcome (80% response rate) to in response to ibrutinib (a small molecule inhibitor of BTK affecting the NF-κB pathway) monotherapy compared to the overall ibrutinib-response rate of 37% in all ABC cases [39]. A multiprotein supercomplex of MYD88-TLR9-BCR (My-T-BCR) residing in endolysosomes together with the mTOR complex was discovered in ibrutinib-responsive cell lines (mainly MCD subtypes) by genome-wide CRISPR-Cas9 screening and functional proteomics technology [40]. The BN2 subtype is sustained by all three gene expression subgroups and is characterized by NOTCH pathway aberrations caused by a *NOTCH2* mutation or amplification, SPEN mutation and NF-κB pathway aberrations involving the NF-κB negative regulators *A20* and its partner *TNIP1*, as well as the BCR-dependent pathway components protein kinase C beta and *BCL10*. BCL6 fusions are also enriched in BN2 subtype. The N1 subtype is largely maintained by ABC cases which represent with by NOTCH1 mutations. The EZB subtype mainly consists of GCB cases and the genetic aberrations including BCL2 translocation, EZH2 mutation, and REL amplification, as well as inactivation of the tumor suppressors TNFRSF14, CREBBP, EP300, and KMT2D, are enriched in the EZB subtype. The different responses to immunochemotherapy that were observed in the individual genetic subtypes, i.e. higher survival rates in the BN2 and EZB subtypes in comparison to the MCD and N1 subtypes, support the medical or therapeutic relevance of further classifying gene expression subgroups into genetic subtypes [30]. Genetic alterations of proteins involving the BCR signaling happen with diverse frequencies in all DLBCL genetic subtypes, which illustrates the concurrent evolution of this phenotype of DLBCL during DLBCL pathogenesis.

1.1.3 B cell receptor signaling

The immune system is a host defense system composed of many biological structures and processes that protect organisms against diseases. Two subsystems of the immune system-the innate immune system and the adaptive immune system exist in vertebrates. In both subsystems, humoral immunity which is mediated by macromolecules present in the body fluids (e.g. secreted antibodies, complement proteins, antimicrobial peptides) and cell-mediated immunity contribute to efficient clearance of pathogens. B cell and T cells are the two types of lymphocytes that constitute the cellular component of the adaptive immune system. T cells are involved in cell-mediated immune response which is not mediated by antibodies, whereas B cells are involved in humoral immunity. The B-cell receptor (BCR) is expressed on the outer surface of B cells which allows the B cells to bind to a specific antigen in order to initiate an antibody response [41].

The BCR is unique in every normal and malignant B cells. It is composed of two parts: a membrane-bound antibody portion consisting of two immunoglobulin heavy (IgH) and two immunoglobulin light (IgL) chains that is non-covalently coupled to a disulphide-linked heterodimer of the CD79A (Igα) and

CD79B (Igβ) subunits [42]. IgH and IgL contain constant regions and the antigen recognition intact variable regions. The assembled variable region is formed by the V(D)J recombination-activating protein 1 (RAG1) and RAG2-mediated recombination of numerous variable (V), diversity (D, only in heavy chain) and joining (J) gene segments. Three hypervariable subgroups termed complementarity-determining regions which own diverse amino acid sequences in the assembled V regions allow the BCR to recognize a huge diversity of antigens from foreign and self. In the case of IgH, the constant region belongs to one of several classes (isotypes), for example, the IgM and IgD constant regions are expressed by naïve B-cells, expression of IgA, IgG and IgE constant regions occurs after antigen mediated activation of mature B-cells and involves CSR. IgM mediated BCR signaling triggers B-cell proliferation, whereas signaling triggered by non-IgM BCRs promotes plasmacytic differentiation [43].

There are several modes of BCR signaling in aggressive lymphoma (**Figure 1.2**). Chronic active BCR signaling was firstly discovered in ABC DLBCL which is characterized by constitutive activation of NF-κB signaling [44, 45]. Two possible NF-κB activation mechanisms were proposed: (1). CARD11 mutations in about 15% of ABC cases induce the formation of the CARD11, BCL10 and MALT1 (CBM) complex which activates IκB kinase beta in the NF-κB pathway [46]. (2). The survival of ABC lines with wildtype CARD11 is retained by signaling from the BCR which involves many effectors of BCR-dependent NF-κB activation (eg, SYK, BTK, BLNK and et al) [44]. Recurrent mutations affecting ITAM motifs of CD79A and CD79B which exist in about 29% of ABC and in only 3% of GCB cases were revealed by targeted sequencing of genes involved in BCR signaling [30, 44]. CD79B mutations of the first ITAM tyrosine to other amino acids are more frequent than complete deletion ITAM region in CD79A. CD79 A/B mutations are not required for the survival of antigen-dependent ABC DLBCL, however, they enhance proximal BCR signaling by increasing the expression of BCR on the membrane and blunting LYN activity [39, 44]. Chronic active BCR signaling is probably also driven by self-antigen in lymphoid malignancies, even in CLL [47] and MCL [48]. In about one third of ABC DLBCL cases, an immunoglobulin heavy chain variable segment V_H4-34 is utilized to bind sugars from the cell surface glycoproteins [49]. After mutating the BCR IgVH segments, ABC cell lines will die due to the failure of self-antigen binding [49]. Moreover, in the ABC cell line OCI-Ly10, signaling is activated by binding of the BCR to an antigen from apoptotic cells' debris [49]. IgM-BCRs are always expressed by B-cell lymphomas which depend on chronic active BCR signaling, even though their origin is from germinal center where B cells mainly express IgG-BCRs. Based on various evidences, it was proposed that IgM can be regarded as the initiating oncogene for malignant growth of normal B cells, but other additional oncogenic aberrations (e.g. *BCL2/MYC* translocation) are required to support the malignancies [50].

In 2011, *MYD88*, an adapter protein involved in Toll-like receptor (TLR) signaling, was found to be an essential gene in ABC DLBCL, not in GCB DLBCL by RNA inference screening [51]. RNA-sequencing of DLBCL tumor samples revealed that MYD88 is mutated in 39% of ABC cases, with the most recurrent point mutation being a leucine to proline change at position 265 ($MYD88^{L265P}$) which is absent in GCB and BL [51]. $MYD88^{L265P}$ has an oncogenic nature in ABC DLBCL since knockdown of it decreases both NF-κB activity and STAT3 phosphorylation in ABC cell lines and ectopic expression of $MYD88^{L265P}$ rescued survival of ABC lines following knockdown of endogenous $MYD88^{L265P}$. $MYD88^{L265P}$ and CD79B mutations have a higher co-occurrence rate than expected by chance. Chronic active BCR signaling and MYD88 signaling have a functional synergy based on the decreased proximal BCR signaling after knockdown of MYD88 in BCR-dependent ABC cell lines. In addition, patients with both $MYD88^{L265P}$ and CD79B mutations showed a relatively high response rate to the treatment with a BTK inhibitor (ibrutinib), however, those patients with MYD88 mutations only

or wild-type MYD88 did not benefit to the same extent, providing another evidence for a functional link between BCR and MYD88 signaling [39].

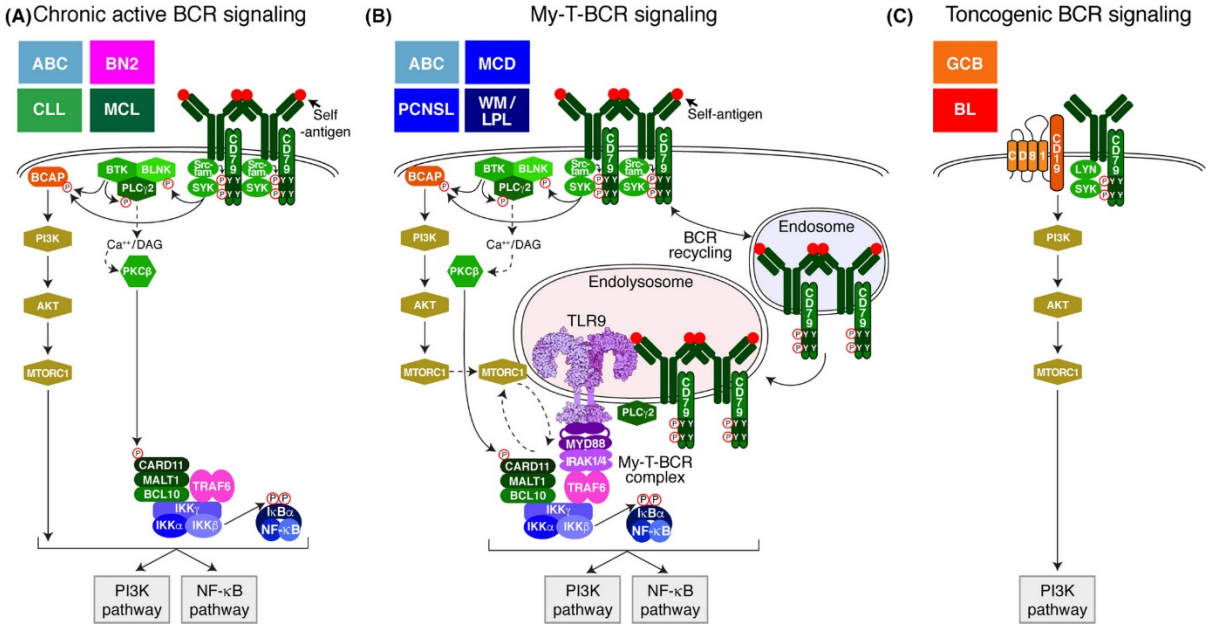


Figure 1.2 Mechanisms of BCR signaling in DLBCL.

This figure is adapted from Young and Staudt, 2019 (Immunological Reviews) [50]. (a) Chronic active BCR signaling. (b) My-T-BCR signaling. (c) Toncogenic BCR signaling.

TLR9 and its chaperones, UNC93B1 and CNPY3 were discovered to be essential for the survival of ABC DLBCL with MYD88^{L265P} and CD79B double mutations (namely MCD genetic subtypes) by whole-genome CRISPR-Cas9 screening [40]. Proteomic analysis of TLR9 interacting proteins identified its interaction with MYD88, CNPY3, CD79A and CD79B subunits of the BCR. This multiprotein signaling module was named as MYD88-TLR9-BCR (My-T-BCR) supercomplex and protein interactions within the complex could be visualized in situ by the proximity ligation assay (PLA) which is used to identify protein-protein interaction within 40 nm between each other as an imaging technology [52]. According to PLAs of the IgM:TLR9 interaction, the My-T-BCR complex localizes on endolysosomal intracellular vesicles, where TLR9 normally locates. Two different pools of BCR play a role in My-T-BCR signaling model. Proximal BCR signaling is firstly initiated at the plasma membrane by the continuous stimulation of the BCR molecules on the cell surface by self-antigens. One pool of BCRs is promoted to internalize and to translocate into Lamp1+ endolysosomes. Then, the BCR binds to TLR9 to form the My-T-BCR. On the other side, proximal BCR signaling activates PLCγ2 and PKCβ, which induce CBM complex formation. Next, the IKK pathway is activated by the CBM complex, together with IRAK1 and IRAK4 which are recruited by My-T-BCR. My-T-BCR formation decreases to a large extent after ibrutinib treatment, providing evidence that NF-κB pathway will be inhibited by ibrutinib. Unexpectedly, the My-T-BCR was found in close association with the mTORC1 complex which also locates at endolysosomes. mTORC1 is thought to be in charge of transducing PI3 kinase signaling to the downstream pathways [53]. Ibrutinib cooperates with mTORC1 inhibitors to kill ABC cell lines [54-56]. My-T-BCR was detected in ABC, primary central nervous system lymphoma (PCNSL) and Waldenström's macroglobulinemia (WM) biopsies by IgM:TLR9 PLAs, but not found in GCB, CLL, MCL and normal B-cells [40]. So the My-T-BCR complex could probably be used in diagnostics to identify malignant B-cells with chronic active BCR signaling.

Based on current knowledge, antigen-dependent NF-κB activation signaling is not the hallmark of GCB DLBCL and Burkitt's lymphoma. BCR, LYN, CD19 and CD81 were identified to be essential for the survival of one subset of GCB DLBCL and BL lines [40]. Moreover, PI3K is engaged as the downstream pathway for the survival of GCB DLBCLs since SYK inhibitors were effective in killing GCB DLBCLs and PI3K activity was significantly decreased after SYK knockdown [40, 57]. The signaling in GCB DLBCLs is similar to tonic BCR signaling, which promotes the survival of mature murine B-cells, due to that PI3K signaling was identified as the essential survival pathway in both cases [58]. Furthermore, some other evidences suggest that BCR signaling in GCB DLBCL seems to be different from tonic BCR signaling. The CRISPR methodology uncovered that the BCR coreceptor CD19 and its associated membrane protein CD81 are essential for the survival of BCR-dependent-GCB and BL cell lines, however, deletion of CD19 or CD81 did not decrease the numbers of mature B cells largely. Likewise, in GCB cell lines, BCR signaling was demonstrated to be LYN-dependent, whereas autoimmune B cells from LYN knockout mice showed hyper-activation of the pathway after BCR stimulation [40, 59, 60]. Recently, the new term of "toncogenic" BCR signaling was proposed for the oncogenic BCR signaling that occurs in an PI3K-dependent and NF-κB-independent pathway in GCB DLBCL and BL, which is distinct from tonic BCR signaling in normal mouse B cells and chronic active BCR signaling in ABC DLBCL [50].

1.1.4 Therapy of lymphoma by targeting BCR signaling

Inhibition of BCR signaling effectors involving either chronic active or toncogenic BCR signaling in lymphoma with drugs is currently common therapy strategy in the clinic. The non-receptor tyrosine kinases BTK and SYK, effectors in PI3K pathway and other essential oncogenic survival pathways are the more prospective targets for BCR pathway inhibition in lymphoma.

Ibrutinib, developed by Pharmacyclics, is an irreversible BTK inhibitor which covalently binds to cysteine 481 near the active site of BTK. Ibrutinib has a high on-target specificity due to the fact that only nine other kinases have cysteine residues in similar positions near their active sites. BTK plays a very important role in chronic active BCR signaling in ABC DLBCL cell lines, but is non-essential in toncogenic BCR signaling in GCB DLBCL cell lines [40, 44]. Early preclinical evidence showed that ibrutinib is highly toxic in ABC DLBCL models at low doses, but has no effect on GCB DLBCL cell lines even at higher concentrations. Later on, a clinical trial of ibrutinib monotherapy in patients with relapsed/refractory DLBCL demonstrated that 37% response rate was achieved in ABC DLBCL, compared to only 5% in the GCB DLBCL subgroup, providing evidence for a role of BTK in chronic active BCR signaling [39]. Moreover, patients with ABC DLBCL treated with ibrutinib have a median overall survival of 10.3 months compared to 3.3 months for GCB DLBCL patients, with three ABC DLBCL patients who survived over 2 years and one patient without disease after 8 years of ibrutinib monotherapy. ABC DLBCL genetic subtypes with certain mutations have distinct ibrutinib responses. ABC tumors with *CD79B* mutations showed higher responses than tumors with wild-type *CD79B* (55% vs 31%), supporting a crucial role of *CD79B* mutation in DLBCL pathology [44]. Meanwhile, many ABC patients with wildtype *CD79B* responded to ibrutinib monotherapy, indicating that chronic active BCR signaling in these tumors is initiated by other mechanisms, such as self-antigen engagement or other genetic abnormalities targeting negative BCR signaling regulators [44, 49]. Tumors with *MYD88* mutations only didn't have clinical response, however, a frequent response (4/5) to ibrutinib was discovered in ABC tumors with both $MYD88^{L265P}$ and *CD79B* mutations. The double-mutant genotype is a hallmark of the MCD genetic subtype and is also prevalent in extranodal aggressive lymphoma such as PCNSL, which obtained 77%-89% objective responses to ibrutinib monotherapy [61]. As the

My-T-BCR supercomplex was only discovered in ibrutinib-sensitive ABC DLBCL, it was suggested that detection of the My-T-BCR complex expression might be more predictive than genetic mutation analysis with respect to ibrutinib responses [40]. My-T-BCR to drive NF-κB pathway may be especially sensitive to ibrutinib in ABC tumors [40]. The My-T-BCR was also found in more than 80% of PCNSL tumors which was consistent with the responses rate to ibrutinib therapy in PCNSL. Therefore, the My-T-BCR might be an available biomarker and could complement genetic analysis for ibrutinib responses in lymphoma. Ibrutinib was already approved by the US Food and Drug Administration (FDA) to treat MCL, CLL and WM in 2013 due to its particular clinical responses in lymphoma as monotherapy. Recently, a second generation BTK inhibitor with a higher specificity-acalabrutinib- which also covalently binds to cysteine 481 of BTK was developed by Acerta Pharmaceuticals and AstraZeneca. It was proved to be similar in therapy efficacy as ibrutinib in CLL and was approved by the FDA to treat MCL in 2017 [62]. Both ibrutinib and acalabrutinib are reported to cause some adverse effects such as bleeding, atrial fibrillation and diarrhea in patients during therapy [63].

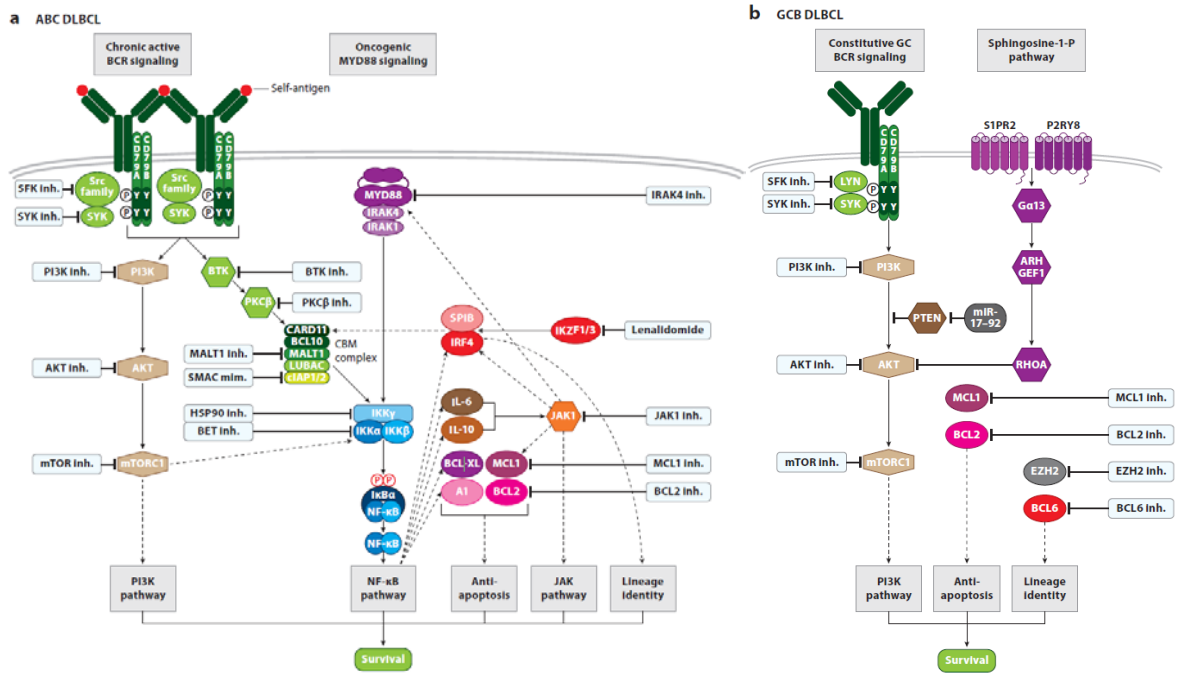


Figure 1.3 Oncogenic survival pathways and drug targets in DLBCL.

This figure is adapted from Young and Staudt, 2019 (Annual Review of Cancer Biology) [38]. (a) Distinct targets inhibitors in oncogenic survival pathways in ABC DLBCL. (b) Drugs targeting effectors in constitutive GC BCR signaling in GCB DLBCL.

SYK inhibitors were firstly paid attention to be studied for the treatment of all BCR-dependent lymphoma in vitro due to the involvement of SYK in both chronic active and toncogenic BCR signaling (**Figure 1.3**). Several known SYK inhibitors such as fostamatinib (R788), R406, PRT062607 and entospletinib (GS-9973) have limited selectivity for SYK and showed evident toxicity [40, 64] in the cell line models of BL [65], ABC DLBCL [44], GCB DLBCL [66] and CLL [67], so that further clinical development is currently precluded. Besides targeting BTK and SYK, many other agents which target the BCR-dependent NF-κB pathway in DLBCL have been developed in preclinical studies, such as inhibitors of PKCβ [68], CBM complex and IKK (**Figure 1.3**).

Lymphomas use a complex network of distinct tumor survival signaling pathways, such as at least five known pathways that promote tumor growth in ABC DLBCL: (1) NF-κB; (2) PI3K/mTOR; (3)

JAK1/STAT3; (4) BCL2 family mediated anti-apoptosis; (5) lineage-defining transcription factors. It is impossible to cure the patients by monotherapy targeting one effector in any of these pathways. Combination therapies by inhibition of parallel survival pathways in order to kill malignant cells cooperatively are without doubt a direction to be developed, under the premise of minimizing side effects and therapy resistance. Preclinical studies revealed that BTK inhibitor ibrutinib synergizes with inhibitors of SYK, BCL2, BET, IRF4, JAK1 and IRAK4, as well as chemotherapy agents to kill DLBCL cell lines [49]. In clinic, combination of ibrutinib with chemotherapy in the DA-TEDDi-R regimen to treat PCNSL patients makes 86% patients survive for more than 2 years, compared to the short progression-free survival of few months after ibrutinib monotherapy or single chemotherapy [61]. Integrated approaches for precision therapy of aggressive lymphoma will be the new direction in the future using the genomics and proteomics technologies.

1.1.5 Glycosylation in lymphoma

Glycosylation is one of the most common, complex and important modifications in organisms. Almost half of proteins expressed in mammalians are glycoproteins [69], which regulate key biological processes such as cell adhesion, cell signaling transduction, receptor activation, immune regulation, endocytosis, and cell proliferation [70, 71]. Aberrant glycosylation have been found to be related with many diseases like infectious and inflammatory diseases, diabetes, cancers and congenital diseases [72]. Currently many known clinical biomarkers are glycoproteins, for example, CA125 in ovarian cancer; CA12-3 and CA27-20 in breast cancer, PSA in prostatic cancer and so on [73]. Glycosylation refers to the covalent attachment of complex carbohydrates to specific amino acid residues on peptide sequences. The attachment of glycans to asparagine (Asn) and serine (Ser)/threonine (Thr) are the two most common and studied types of glycosylation, which were defined as N-linked glycosylation and O-linked glycosylation respectively. A consensus amino acid sequence (Asn-X-Thr/Ser/Cys (X is any amino acid except proline)) on polypeptide chains and a common trimannosyl chitobiose structure ($(GlcNAc)_2Man_3$) (GlcNAc: N-acetyl-glucosamine; Man: mannose) on glycan chains are the typical features of N-glycosylation. Other monosaccharides including Man, galactose (Gal), GlcNAc, fucose (Fuc), and sialic acids (SA), usually N-acetylneuraminic acid (NeuAC) or N-glycolylneuraminic acid (NeuGC) are attached to the core structure in Endoplasmic reticulum and Golgi to form the mature structure. N-glycans are divided into three main types: in the high-mannose type, only mannose units extend from the core structure; in the complex type, GlcNAc is attached to the core units followed by extensions with galactose and lactosamine structures and sialic acids at the ends (Fucose may be attached to any oligosaccharide on the glycan chain); the hybrid type is the combination of complex and high-mannose type (**Figure 1.4**) [74]. However, no consensus motif on peptide sequences and the substitution of the trimannosyl chitobiose core by eight types of core structures are found in O-glycosylation.

Many research reported that glycosylation was a key regulator for the metastasis of mouse lymphoma cells using mouse lymphoma cell lines. Until now, there are only a few reports of the role of glycosylation in human lymphoma cell metastasis [75-79]. The B-cell receptor, which consists of the glycoproteins immunoglobulin M or G (IgM or IgG), is central to the development and maintenance of normal B-cells and co-regulates the malignant growth and survival of B-cells. Aberrant N-glycosylation in CD79a, IgG, IgM and CD95 has been found to be associated with lymphoid malignancies [80-82]. In some ABC DLBCL cases, immunoglobulin heavy chain variable (IgVH) of BCR binds to itself glycans as self-antigen instead of foreign antigen to promote BCR micro clustering and initiate downstream BCR signaling, including the PI3K, MAP kinase, NF-AT, and NF-κB pathways. The self-antigen involvement to maintain malignant B cells survival was validated by the induced death of ABC cells after the mutation of BCR IgVH to inhibit self-antigen binding. In addition, many other cell surface

glycoproteins such as CD22, CD166, CD44, CD48 and CD54 interact with BCR and affect the BCR signaling which is mediated by their N-glycan. For example, the heavily glycosylated CD22 molecule inhibits BCR signaling after binding to Galectin-9, whereas, the inhibition of BCR signaling by interacting with CD22 decreases after mutation of glycosites of CD22 [83, 84].

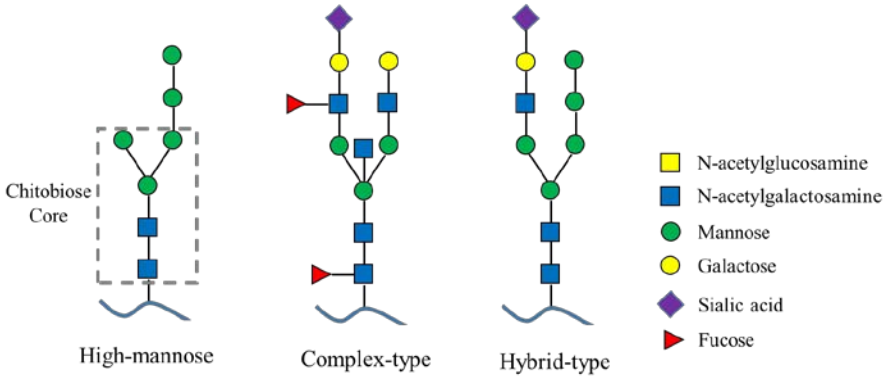


Figure 1.4 Main subtypes and the core structure of N-glycans.

In malignant tumors, aberrant glycosylation such as sialylation and fucosylation occur in most cases due to the aberrant expression of fucosyltransferases and sialyltransferases. Sialic acid was found to be related with tumor cell aggressiveness in lymphoma due to the higher expression of sialic acid in the serum of lymphoma patients [85]. Moreover, the loss of *Phaseolus vulgaris* leukoagglutinating lectin (L-PHA) reactive oligosaccharides and their α2, 6-sialylation are closely related with a worse prognosis for DLBCL patients [86]. Cell adhesion to extracellular matrix in the context of lymphoma cell metastasis was enhanced by the knockdown or neuraminidase-mediated inhibition of β-galactoside α-2, 6-sialyltransferase (ST6Gal1) in the anaplastic large cell lymphoma cell line H-ALCL [77, 78, 87]. On the other hand, the sialylated glycan on integrins regulates integrin signaling in lymphoma, resulting in increased invasiveness of lymphoma cells [78]. We can say that sialic acid is related with cell invasion and metastasis in lymphoma.

1.1.6 Phosphorylation in lymphoma

Phosphorylation is one of the most common and important PTMs whose discovery went back to 1906 by Phoebus Levene [88]. It is a reversible process which involves transfer of a phosphate group from adenosine triphosphate (ATP) or guanosine triphosphate (GTP) to a specific amino acid residues. Most of the phosphorylation events happen on Ser, Thr and Tyr, of that, Ser is the most commonly phosphorylated amino acid residue, followed by Thr, whereas tyrosine phosphorylation is relatively rare. Phosphorylation is an important regulatory mechanism. Many enzymes and receptors are activated and deactivated via phosphorylation and dephosphorylation events by various kinases and phosphatases. Phosphorylation events which are regulated by more than 500 protein kinases and 150 phosphatases in humans, are vital for various biological processes such as proliferation, differentiation, apoptosis [89]. Most of the protein kinases are activated by the kinase itself with a cis-phosphorylation or auto-phosphorylation followed by the activation of a series of downstream phosphorylation events. In general, the phosphorylation of different amino acid residues requires specific kinases, such as, STKs (serine/threonine kinases) are only responsible for the phosphorylation of serine and threonine, TKs (tyrosine kinases) act on tyrosine. Phosphatases are responsible for the removal of phosphate group from phosphoproteins by hydrolyzing phosphoric acid into a phosphate group and a molecule with a free hydroxyl group.

Dysregulated phosphorylation is involved in many abnormal cellular processes in various cancers. Almost all signaling pathways are mediated by multiple kinases, such as, tyrosine kinase and MAPK, which are responsible for the phosphorylation of effectors and auto-phosphorylation. Abnormal kinase expressions are found in many cancers and cause changes of the phosphorylation status of downstream effectors. NF-κB pathway, which is essential for cell proliferation and inhibition of apoptosis is a phosphorylation-mediated process. NF-κB activation is involved in the chronic active BCR signaling in ABC DLBCL. In normal B cells BCR signaling pathway involves a series of tyrosine kinases and their regulated phosphorylation. Briefly, BCR signaling is initiated by the phosphorylation of tandem tyrosine residues within ITAMs located on CD79A and CD79B by the SRC-family kinases which is a family of non-receptor tyrosine kinase (LYN, FYN and BLK). The dually phosphorylated ITAMs then recruit the tyrosine-protein kinase SYK, also known as spleen tyrosine kinase, by its tandem SRC homology 2 (SH2) domains leading to the activation of SYK. B-cell linker protein encoded by the *BLNK* and BTK are recruited and bind to the auto-phosphorylated SYK which induces BTK activation. Activated BTK phosphorylates PLCγ2 at Y753 and Y759. Activated PLCγ2 hydrolyses PIP2 into inositol triphosphate (IP3) and diacylglycerol (DAG). IP3 and DAG furtherly regulate other proteins. They activate the signaling cascades in NF-κB, NF-AT and MAPK pathways which promote B cell survival, proliferation and differentiation [90]. Genetic alterations of many kinases and adaptors involving BCR signaling in normal B-cells are discovered in lymphoma, which results in dysregulated and constitutive oncogenic signaling [30].

1.1.7 Ubiquitination in lymphoma

Ubiquitination is another common PTM found in the majority of eukaryotic organisms. It is a process in that a 76 amino acids polypeptide (ubiquitin) is attached to via a C-terminal glycine the ε-amino group of lysine in a substrate protein by the isopeptide bond linkage [91]. Ubiquitination is a three-step enzymatic process, which starts with ATP-dependent ubiquitin activation by a ubiquitin-activating enzyme (E1), followed by the transfer of activated ubiquitin to a ubiquitin-conjugating enzyme (E2), and final conjugation of the ubiquitin to a lysine residue in the target protein by a ubiquitin-protein ligase (E3) [92]. Each of the seven lysines in ubiquitin enables with the attachment of another ubiquitin polypeptide, resulting in the formation of poly-ubiquitin. M1-linked poly-ubiquitin which means the amino-terminal methionine of one ubiquitin is linked with another ubiquitin was found [93]. Meanwhile, deubiquitinating enzymes (DUBs) regulate the removal of ubiquitin from proteins in human [94]. The proteasome-dependent degradation is mainly mediated by K48- or K11-linked poly-ubiquitination, whereas, mono- or multi-ubiquitination, i.e., a single ubiquitin moiety attached to one or more lysine residues in a protein, is related to vesicle trafficking, protein location and protein-protein interaction.

The BCR-dependent NF-κB pathway, which is the main regulatory mechanism for cell survival in ABC DLBCL, is mediated by ubiquitination. For example, the mono- or poly-ubiquitination of MALT1, the linear ubiquitin chain assembly complex (LUBAC) and IKKγ mediate the activation of the NF-κB pathway [95]. Inactivating mutations of A20 and cylindromatosis (CYLD), as two important DUBs, were also shown to be involved the NF-κB activation in many types of lymphomas [96]. However, the knowledge of the functions of ubiquitin enzymes and the regulatory mechanisms of ubiquitination in lymphomas are still limited.

1.2 Mass spectrometry-based proteomics

Mass spectrometry-based proteomics is a robust and powerful tool to analyze almost all proteins in one sample and mainly serves the identification and quantification of proteins, their post-translational

modifications (PTMs) and interactions with high sensitivity and selectivity [97]. It enables to decipher the molecular mechanisms of diseases by mining regulated effectors mediating signaling pathways in cancers in cooperation with other omics techniques, such as genomics, transcriptomics and so on.

1.2.1 Tandem Mass spectrometry

The general principle of mass spectrometry (MS) is to detect the ionized analytes qualitatively and quantitatively by their individual mass to charge ratio (m/z) and abundance in the gas phase. Ion source, mass analyzer and detector are the requisite parts for a mass spectrometer. Two soft ionization techniques, namely matrix-assisted laser desorption ionization (MALDI) and electrospray ionization (ESI) are the most commonly used methods to ionize analytes. MALDI was firstly introduced in 1985 by Franz Hillenkamp, Michael Karas and their colleagues [98] MALDI uses laser irradiation to ionize analyte molecules out of matrix material. ESI uses a high voltage between the tip of a stainless steel or quartz silica capillary and the entrance of the mass spectrometer to make the liquid form an aerosol [99]. ESI was chosen as the ionization method for liquid chromatography coupled mass spectrometry (LC-MS) because samples can be ionized out of a liquid phase eluting from the LC column. The ESI and MALDI inventors John B. Fenn and Koichi Tanaka were jointly awarded the Nobel Prize in Chemistry in 2002. The mass analyzer plays a vital role in the acquisition of m/z with respect to mass accuracy. There are few commonly used mass analyzers in the proteome analysis, such as, time-of-flight (TOF), ion traps (cubic or linear ion trap), quadrupole, orbitrap and Fourier-transform ion cyclotron (FT-ICR). Each analyzer has its own benefits and drawbacks, mainly reflecting in mass accuracy, sensitivity, resolution and dynamic range. For example, orbitrap and TOF own a high resolution, IT and quadrupole belong to the low resolution analyzers. Tandem mass spectrometry, also named as MS/MS, is a two-step technique used for a sample analysis either by using two or more connected mass spectrometers or a single mass spectrometer in which few analyzers are arranged one after another. For example, the Q Exactive HF mass spectrometer includes quadrupole and orbitrap; Orbitrap Fusion and Lumos instruments combine linear ion trap, quadrupole and orbitrap. Since orbitrap was used as the mass analyzer for the acquisition of all MS raw data files in this study, I will give an introduction to orbitrap in more details. Orbitrap was developed by Alexander Makarov in 2000 based on the Kingdon trap [100]. It consists of a central spindle-like electrode and an outer barrel-like electrode which is split into half by an insulating ceramic ring. Trapped ions in the electrostatic field move around and oscillate along the central electrode, whose trajectory in space seems like helices. The axial motions of ions only depend on their m/z, and are not related to other parameters of the ions such as initial energy, angels and positions. The frequencies of oscillation are detected as the image current via an amplifier between the two halves of the outer electrode and are recorded as ion's m/z by Fourier transform [101]. Later, a compact high-field orbitrap was invented by reducing the dimensions including inner diameter of outer and inner electrode. A higher field strength was obtained by increasing the ratio of inner to outer electrode, leading to a higher resolution (almost double) at the same transient time. It was used as the mass analyzer in Orbitrap Velos and all the later-developed Thermo Orbitrap series mass spectrometers [102-104].

1.2.2 Peptide and protein identification by MS

Bottom-up proteomics (or shotgun proteomics) is currently the most widespread strategy by analyzing the enzymatic digested peptides from proteins. In traditional bottom-up workflow (**Figure 1.5**), proteins are digested into peptides by specific enzymes, e.g., trypsin, Glu-C. After separating the peptides mixtures by reverse-phase chromatography, the ionized peptides by ESI fly into the mass spectrometer and move to the mass analyzer under the vacuum environment. The instrument firstly scans all the

peptide ions and the accurate mass of peptide ions is recorded in one spectrum, the so called MS1 spectrum. Subsequently, the precursor ions are isolated and fragmented. Only few different fragmentation methods have been developed until now. Collision induced dissociation (CID) is the most common fragmentation method. For this, the isolated peptide ions collide with an inert gas like nitrogen and helium, resulting in breaking of peptide bonds (-CO-NH-) to generate N-terminal b-ion and C-terminal y-ion series [105, 106]. Two types of CID fragmentation are used indeed, ion trap CID and high energy collisional dissociation (HCD). Neutral loss of H_2O, NH_3 and H_3PO_4 and so on sometimes happens when performing peptides fragmentation by ion trap CID. Moreover, ion trap can't detect the low mass fragments, resulting in a low mass cut-off [107]. Higher energy collisional dissociation (HCD) was firstly introduced in 2007, which takes place in separate collision cells instead of in the analyzer [108]. Likewise, b- and y-ions are obtained after peptides fragmentation by HCD. Due to the higher energy for the collision, it reduces neutral loss of molecules. In general, the fragments ions are detected in the Orbitrap analyzer at high resolution and mass accuracy. Electron transfer dissociation (ETD) is another common fragmentation method which is useful for entire protein or longer peptide fragmentation [109]. During ETD, highly positive charged peptide ions obtain an electron from radical anions, resulting in fragmentation that occurs along the peptide backbone at a NH-Cα bond to generate c- and z- type fragment ions while leaving post translational modifications intact [110, 111]. All fragment ions are recorded in a spectrum called MS/MS or MS2 which is used to decipher peptide sequence by de novo sequencing initially [112]. Peptide identification is achieved by matching the experimental spectra with theoretical fragment ion spectra from a sequence database after in silico digestion of proteins which is done automatically by searching platforms [97, 113].

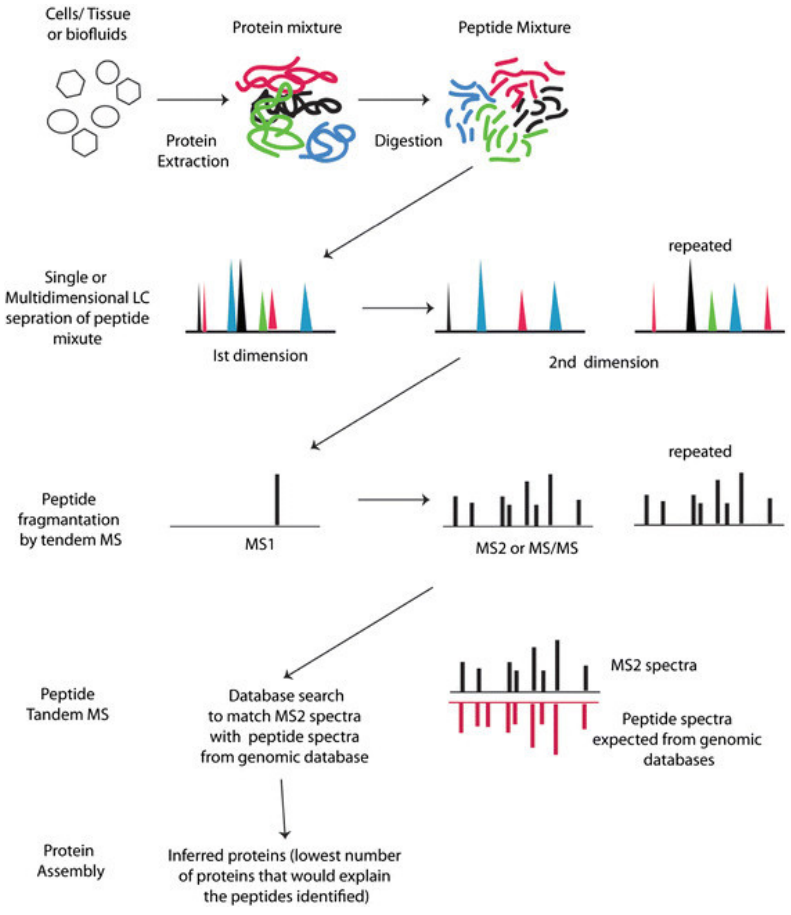


Figure 1.5 Workflow of bottom-up proteomics.

This figure is adapted from Bhargava, M., 2014 (Clin Transl Med) [114].

With respect to the selection of precursor ions for MS2 analysis, data-dependent acquisition (DDA) mode was used in this study. In the DDA mode [115], the mass spectra of all the peptide ions that co-elute at a specific time point in the gradient elution are recorded at the MS1 level. In typical, top N intense precursor ions from one MS1 scan are selected in a first stage of tandem mass spectrometry. Then they are fragmented and analyzed at the MS2 level.

1.2.3 Quantitative proteomics

Quantitative proteomics is becoming a crucial and basic tool in biological or clinical research to detect protein levels in different samples or from same samples under different conditions. Due to the varying peptide ionization efficiency [116, 117] and the detectability of peptides in a mass spectrometer, the intensity of the peaks in a mass spectrum cannot be used to compare the different peptide amounts in a sample directly. For comparison of the peaks of same peptides with same ionization efficiency among different samples, relative quantitation has been developed in the last decades. Stable isotope labeling and label free are the methods used for relative quantitation. Label-free quantitation is based on measuring the MS1 signal of peptides without modifying the peptides by any tag. Spectral counts and peak area of each peptide are the two main approaches for comparison [118, 119]. Some major limitations in label-free method exist till now: e.g., 1. Lower throughput due to the necessity for multiple replicates for each sample; 2. Poor precision based on the large protein coefficient of variations between replicates, especially for less abundant proteins; 3. Problems of missing values in replicates [120]. In stable isotope labeling methods, a known quantity of the analog, which is synthesized with nonradioactive isotope labels such as ^{13}C, ^{15}N and ^{2}H, is introduced into the sample. Due to the same ionization efficiencies, the peptides with or without isotopes obtain the same MS signaling responses in the mass spectrometer, only balance by the mass difference introduced by the stable isotopes [121]. It can be divided into chemical labeling, metabolic labeling and enzymatic labeling further. Enzymatic labeling is performed by digestion of proteins in the environment of isotopically enriched water (^{18}O). One or two ^{18}O atoms can be incorporated into peptides resulting in a mass shift of 2 or 4 Da between two samples in the MS1 spectrum [122]. Isotope-coded affinity tags (ICAT) [123], isobaric tags for relative and absolute quantitation (iTRAQ) [124] and isobaric tandem mass tags (TMT) [125] are the commonly used chemical labeling strategies (**Figure 1.6a**). In general, a mass tag is incorporated into the peptides in vitro and attached to the specific amino acid group. In the case of TMT labeling, the TMT reagent consists of a mass reporter, a mass normalizer and an amine-active NHS-ester group. The mass reporter and mass normalizer in each isobaric label reagent with the same mass remain attached to the peptides during MS1 analysis and the tagged peptides from different samples represent as a signal peak in MS1 scan. After fragmentation, the abundance of reporter ions reflects the relative ratio of the peptides in the combined samples. The common fragmentation method in MS2 level induces incorrect quantitation involving ratio distortion due to the interference of co-eluted contaminant peptides into targeted peptides. Detection of reporter ion intensity in an MS3 spectrum obtained from one of the TMT labelled MS2 fragment ions is able to eliminate the ratio distortion, however, the sensitivity decreases as a penalty [126]. With the advent of Orbitrap Fusion capable of synchronous precursor selection (SPS) in 2014, the sensitivity problem in the MS3 method was eliminated by using isolation waveforms with multiple frequency notches for synchronous precursor selection of multiple MS2 fragment ions [127]. Therefore, TMT based quantitation using the SPS-MS3 strategy wass adopted for intact N-glycoproteome quantitation in this study. Metabolic labeling enables the introduction of the mass tag into the samples in vivo during cell cultivation and growth (**Figure 1.6b**). Stable isotope labeling with amino acids in cell culture (SILAC) is one of the most common metabolic labeling methods for proteome quantitation with high labeling efficiency, low labeling error and high reproducibility, which

is prevalently applied to investigate the dynamics in signal transduction, protein interaction, PTMs or for cancer biomarker identification. In general, cells are cultured in medium supplemented with the isotope-labelled amino acids arginine (R) and lysine (K) instead of the natural ones for at least 5-6 generations to achieve complete incorporation of the isotope-labelled amino acids [128-130]. The quantitation among different labelled samples is based on the intensity of precursors in MS1 spectrum during MS analysis of the mixed samples. Triple labeling was performed in conventional SILAC experiments, for which the cells were cultivated in light (K+0, R+0), medium (K+4, R+6) and heavy (K+8, R+10) culture mediums as three different conditions. The quantitation bias decreases to a large extent compared to chemical labeling because of the early mixing of different samples before protein digestion. We adopted the SILAC strategy for the phosphoproteome and ubiquitinome quantitation in this work.

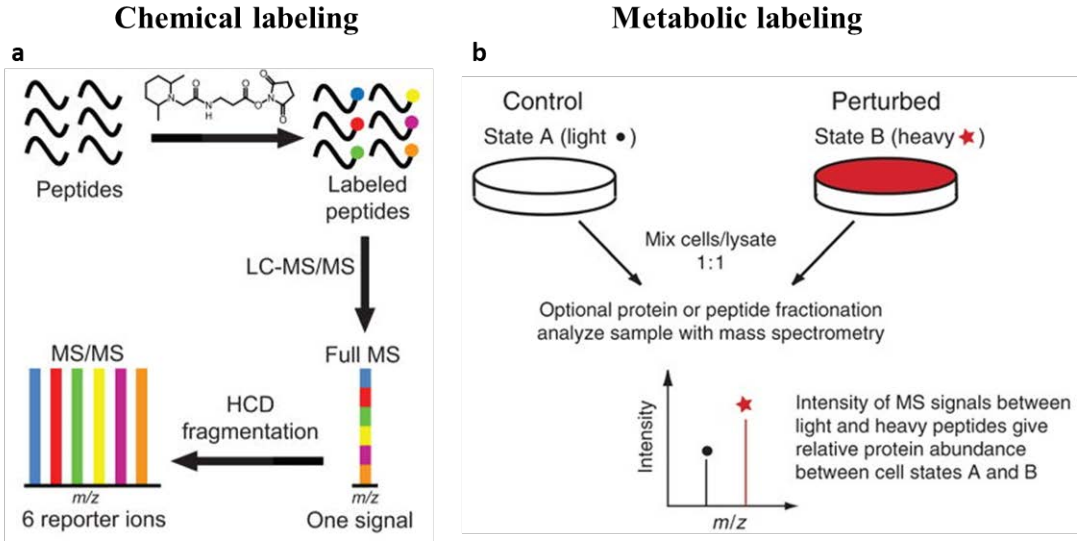


Figure 1.6 Schematic representation of chemical labeling and metabolic labeling workflow.

(a) Chemical labeling workflow. This figure is adapted from Ting et al., 2011 (Nat Methods) [126]. (b) Metabolic labeling workflow. This figure is adapted from Ong and Mann, 2007 (Nature Protocol) [129].

1.3 Post-translational modification analysis by MS

Post-translational modifications (PTMs) occur in almost any protein and increase the complexity of proteins to a large extent. PTMs are defined based on the functional groups that are covalently or enzymatically attached to the amino acid residues during protein synthesis. Till now, more than 200 PTMs have been found in organisms, affecting proteins' conformations in biological systems. Many diseases are derived from the deregulation of the modifications, e.g., aberrant phosphorylation in the BCR signaling pathway in lymphoma. MS based proteomics has been demonstrated to be a powerful tool to qualitatively and quantitatively analyze various PTMs in a systematical level regardless of the low abundance of most PTMs [131, 132]. In this study, we focused on three PTMs analysis, including phosphorylation, glycosylation, ubiquitination. Each of them will be described in more details in the following with respect to the biological function, sample preparation strategy and mass spectrometric method.

1.3.1 Glycosylation

1.3.1.1 Characterization of MS-based N-glycoproteomics

Due to the complexity of glycan structures on glycopeptides, glycosylation research becomes much more difficult than other modifications. In the early days, due to the limitation of technology, people mainly focused on the deglycopeptides level, whose glycans are removed enzymatically or chemically. Worthwhile, the released glycans are used for the interpretation of the composition and structure of the glycan chains, which is named as glycomics. However, the information of linkage between specific glycans and the position of the modified amino acid residues in proteins is lost. The final goal is to accomplish the identification and quantitation of intact glycopeptides which means the protease-digested peptide still linked to its complete glycan moiety. This is defined as glycoproteomics or site-specific glycoproteomics. The difficulties for characterization of intact glycopeptides in MS are listed as following: 1. Low abundant expression of the majority of glycoproteins in organisms and very low percentage (2%~5%) of glycosylated peptides compared to non-glycosylated peptides [133]. 2. The low ionization efficiency of intact glycopeptides and easy suppression by non-glycopeptides. 3. Micro-heterogeneity, namely a single glycosite is modified by many different glycan chains. The term glycoform defines a specific glycosite in a specific protein which a specific glycan moiety is attached regardless of missed cleavage of digested glycopeptides. For example, the 46^{th} asparagine in IGHM and 380^{th} asparagine in PTPRC are modified by more than 100 types of glycan moieties and the abundance among each glycoform varies greatly. 4. Macro-heterogeneity, glycosite is not always occupied by glycans. 5. Complex and insufficient fragment ions information in the MS spectrum obtained from the fragmentation of intact glycopeptides. 6. The diverse linkages among monosaccharides on glycan chain like α or β linkage, 1-3 or 1-4-linked. 7. Lack of powerful search engines for confident identification of intact glycopeptides. In the past few years, site-specific N-glycoproteomics research has been developing rapidly along with the breakthroughs in the following fields: N-linked glycopeptide enrichment strategy, intact glycopeptide fragmentation method and new search engine for the interpretation of spectrum of intact glycopeptides. All above mentioned progresses will be described in more detail.

1.3.1.2 N-linked glycopeptide enrichment strategy

Due to the low abundance, the micro-heterogeneity and the low ionization efficiency of glycopeptides, a highly efficient glycopeptide enrichment from complex samples is the first and critical step for glycoproteome research. In the past few years, few methods have been developed to use for glycopeptide enrichment successfully, such as lectin affinity [134, 135], TiO_2 affinity [136, 137], hydrazide chemistry [138, 139] and hydrophilic interaction chromatography (HILIC) [140, 141] (**Figure 1.7**). Lectins are ubiquitous in nature as glycoproteins, which have a high specificity of binding sugar groups. Different lectins recognize specific monosaccharides or a carbohydrate moiety from a part of sugar chains on glycoproteins. For example, concanavalin A (ConA) binds to high mannose type N-glycan specifically, while RCA recognizes galactose exclusively. In order to obtain more types of glycopeptides, several lectins are used together for glycopeptide enrichment in general [134, 142]. Titanium dioxide specifically attaches to the negatively charged sialic acids moiety at the end of N-glycan chains by multidentate binding to serve the purpose of glycopeptide enrichment. In hydrazide chemistry, carbohydrate cis-diol groups are oxidized to aldehydes by periodate followed by hydrazone formation between aldehydes and hydrazide groups. Deglycosylated peptides are collected by the use of PNGaseF for cleaving N-glycan from glycoproteins after removing the non-glycosylated peptides by digestion of immobilized proteins. The lack of intact glycopeptides is the drawback of the hydrazide

chemistry method. The principle of HILIC is that the hydrophilic analytes are retained to the water-enriched liquid layer established on the stationary phase by the hydrophilic partitioning. Due to the hydrophilic polyhydroxyl groups of the glycan chains, glycopeptides are retained and nonglycosylated peptides are washed away by high organic mobile phase. The commonly used silica-based ZIC-HILIC stationary phase carries a covalently bonded, zwitterionic, functional group of the sulfobetaine type including a quanternary amine cation with one positive charge and a sulfonate anion with a negative charge. The formation of electrostatic interactions between analytes with quanternary amine cation and sulfonate anion provides the secondary retaining mechanism. Later on, people found that the addition of trifluoroacetic acid (TFA) to the mobile phase for glycopeptide enrichment enables to decrease the coenrichment of nonglycosylated peptides, whose hydrophilicity is overlapped with some of the glycosylated peptides. TFA as ion-paring (IP) reagent neutralizes charged peptides by protonating the charged amino acids. Therefore, the electrostatic interactions between peptide and the stationary phase are impaired and the hydrophilicity of peptides decreases by lowering the hydrogen bonding potential. However, because of the existence of numerous hydroxyl groups, the hydrophilicity of glycosylated peptides is affected in a lower extent compared to nonglycosylated peptides, which results in a larger hydrophilicity difference [140]. IP-ZIC-HILIC strategy enables the intact glycopeptide enrichment with high specificity. In addition, it was proved that more glycopeptides were enriched using IP-ZIC-HILIC than other enrichment methods under the same condition [143, 144]. In this study, IP-ZIC-HILIC was adopted as the single strategy for glycopeptide enrichment.

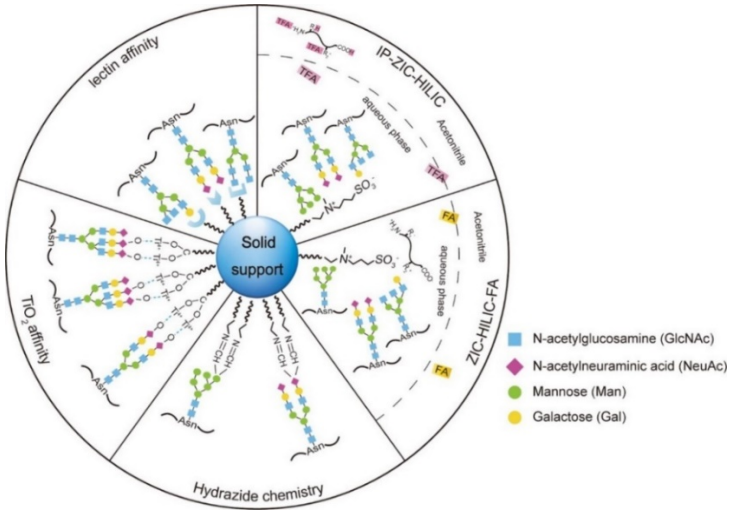


Figure 1.7 Main N-glycopeptides enrichment methods.

This figure is from Zhang et al., 2016 (J Proteome Res) [143].

1.3.1.3 Characterization of intact glycopeptide fragmentation in MS

Large scale site-specific N-glycoproteomics analysis is restricted by the quality of the MS spectrum. On the spectrum, the complex and incomplete fragment ion information from intact glycopeptides is recorded. For example, besides the regular peptide ion series of b/y, c/z, it still includes the intact peptide sequence plus part of the glycan chain ion series of B/Y and C/Z, oxonium ions from dissociated single or multi-monosaccharides, and A/X type ions from the cross-ring cleavages of monosaccharides (**Figure 1.8**) [145]. However, only limited fragment ions are actually obtained after glycopeptides dissociation in the mass spectrometer based on the preference of different ion dissociation methods [146-148]. When performing intact glycopeptide fragmentation by CID in an ion trap [149, 150], more B and Y ions and few b/y ions are acquired due to the preference of glyosidic bond cleavage compared to amide bond cleavage. Moreover, the lack of diagnostic oxonium ions, the cut-off in the low m/z

region and lower resolution and mass accuracy limit the intact glycopeptide identification in a systematic level. HCD enables to produce abundant diagnostic oxonium ions, partial B- and Y-ion series for glycan chain assignment and solid b/y ions for confident peptide identification. Successful large-scale intact glycopeptide identification performed by HCD has been reported by many research groups [151-153]. Moreover, analysis using step collision energy in HCD enables to provide more complementary fragment ions in one single spectrum for intact glycopeptide identification [154, 155]. ETD enables to detect the c/z ions for peptide sequence identification and glycosylation site assignment, but provides only little information on glycan chain composition due to the difficult fragmentation of the glycan moiety. Recently, Coon reported AI-ETD, which is the combination of simultaneous vibrational activation from infrared (IR) photon bombardment and electron-driven dissociation via ETD, for intact glycopeptide dissociation. In the study, more than 5600 unique glycoforms were identified in mouse brain [156]. In fact, higher energy collision dissociation-product dependent-activated ion electron transfer dissociation manner (HCD-pd-AI-ETD) was used in this study. HCD was performed firstly for intact glycopeptide fragmentation. Once glycan oxonium ions were detected in the HCD MS/MS scans, the same precursor ion was triggered for AI-ETD fragmentation. Longer duty cycles is needed and IR photoactivation accessory must be installed on the mass spectrometer separately in advance. Moreover, the availability for glycopeptide quantitation using AI-ETD is still waiting for further investigation. Recently, electron-transfer/higher-energy collision dissociation (EThcD) was applied to intact glycopeptide characterization which incorporates the main glycan fragment ions from HCD and most peptide fragments from ETD into one spectrum [157-159]. Last year, one paper submitted on bioRxiv showed the large-scale glycoproteome analysis in human serum and brain tissue using HCD and EThcD for glycopeptide fragmentation. It was performed using a similar strategy to HCD-pd-AI-ETD, only replacing AI-ETD for secondary product ion mass scan by EThcD [159]. Associated problems such as the longer duty cycle, unknown availability for quantitation and limited amount of search engines for the deciphering of EThcD data still need to be tackled. In this study, we developed the SPS-MS3-HCD method for glycoproteome analysis, in which the comparable glycopeptide identification was obtained based on step collision energy HCD, but allowed us to acquire more accurate quantitation information than using MS2-HCD method.

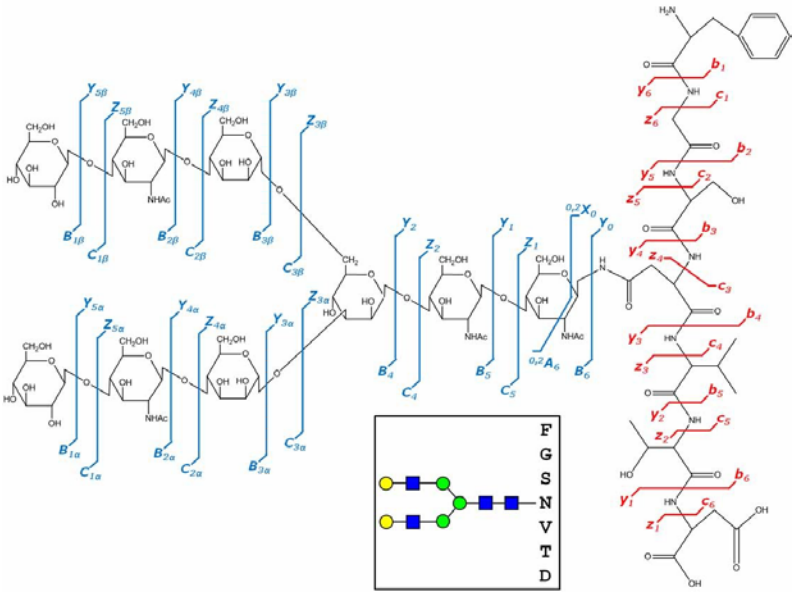


Figure 1.8 Diagram of fragment ions types dissociated from the intact glycopeptide.

It is from the composition of FGSNVTD+$GlcNAc_4Man_3Gal_2$ using different ion dissociation methods. This figure is adapted from Eric D. Dodds, 2012 (Mass Spectrom Rev) [145].

1.3.1.4 Search engine development for intact glycopeptide analysis

Similar to common peptides identification using database search-based methods, eg, SEQUEST, Mascot, X! Tandom etc, almost all the developed software for intact glycopeptide analysis also adopted the database search-based strategy. However, separate identification of peptides and glycans is necessary for confident glycopeptide identification. The most frequently used glycan databases are GlycomeDB which is a huge and comprehensive database of carbohydrates [160] or the database from the consortium for functional glycomics (http://www.functionalglycomics.org/). Several search engines for intact glycopeptide identification have been developed. GlycoPepGrader [161], GlycoPeptideSearch [162], Sweet-Heart [163] and MAGIC [164] are adapted to assign glycopeptides from low resolution CID spectra, whereas GlycoPepDetector [165] and GlycoPep Evaluator [166] are used for glycopeptide identification based on ETD spectra. Due to the low quality of CID and ETD spectra for limited intact glycopeptide fragment ions, a high false positive identification rate always occurs in the result when searching using the above mentioned software. GlycoFragwork [167] and GlycoMaster DB [168] are used for the identification of intact glycopeptides using the combination of several fragmentation types. Byonic [169] as a commercial software is able to search intact glycopeptides from any fragmentation method. GPQuest [153, 170] is able to assign glycopeptides on HCD spectrum. It needs experimental spectral libraries, including a peptide database from deglycoproteome analysis and a glycan database from glycomic analysis, for the matching of HCD-fragmented glycopeptide spectrum. SugarQB [151] is an algorithm used for analyzing glycopeptide spectrum from HCD. $Y1^+$ [peptide+ HexNAc]$^+$ fragment ion is identified by iteratively changing the precursor-ion mass of a given MS2 spectrum by subtracting all glycan masses considered. Then the peptide sequence is assigned against a protein database using the conventional MS/MS search engine, with the setting of HexNAc moiety as a variable modification to any asparagine, serine and threonine residue.

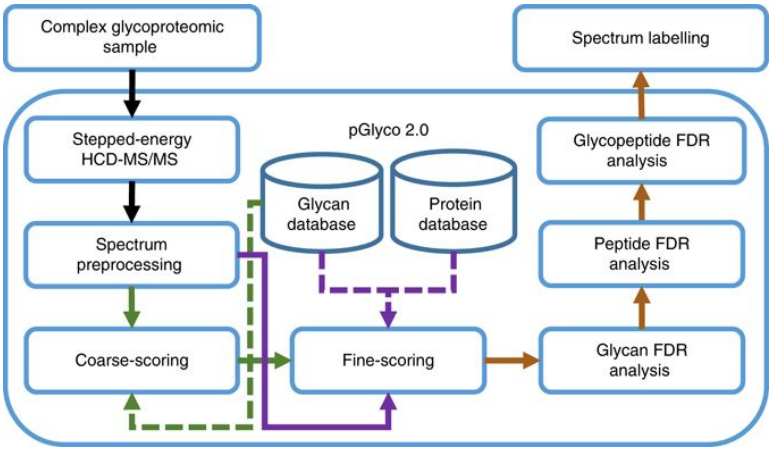


Figure 1.9 Schematic workflow of pGlyco 2.0.

This figure was adapted from Liu et al., 2017 (Nat Commun) [154].

pGlyco 2.0 [154] was developed for glycopeptide identification by using the fragment ions on Step collision energy (SCE)-HCD MS spectrum but is not limit to this fragmentation type. Each spectrum was used to identify glycan candidates by scoring first against the glycan database from GlycomeDB and then against the protein database to identify the candidate peptides. Comprehensive false positive identification rate (FDR) control was performed followed by spectrum annotation automatically (**Figure 1.9**). It is the first software for intact glycopeptide identification which considers the comprehensive quality control for three levels' FDR analysis of glycan, peptide and glycopeptide separately. A large scale glycoproteomics analysis in mouse tissues was achieved by pGlyco 2.0

searching under the strict FDR control. There is also reported software for intact glycopeptide analysis which is available for the assignment of glycopeptides on HCD spectrum, such as, Integrated GlycoProteome Analyzer (I-GPA) [171], pMatchGlyco based on open mass spectral library search [172], GlycoProteomics Analysis Toolbox (GlycoPAT) [173]. Except for pGlyco, all above mentioned search engines don't consider the independent glycan FDR control, easily resulting in a lower FDR obtained from each software compared to the virtual ones. No matter which search engine is used for intact glycopeptide identification, the quality of mass spectrometry spectrum which contains as much fragment ions information as possible, such as oxonium ions, peptide ions and peptide plus partial glycan ions, is important for the confident identification of intact glycopeptides. After incorporating the quantitative information, the MS spectrum becomes much more complex and more modifications need to be investigated to improve spectrum quality. In this study, we used pGlyco as the core search engine for intact glycopeptide identification, supplementing other software to achieve the quantitation of glycopeptides.

1.3.1.5 Intact glycopeptides quantitative methods

Due to the heterogeneity of protein glycosylation, quantitation of intact glycopeptides is still very difficult and complicated at present, even on the premise of recent huge developments in intact glycopeptides identification. In the past few years, there are some publications about the quantitation of intact glycopeptides by label-free approach [171, 174, 175], metabolic labeling and isobaric chemical labeling strategies. The existing problem in each glycoproteomics quantitation strategy is described in the following. Label-free method uses MS1 extracted ion currents (XICs) or spectral counts for the quantitation. Due to the varied ionization efficiency of glycopeptides owing to different glycan moieties and lower abundance of glycopeptides, serious MS response variations and missing values in data-dependent acquisition analysis (DDA) more frequently occur compared to common peptide quantitation by label-free approach. Recently, a data-independent acquisition (DIA) method was used for the quantitation of the O-glycoproteome with less missing values and higher sensitivity [176]. However, it still cannot be used for the N-glycoproteome quantitation in a large-scale level due to the lack of a universal spectral library of N-glycopeptides. Metabolic labeling such as SILAC used for quantitation has the drawback of a limited number of samples to be compared in one measurement (in general up to three) [177]. What's worse, the heterogeneous glycoforms linking to one peptide sequence with related mass are often recorded in one MS1 scan, which will be more complicated after incorporating SILAC pairs to each glycoform. It results in incorrect SILAC pair detection and XIC extraction. Isobaric chemical labeling such as using TMT or iTRAQ reagents is suitable for the quantitative proteomics in all types of sample with the benefits of sample-multiplexing, reducing overall experiment time and variations from replicates. The signal response of low abundant peptides, which may not be detected in each separate sample, will be boosted after pooling all labelled samples together [178]. Moreover, TMT labeling enables to increase the ionization efficiency of peptides or glycans [179]. Based on the low stoichiometry of glycopeptides and the above mentioned advantages in isobaric chemical labeling methods, it has already been used for glycoproteome quantitation successfully [151, 153, 180, 181]. However, a systematic optimization of experimental parameters for chemically labelled glycopeptides is still lacking due to the mutual influence of physical and chemical properties in glycopeptides with and without isobaric tags. The conventional DDA MS2-based isobaric quantitation analysis often suffers from co-isolation interference, resulting in ratio compression for the incorrect quantitation [182]. This interference is expected to be even worse in quantitative glycoproteomics analysis due to the heterogeneity of glycosylation and the adjacent masses of glycoforms. MS3 analysis enables to eliminate ratio distortion and to obtain accurate and sensitive quantitation [126, 183] and was already

applied to quantitative phosphoproteome analysis successfully [184]. In this study, we developed one glycoproteome quantitation method based on SPS-MS3 for higher accuracy and sensitivity, which is named Glyco-SPS-MS3.

1.3.1.6 Fucosylation

Fucosylation (fucose attached to glycan chain) is one important form of glycosylation related to cancer progression. Fucose is a deoxyhexose sugar, which lacks a hydroxyl group on the carbon at position-6 compared to the other hexoses. It includes L- and D-configurations, however, only L-fucose exists in mammals. Many known evidences show that fucosylated glycans are deregulated in many cancer types, resulting in the change of protein functions, which regulates metastatic cancer cells adhesion, mobility, growth and transcription [185, 186]. Donor substrate GDP-fucose, N-glycans and fucosyltransferases (FUTs) are required for the biosynthesis of fucosylated N-glycans. In mammalian, two pathways including the de novo and the salvage pathway are involved in the synthesis of GDP-fucose in the cytoplasm. In the de novo pathway, GDP-mannose is converted into GDP-fucose via enzymatic reactions catalyzed by GDP-mannose 4, 6-dehydratase (GMD) and NADP(H)-binding epimerase-reductase FX protein. In the salvage pathway, free L-fucose, which is transported into the cytosol from the extracellular space by an unknown mechanisms, is firstly phosphorylated by fucokinase (FUK). The resulting fucose-1 phosphate is converted into GDP-fucose by GDP-pyrophosphorylase. Finally, GDP-fucose from these two pathways is transported into the Golgi apparatus, where it is used for fucosylation of glycans as substrates [187]. Till now, 11 fucosyltransferases (FUT1-11) have been identified, responsible for the conjugating fucose onto N-linked oligosaccharides in four different linkages: α (1, 2), α (1, 3), α (1, 4) and α (1, 6), where the first carbon of fucose is bound to the (1, 2), (1, 3), (1, 4) or (1, 6) carbon of galactose or N-acetylglucosamine (GlcNAc) [188]. L-fucose conjugations occur at core or terminal/subterminal positions along the glycan chain. FUT8 regulates core fucosylation, which means fucose is attached to the initial N-acetylglucosamine (GlcNAc) on N-glycan in an α (1, 6) linkage [189]. FUT1 and FUT2 mediate fucosylation on terminal galactose on N-glycans in α (1, 2) linkage. The other FUTs regulate the addition of fucose to subterminal GlcNAc in α (1, 3) and α (1, 4) linkages on N-glycans.

In order to investigate the role of either core- or terminal/subterminal fucosylation in lymphoma, there are several feasible ways, such as, knockdown or knockout of the FUTs, knockout (KO) the key enzymes for GDP-fucose generation like GMD or FX, weaken the Golgi GDP-fucose transporter, inhibition of FUTs by inhibitors, use of GDP-fucose analogs. Of that, 2-fluoro-L-fucose (2FF), a specific fluorinated analog of fucose, is metabolized into the donor substrate analog of GDP-fucose, GDP-2FF instead of the normal GDP-fucose synthesis via the salvage pathway. In addition, GDP-2FF blocks the natural GDP-fucose synthesis in the *de novo* pathway. Therefore, 2FF has been used to inhibit fucosylation in many types of cancers, such as in a mouse mammary tumor cell line representing a stage 4 metastatic breast cancer [186], in colon carcinoma cells [190], in HL-60 cells [191] and in human liver cancer HepG2 cells [192]. However, the effect of inhibition of fucosylation by 2FF in lymphoma cell lines remains unclear. FUT8, as the known single fucosyltransferase to date, regulates the addition of fucose onto the core GlcNAc via an α(1,6) linkage onto N-glycans, as the core-fucosylation. It was reported that core-fucosylation showed distinct expression in cancers, for example, increased core-fucosylation in breast, liver, and lung, etc., decreased in gastric and prostate [185]. In addition, core-fucosylation was found to be involved in several important growth factors like TGFβ, EGFR, VEGFR and c-Met signaling pathways. Knockdown or inhibition of FUT8 enhances these signaling pathways, resulting in inhibiting cell growth or survival in lung cancer [193] and hepatocellular carcinoma [194].

Knockdown of FUT8 to decrease core fucosylation hinders cancer growth and survival. Till now, there is still no publication about the role of fucosylation in B cell lymphoma. More detailed investigations are demanding to reveal the underlying mechanisms about how the altering fucosylated glycopeptide level mediates the survival of lymphomas. The final goal is to optimize treatment strategies and develop new ones for the improvement of clinical outcomes of patients based on the expanding understanding of the role of glycosylation in lymphomas.

1.3.2 Phosphorylation analysis by MS

Similar to glycopeptides, phosphopeptides are also very low abundant on the peptide level in organisms. Efficient enrichment workflows for phosphopeptides from complex protein digests are required prior to large-scale phosphopeptide identification using mass spectrometry. In the past decades, several mature enrichment methods have been developed, such as, metal oxide affinity chromatography (MOAC), immobilized metal ion affinity chromatography (IMAC), and immunoprecipitation (IP)-based enrichment. MOAC is based on the affinity between phosphate groups and metal oxides, like TiO_2, ZrO_2, etc. TiO_2 is the most commonly used metal oxide for phosphopeptide enrichment with high selectivity and specificity. Briefly, in acidic condition, the protonation of acidic residues of non-phosphorylated peptides hinders their affinity to TiO_2 resulting in their separation from phosphopeptides which can be eluted from TiO_2 beads using alkaline buffers [195]. The affinity mechanism between phosphopeptides and IMAC resins is based on the electrostatic interactions between the negatively charged phosphate groups of phosphopeptides and the positively charged metal ions which are bound to a solid support. The procedures in IMAC and MOAC are similar. Three main steps of loading samples to IMAC column, the removal of non-phosphopeptides by binding buffer, elution of phosphopeptides from beads with suitable elution buffers are included. The immunoprecipitation-based method is specific for the enrichment of tyrosine phosphorylated peptides (pY) using phospho-Tyrosine antibodies due to the much lower abundance of pY peptides compared to phospho-serine (pS) and phospho-threonine (pT) peptides. Due to the non-existence of highly specific antibodies for pS and pT, IP-based approaches are not suitable for serine- and threonine-phosphorylated peptide enrichment. Furthermore, the fractionation of enriched phosphopeptides is performed prior to MS analysis in order to reduce sample complexity to improve phosphopeptide identification. Ion-exchange chromatography such as strong cation exchange (SCX) or strong anion exchange (SAX), basic reverse phase chromatography, hydrophilic interaction liquid chromatography and electrostatic repulsion hydrophilic interaction chromatography (ERLIC) are the common approaches for the pre-fractionation of enriched phosphopeptides [196].

The mass shift induced by the addition of a phosphate group to the amino acid on a peptide sequence is +79.9663 Da (HPO_3) [197]. A neutral loss of phosphoric acid (HPO_3) with 79.9663 Da is prone to happen for pY peptides under CID fragmentation, while, loss of meta-phosphoric acid (H_3PO_4) with 97.976 Da is more often in phosphorylated serine and threonine residues. This is because of the stronger C-O bond in phosphorylated tyrosine residues than in phosphorylated serine and threonine residues [198]. The additional MS3 scan and multistage activation (MSA) are performed the ion trap CID fragmentation along with the longer cycle time, which enables to improve the phosphopeptide identification with the neutral loss on MS2 scan [199]. In HCD fragmentation, the peptide-specific fragment ions (b- and y-ions) without the neutral loss tends to occur. In addition, some immonium ions with small m/z are prone to produce, such as the phosphotyrosine-specific immonium ion at m/z 216.0426, which help to identify specific modified residues due to the overcoming of low mass cutoff under HCD fragmentation [200].

1.3.3 Ubiquitylation analysis by MS

With the advance of high-resolution mass spectrometry (MS) and efficient ubiquitinated peptides enrichment methods, large-scale identification of ubiquitylation sites in many cells and tissues samples was achieved using MS-based proteomic strategies [201-204]. Briefly, after the proteolytic digestion with trypsin, ubiquitin is cleaved between the arginine at position 74 and glycine at position 75, leaving Gly75 and Gly76 residues on the modified lysine of the substrate peptide. Di-glycine modified peptides can be recognized and targeted by a specific antibody due to the K-GG motif. The mass shift of 114.0429 Da on the lysine residue allows the identification of ubiquitin modified sites from the fragment ions by MS. Unfortunately, the poly-ubiquitination information is lost by this method. Moreover, two ubiquitin-like proteins (UBLs, NEDD8 and ISG15) were found to own the same di-glycine modified lysines peptides after trypsin digestion, which is indistinguishable by antibody and MS. Later on, a UbiSite method based on a new antibody to enrich the lysine-c digested ubiquitinated peptides was developed which enables to eliminate the interference of UBLs [201]. However, since ISG15 and NED88 are less abundant with respect to their expression level [205], their effect on large-scale ubiquitination sites analysis is very limited. The mimic of ubiquitination by iodoacetamide-induced artifacts were observed in MS analysis. The use of chloroacetamide instead of iodoacetamide as an alkylating reagent can avoid the production of di-glycine-like artifacts [206]. HCD fragmentation is widely used and the most efficient dissociation method for the identification of di-glycine modified peptides by MS. Due to the missed cleavage at the modified lysine by ubiquitin after trypsin digestion, the majority of di-glycine modified ubiquitinated peptides own higher charges of at least +3. ETD fragmentation was demonstrated to be an available choice for the identification of ubiquitinated peptides [207].

2. Objective

The classification of cancer subtypes is a very important basis for rational treatment of any heterogeneous tumors. DLBCL, as one of the highly aggressive B cell lymphoma, was classified into two main subtypes based on gene expression profiling. They represent the different pathogenesis and clinical responses to treatment. Previous studies uncovered that the protein expression and glycosite level (deglycopeptides expression) from the cell line model of DLBCL enabled to segregate these two subtypes using clinical proteomics analysis. It helps to reveal cancer related characteristics and the developmental stage of B cell. In many DLBCL cases, $IgV_{H4\text{-}34}$ BCR binding sugars from cell surface initiates BCR signaling in lymphoma. In order to obtain the glycan information in DLBCL, we decided to investigate the site-specific glycosylation profiling (intact glycopeptide expression) in the cell lines of DLBCL subtypes (ABC DLBCL and GCB DLBCL) using quantitative site-specific N-glycoproteomics analysis. However, accurate quantification of intact glycopeptides on a large scale remains technically challenging, and software tools for quantitative data processing are still lacking. Therefore, we firstly established a streamlined pipeline for multiplexed quantitative site-specific N-glycoproteomics. The accuracy of quantitation for mapping protein glycosylation patterns using our developed platform was evaluated in Burkitt's lymphoma cells treated with varying doses of 2FF. We also demonstrated the capability of this platform for site-specific characterization of reduced expression of fucosylated N-glycans in 2FF-treated and FUT8 KO TMD8 cell line.

The survival of DLBCL is thought to be driven by antigen-mediated BCR signaling, which is the PTM-based signaling pathway. Phosphorylation and ubiquitylation were found to play important roles in BCR signaling. However, the phosphorylation and ubiquitylation modified proteins involved in BCR signaling and the extent of their involvement are still not fully known. Moreover, some tyrosine kinases have been identified as the essential effectors in the upstream of BCR signaling, such as, SYK and BTK. Upon the inhibition of BTK and SYK in DLBCL, it would be very interesting to detect the change of downstream signaling outputs. Thus, we performed systems-wide analyses of signaling networks by monitoring the dynamics of phosphorylation and ubiquitylation in BTK and SYK inhibited DLBCL cell lines using quantitative MS-based proteomics technology.

3. Materials and Methods

3.1 Materials

Commonly employed chemicals like salts were obtained from either Sigma-Aldrich (St. Louis, USA), Carl-Roth (Karlsruhe, Germany), Merk Millipore (Billerica, USA), Fluka (Buchs, Germany), or Thermo Fisher Scientific (Waltham, USA).

3.1.1 Chemicals and reagents

Item name	Manufacturer
2-Deoxy-2-fluoro-L-fucose (2FF)	Carbosynth, UK
Acetic acid, ≥99.5%, FG	Merck, Germany
Acetone	Merck, Germany
Acetonitrile (ACN), LichroSolv	Merck, Germany
Ammonium acetate (NH_4Ac)	Sigma-Aldrich, Germany
Ammonium bicarbonate (ABC)	Merck, Germany
Ammonium hydroxide (NH_4OH) 28-30% (w/v)	Acros Organics, Belgium
β-Glycerophosphate	Sigma-Aldrich, Germany
Biotinylated Aleuria Aurantia Lectin (AAL)	Vector laboratories, USA
Boric acid	Sigma-Aldrich, Germany
Bovine serum albumin (BSA)	Sigma-Aldrich, Germany
Complete Protease Inhibitor Cocktail, EDTA-free (Tablets)	Merck, Germany
Chloroacetamide (CAA)	Sigma-Aldrich, Germany
Coomasie Brilliant Blue G-250	Sigma-Aldrich, Germany
Dithiothreitol (DTT)	Roth, Germany
Dimethyl Sulfoxide (DMSO)	Sigma-Aldrich, Germany
EDTA Solution (0.5 M, 100X)	Thermo Scientific, USA
Ethanol, LiChroSolv	Merck, Germany
Formic acid (FA)	Fluka, Germany
Glycerol, >99.5%, p.a.	Merck, Germany
Glycine	Merck, Germany
HEPES	Merck, Germany
Halt™ Protease and Phosphatase Inhibitor Cocktail (100X)	Thermo Scientific, Germany
HILIC amphion	Welch, China
Hydrochloric acid (HCl, 37%)	Merck, Germany
IAP buffer (10x)	CST, USA
IgM from human serum	Sigma-Aldrich, Germany
Iodoacetamide (IAA)	Sigma-Aldrich, Germany
MagReSyn HILIC (cat.no. MR-HLC005)	ReSynBio, South Africa
MS Compatible Yeast Protein Extract, Intact (V7341)	Promega, Germany
Methanol, LiChroSolv	Merck, Germany
NP-40 (10% in H_2O)	Biovision, USA
NuPAGE LDS sample buffer (4X)	Invitrogen, USA
NuPAGE MOPS SDS running buffer (10X)	Invitrogen, USA
NuPAGE Sample reducing agent (10X)	Invitrogen, USA
Phosphate buffered saline (PBS, tablet)	Sigma-Aldrich, Germany

Phosphoric acid, 85 wt. % in H_2O, FCC, FG	Sigma-Aldrich, Germany
Quenching buffer (50% hydroxylamine)	Thermo Scientific, USA
RapiGest	Waters, USA
ReproSil-Pur 120 C18-AQ, 120 Å, 1.9 μm pore size	Dr. Maisch, Germany
Sera-Mag SpeedBeads (cat.no. 45152101010250)	GE Healthcare, Germany
Sera-Mag SpeedBeads (cat.no. 65152105050250)	GE Healthcare, Germany
Sodium chloride (NaCl)	Merck, Germany
Sodium dodecyl sulphate (SDS)	SERVA, Germany
Sodium fluoride (NaF, 0.5 M)	Sigma-Aldrich, Germany
Sodium hydroxide (NaOH)	Merck, Germany
Sodium orthovanadate (Na_3VO_4)	Sigma-Aldrich, Germany
Sodium pyrophosphate	Sigma-Aldrich, Germany
Strep-Tactin®-HRP conjugate	IBA, Germany
Triethylammonium bicarbonate buffer (TEAB, 1.0 M)	Sigma-Aldrich, Germany
Titanium dioxide (TiO_2, 10 μm)	GL Sciences, Japan
Trifluoroacetic acid (TFA)	Roth, Germany
Tween-20	Roth, Germany
Bond-Breaker ™ TCEP Solution (0.5 M)	Thermo Scientific, Germany
Tris(hydroxymethyl)aminomethane (Tris)	VWR, Germamy
Urea	Sigma-Aldrich, Germany
Water, LiChroSolv	Merck, Germany

3.1.2 Buffers and solutions

Purpose	Buffer/Solution	Composition
General MS sample preparation	Reduction buffer	1.25 M DTT stock in water, 0.5 M TCEP
	Alkylation buffer	100 mM CAA/IAA in 50 mM ABC/TEAB (freshly prepared, in the dark)
	Incubation buffer	50 mM ABC/TEAB in water
	Sample loading buffer for MS	A: 0.1 % FA; B: 50% ACN, 0.1% FA
Lysis buffer	NP40 lysis buffer	50 mM Tris-HCl pH 7.5-7.8, 150 mM NaCl, 0.5% NP40, 5 mM NaF, 1 mM Na3VO4 , 1XComplete Protease Inhibitor Cocktail, EDTA-free (Roche)
	Urea lysis buffer	20 mM HEPES/NaOH pH 8.0, 9 M urea, 1 mM sodium orthovanadate, 2.5 mM sodium pyrophosphate, 1 mM β-glycerophosphate (freshly prepared)
	SDS lysis buffer	4% (w/v) SDS in 50 mM TEAB
	UA lysis buffer	8 M urea in 50 mM ABC
	Rapigest lysis buffer	1% or 0.1% RapiGest in 50 mM TEAB
Phosphopeptides enrichment (TiO_2)	Equilibration/binding buffer	5% glycerol in 80% ACN, 5% TFA
	Washing buffer 1	80% ACN, 5% TFA
	Washing buffer 2	60% ACN, 0.1% TFA
	Elution buffer	1~5% NH_4OH (pH ≥10.5)

Immunoaffinity Purification (IAP)	1X PBS	One tablet dissolved in 200 ml of deionized water yields 0.01 M phosphate buffer, 0.0027 M KCl and 0.137 M NaCl, pH 7.4
	1X IAP	50 mM MOPS/NaOH pH 7.2, 10 mM Na_2HPO_4, 50 mM NaCl
	Washing buffer	chilled HPLC water and 1X IAP buffer
	Elution buffer	0.2% TFA
Glycopeptides enrichment	Loading buffer	80% ACN, 1% TFA
	Elution buffer	0.1% TFA
Basic C18 Fractionation	Mobile phase A	10 mM NH_4OH, pH 10
	Mobile phase B	10 mM NH_4OH in 80% ACN, pH 10
Desalting	Solvent A	0.1% TFA
	Solvent B	40% ACN, 0.1% TFA
	Solvent C	50% ACN, 0.1% TFA
Acidification	Solvent	TFA, 20% TFA, 10% TFA
SCX Buffer	Washing buffer	50% ACN, 0.1% TFA
	Elution buffer 1	20 mM acetic acid, 20 mM boric acid, 20 mM phosphoric acid in 40% ACN, pH 4.5
	Elution buffer 2	20 mM acetic acid, 20 mM boric acid, 20 mM phosphoric acid in 40% ACN, pH 5.5
	Elution buffer 3	20 mM acetic acid, 20 mM boric acid, 20 mM phosphoric acid in 40% ACN, pH 8.5
		* pH adjustment by 5 M NaOH for elution buffer
SDS-PAGE	Coomassie staining solution	0.08% (w/v) Coomassie Brilliant Blue, 1.6% (v/v) ortho-phosphoric acid, 8% (w/v) ammonium sulphate, 20% (v/v) Methanol
Western/lectin blot	TBS	0.1 M Tris-HCl pH 7.5, 150 mM NaCl
	TBST	0.1% Tween-20 in TBS (freshly prepared)
	Blocking buffer	5% (w/v) BSA in TBST
	Transfer buffer	190 mM Glycine, 25 mM Tris, 0.05% SDS, 20% Methanol, pH 8.5
Liquid chromatography (LC)	Buffer A and B	0.1% FA (A); 80% ACN, 0.08% FA (B)

3.1.3 Commercial kits

Item name	**Manufacturer**
Pierce BCA Protein Assay Kit	Thermo Scientific, Germany
PTMScan Phospho-Tyrosine Rabbit mAb (P-Tyr-1000)	CST, Germany
PTMScan Ubiquitin Remnant Motif (K-ε-GG)	CST, Germany
Quick Start™ Bradford Protein Assay	Bio-Rad, Germany
TMT Mass Tag Labeling Kits (TMT^0, 6plex,10plex)	Thermo Scientific, Germany

3.1.4 Consumables

Item name	Manufacturer
CL-Xposure film (18 × 24 cm)	Thermo Scientific, Belgium
Diamond Tower Pack pipet tips (10 µL, 20-200 µL, 1000 µL)	Gilson, France
Falcons (15 ml, 50 ml)	Sarstedt, Germany
GELoader tip	Eppendorf, Germany
Empore C18/Cation 47 mm Extraction Disks	3M, USA
Empty Micro SpinColumns, 5 µm frit	Harvard Apparatus, USA
Micro SpinColumns, C18, 5 µm frit	Harvard Apparatus, USA
Nitrocellulose Western blotting membranes	GE Healthcare, Germany
NuPAGE Novex 4-12% Bis-Tris Mini Gels	Invitrogen, USA
Parafilm	Bemis, USA
Safe-Lock Tubes (0.5 ml, 1.5 ml, 2 ml, 5 ml)	Eppendorf, Germany
Oasis HLB Vac Cartridge (1 cc, 3 cc, 6 cc)	Waters, Germany
Syringe (50 µL)	Hamilton, USA
Whatman paper	GE Healthcare, Germany

3.1.5 Equipment

Category	Item name	Manufacturer
Centrifuge	Biofuge Pico	Heraeus, Germany
	Centrifuge 5424	Eppendorf, Germany
	Heraeus Fresco 17	Thermo Scientific, Germany
	Heraeus Pico 17	Thermo Scientific, Germany
	Multifuge X3R	Thermo Scientific, Germany
	Sorvall Lync 6000 (rotors F14-14X50 cy)	Thermo Scientific, Germany
	Qualitron DW-41	Artisan, USA
Computation (MaxQuant server)	PowerEdge R815	Dell, Germany
Electrophoresis	Electrophoresis power supplies	Biorad, Germany
	XCell SureLock Mini-Cell Electrophoresis System	Thermo Scientific, Germany
	Epson perfection V700 photo scanner for gel	Epsom, Germany
LC system	Agilent 1100 series	Agilent, Germany
	EASY n-LC 1000	Thermo Scientific, Germany
	Ultimate 3000 RSLC UHPLC	Thermo Scientific, Germany
Lyophilization	Lyophilization apparatus alpha 2-4	Christ, Germany
	SpeedVac (Concentrator 5301)	Eppendorf, Germany
pH meter	MPC227 pH/Conductivity Meter	Mettler Toledo, Germany
	Orion 2-Star pH Benchtop Meter	Thermo Scientific, Germany
Rotator	Rotator SU 1500	Sustainable lab, Germany
	Test-tube-rotator	Schuett labortechnik, Germany
Sonicator	Bioruptor sonicator	Diagenode, Belgium
	Microtip sonication Sonifier W-250D	Branson Digital Sonifier, Germany
	Sonication bath SONOREX Super	BANDELIN electronic, Germany

Other	HERAfreeze basic (-80 °C)	Thermo Scientific, Germany
	Ice machine	Ziegra Eismaschinen, Germany
	Nanodrop ND-1000 UV/Vis spectrophotometer	PqQLab, Germany
	Thermomixer Comfort	Eppendorf, DE
	Hybaid Shake 'n' Stack (37 °C incubator)	Thermo Scientific, Germany
	Amersham Hyperprocessor	Amersham, UK

3.1.6 HPLC column

Item name	Manufacturer
Xbridge C18 1.0 × 150 mm, 3.5 μm	Waters, Ireland
In-house-packed C18 column (ReproSil-Pur 120 C18-AQ, 1.9 μm, 300 × 0.075 mm)	Dr. Maisch, Germany

3.1.7 Enzymes

Item name	Manufacturer
Glu-C, sequencing grade	Promega, USA
Sequencing Grade Modified Trypsin	Promega, USA
PNGase F (Glycerol-free)	New England Biolabs,USA

3.1.8 Cell lines

Type	Name	Species	WT	Cas9-positive derivatives
Burkitt's lymphoma	DG75	*Homo sapiens*	+	
	Daudi	*Homo sapiens*	+	
ABC	HBL1	*Homo sapiens*	+	+
	TMD8	*Homo sapiens*	+	+
	OCI-Ly3	*Homo sapiens*		+
	OCI-Ly10	*Homo sapiens*	+	+
	RIVA	*Homo sapiens*		+
GCB	U2932	*Homo sapiens*	+	
	DOHH2	*Homo sapiens*	+	+
	SUDHL4	*Homo sapiens*	+	+
	SUDHL5	*Homo sapiens*		+
	WSU-FSCCL	*Homo sapiens*	+	+

1. WT of DLBCL cell lines were used for ubiquitinome;
2. Cas9-positive derivatives of DLBCL were used for phosphoproteome and glycoproteome;
3. Burkitt's lymphoma is only used for glycoproteome analysis in this study.

3.1.9 Mass spectrometers

Item name	Manufacturer
Q Exactive HF Hybrid Quadrupole Orbitrap	Thermo Scientific, Germany
Orbitrap Fusion Tribrid Mass Spectrometer	Thermo Scientific, Germany
Orbitrap Fusion Lumos Tribrid Mass Spectrometer	Thermo Scientific, Germany

3.1.10 Software

Name	Company/Institution
Adobe Illustrator CS6	Adobe, USA
Byonic	Protein Metrics, USA
Cytoscape	Cytoscape Consortium
GlycoBinder (self-developed)	Max Planck Institute for Biophysical Chemistry, Germany
KEGG	Kyoto University, Japan
MaxQuant	Max Planck Institute for Biochemistry, Germany
Office 2016	Microsoft, USA
Origin 2019	OriginLab, USA
Perseus	Max Planck Institute for Biochemistry, Germany
pGlyco	Institute of Computing Technology, Chinese Academy of Sciences, China
Proteome Discoverer	Thermo Scientific, Germany
R	R Core Team
R studio	RStudio, Inc.
STRING	Academic Consortium
Xcalibur	Thermo Scientific, Germany

3.2 Methods

Cell lines which were used for ubiquitinome and glycoproteome analysis in this study were cultured by Silvia Münch and Dr. Carmen Döbele in Frankfurt. Cell lines which were used for pYome and global phosphoproteome analysis were cultured by Dr. Thomas Oellerich at the NCI. Sample preparation for LC-MS/MS was performed in Goettingen. Western blotting and lectin blotting for the DLBCL glycosylation experiments were done by Silvia Münch and Dr. Carmen Döbele in Frankfurt. GlycoBinder was developed by Ivan Silbern, a PhD student in Prof. Dr. Henning Urlaub lab in MPI-BPC. IgM-yeast interference model was done by Dr. Pan Fang in Prof. Dr. Henning Urlaub lab in MPI-BPC.

3.2.1 Cell-based sample preparation

3.2.1.1 Cell culture

For regular cultivation of cells, we used RPMI 1640 (Cat# 21875, ThermoFisher Scientific) supplemented with 10% heat-inactivated (h.i.) FCS (Sigma-Aldrich), 2 mM L-Glutamine (Thermo Fisher Scientific) and 100 U/ml Penicillin and 100 µg/ml Streptomycin (Thermo Fisher Scientific) or advanced RPMI 1640 (Cat# 12633, Thermo Fisher Scientific) supplemented with 5% h.i. FCS (Sigma-Aldrich), 2 mM L-Glutamine (Thermo Fisher Scientific) and 100 U/ml Penicillin and 100 µg/ml Streptomycin (Thermo Fisher Scientific). Daudi (DSMZ no.: ACC 78) were purchased from the Leibniz Institute DSMZ-German Collection of Microorganisms and Cell Cultures GmbH in Braunschweig, Germany. DG75 were provided by Prof. Dr. Thorsten Zenz at the National Centre for Tumor Diseases (NCT) in Heidelberg, Germany. Both cell lines were authenticated using Multiplex Cell Authentication by Multiplexion (Heidelberg, Germany) as described recently (Castro F, Dirks WG, Fähnrich S, Hotz-Wagenblatt A, Pawlita M, Schmitt M. 2012. High-throughput SNP-based authentication of human cell lines. Int J Cancer. 2012 Jun 15. doi: 10.1002/ijc.27675.). The SNP profiles matched known Daudi and DG75 profiles. Cell lines were regularly tested for mycoplasma according to a PCR protocol published

by Uphoff and Drexler (Uphoff CC, Drexler HG. Detection of mycoplasma contaminations. Methods Mol Biol 2005; 290:13-23).

3.2.1.2 SILAC labeling in DLBCL

For the ubiquitinome, pYome and global phosphoproteome analysis, we cultured the DLBCL cell lines in SILAC RPMI (SILANTES Cat# 283001300) containing 2 mM L-Glutamine supplemented with 10% h.i. dialyzed FCS (Bio&Sell), 100 U/ml Penicillin and 100 µg/ml Streptomycin (Thermo Fisher Scientific), 200 mg/l L-proline (Roth) and either the natural (light) isotopes of arginine (250 mM) (Sigma-Aldrich) and lysine (600 mM) (Sigma-Aldrich) or heavy isotopes of (L-ARGININE:HCL (^{13}C6, 99%; ^{15}N4, 99%), Cat# CNLM-539-H-1; L-LYSINE:2HCL (^{13}C6, 99%; ^{15}N2, 99%), Cat# CNLM-291-H-1; both Cambridge Isotope Laboratories; heavy) for at least 6 cell doublings.

3.2.1.3 BTK and SYK inhibition in DLBCL

For BTK and SYK inhibitor treatment of DLBCL cells, we seeded 1,2x10^8 cells in 50 ml fresh SILAC medium and added either 25 µl of 0.5 mM PRT062607 (250 nM final) or 25 µl 20 µM Ibrutinib (10 nM final, both inhibitors dissolved in DMSO) to the heavy labelled cells and the same amount of DMSO (vehicle control, 1:2000, 0.05% final) to light labelled cells. After 3 h, cells were washed once with cold PBS and lysed in 10 ml (later 7 ml) urea lysis buffer, shock froze them in liquid nitrogen and stored them at -80°C for pYome and ubiquitinome analysis. Cells lysed in NP40 lysis buffer stored at -80°C for GPome analysis.

3.2.1.4 Fucosylation inhibition in DL75 and DLBCL

For the DG75 glycosylation experiment, we seeded 1.5x10^6 DG75 cells in 15 ml 10% RPMI and treated them with 15 µL 600 mM (600 µM final), 480 mM (480 µM final), 240 mM (240 µM final), 120 mM (120 µM final) and 60 mM (60 µM final) 2FF in DMSO or DMSO only (vehicle control, 1:1000, 0,1%) for 3 days. We harvested 1.5x10^7 cells per condition, washed them once with PBS, shock froze them in liquid nitrogen and stored them at -80°C. For the DLBCL glycosylation experiment, cells were harvested at day 1, day 3 and day 5 separately. Day1 samples: 3,75x10^6 cells were treated with 600, 480, 240, 120, 60, 30, 15 and 7.5 µM 2FF in 5 ml Advanced RPMI separately. After one day (approx. 22 hours), cells from 5 ml cell suspension were harvested, washed once in PBS and lysed in Triton X-100 Lysis buffer (20 mM Tris-HCl (pH 7.5), 1% Triton X-100, 150 mM NaCl, 5 mM NaF, 1 mM Na_3VO_4, Complete Mini Protease Inhibitors) using 10 µL lysis buffer per 10^6 cells.
Day 3 and Day 5 samples: 4x10^6 cells were treated with 600, 480, 240, 120, 60, 30, 15 and 7.5 µM 2FF in 10 ml Advanced RPMI separately (5 ml per well in a 6-well plate). At day 3, cells from 5 mL cell suspension were harvested, washed once in PBS and lysed in Triton X-100 lysis buffer using 10 µL lysis buffer per 10^6 cells. Splitting of remaining cells for day 5 was achieved by addition of medium containing the respective amount of 2FF to reduce cell density to approx. 5×10^5/ml, then cell harvest at day 5 was done as described above.

3.2.1.5 FUT8 knockout

For the DLBCL glycosylation experiment, TMD8-Cas9 pLKO-Puro-GFP-sgNT, -sgFUT8-1, sgFUT8-2 and sgFUT8-3 were incubated for 3 days in Advanced RPMI containing 500 ng/ml Doxycyclin to induce Cas9-Expression. All of the three distinct small guide RNA molecules contain the homologous sequence with the targeted FUT8 DNA sequence. Once the DNA sequence matches the sgRNA, Cas9 nuclease will come to break the strands. Inserted or deleted bases will occur after the imprecise repairing by the non-homologous end joining repair (NHEJ) pathway, resulting in gene knockout.

3.2.2 Biochemical methods

3.2.2.1 Protein and peptide concentration measurement

Two protein concentration measurement kits were used in this study including Quick Start™ Bradford Protein Assay (BIO-RAD) and Pierce™ BCA Protein Assay Kit (Thermo Scientific). It was performed according to the provided individual protocol.

3.2.2.2 SDS-PAGE

SDS-PAGE which was performed in Goettingen used the NuPAGE™ 4-12% Bis-Tris Protein Gels, 1.0 mm, 10-well, provided by Invitrogen™ 4–15% Mini-PROTEAN® TGX™ Precast Protein Gels, 10-well, 30 µl (#4561083, BIO-RAD) was used for SDS-PAGE in Frankfurt. Either was done based on the manufacturer's instruction.

3.2.2.3 Western blotting

In the FUT8 knockout experiment, TMD8-Cas9 pLKO-Puro-GFP-sgNT (*TAAAGCAGAAGAATATACAG*), pLKO-Puro-GFP-sgFUT8-1 (*CTAAACTACAGGATACCAGA*), pLKO-Puro-GFP-sgFUT8-2 (*AATTGGCGCTATGCTACTGG*), and pLKO-Puro-GFP-sgFUT8-3 (*ACAGCCAAGGGTAAATATGG*) cells were harvested at day 3 after Doxycycline induction followed by washing with PBS and lysed in 2x Lämmli buffer. 30 µl of the crude lysate were separated by SDS-PAGE using 4–15% Mini-PROTEAN® TGX™ Precast Protein Gels, 10-well, 30 µL and transferred to a PVDF membrane. Blocking of membranes was achieved using 5% BSA in TBS-T (4°C, O/N), followed by incubation of the membranes with 3 µg/ml FUT8 antibody from Abcam (ab198741) in 5% BSA in TBS-T for 1 h at RT. Membranes were then washed three times with TBS-T followed by incubation with HRP-conjugated streptavidin (1:2000) in 5% BSA in TBS-T for 1h at RT. After three TBS-T washes, signals were detected using SuperSignal™ West Femto Maximum Sensitivity Substrate (Thermo Fisher) and a Licor Odyssee device. Actin served as loading control.

3.2.2.4 Lectin blotting

For the DG75 glycosylation experiment, DG75 cell pellets with and without 2FF treatment were lysed in 4% SDS, 0.1 M Tris-HCL, pH 8.0. Protein concentrations were determined by BCA. For each condition, 20 µg proteins were reconstituted in 1 × NuPAGE LDS Sample Buffer (Invitrogen) and separated on 4%–12% NuPAGE Novex Bis-Tris Minigels (Invitrogen). The proteins separated in gel were transferred onto a nitrocellulose membrane (GE Healthcare). The membrane was blocked with 5% (w/v) bovine serum albumin (Sigma) in TBST overnight at 4 °C. The membrane was then incubated with biotinylated lectins AAL (vector laboratories) (3 µg/ml in blocking buffer) for 1 h at room temperature. The membrane was washed three times using TBST. A final incubation of the membrane with Strep-Tactin®-HRP conjugate (iba) in TBST was done for 1 h. After washing, the signal was visualized using a chemiluminescence detection system (ECL, GE Healthcare) and detected on X-ray film.

For the DLBCL glycosylation experiment, at day1, day3 and day5 obtained TMD8 cell samples were lysed in 2x Lämmli buffer. Protein concentration was determined using the BCA protein assay kit. 50 µg protein per sample were separated by SDS-PAGE. Proteins were blotted onto nitrocellulose membrane. The membranes were probed with 3 µg/ml biotinylated aleuria aurantia lectin. Actin served as loading control. The detailed procedures refer to section 3.2.2.3.

3.2.3 Sample preparation for LC-MS/MS

3.2.3.1 MS Sample preparation for ubiquitinome analysis

Cell lysates in 10 ml (later 7 ml) urea lysis buffer were sonicated to break DNA/RNA using a Sonifier W-250D (Branson Digital Sonifier) with a microtip. It was performed on ice with the setting of 15s on / 59.9s off per cycle at 40% output, 3-4 cycles in total. Lysates were cleared by centrifugation at 14,000 g for 15 min at RT and the supernatants were transferred into a new tube. The protein concentration of each SILAC state (light and heavy labelled ones) was measured by Bradford assay. Equal protein amounts (7.5 mg or 10 mg) of light labelled and heavy labelled samples were combined into one tube. Next, mixed protein lysates were reduced with 4.5 mM DTT at 37 °C for 1 h, then alkylated with 10 mM CAA for 40 min at RT in the dark. The reduced and alkylated protein lysates were diluted by addition of 50 mM ABC to a final concentration of urea to less than 1 M. Sequence grade modified trypsin (Promega) was added at an enzyme to protein ratio of 1:50 for digestion at 37°C overnight. The digested sample was acidified by adding 20 % TFA to a final concentration of 1 % TFA followed by placing sample stand on ice for 15 min to form pellets. Centrifuge the acidified peptide solution for 15 min at 2,000 g at RT to remove the precipitate, then the supernatants were transferred to a new fresh tube without taking out the pellets. Peptides purification was performed using Waters Oasis HLB column (6 cc, 200 mg sorbent per cartridge) immediately. Application of all solutions during desalting steps, should be performed by gravity flow. Firstly, the column was activated with 6 ml 100% ACN, followed by washing 3 times using 6 ml 0.1% TFA. The acidified sample was loaded onto the column twice. Then the column was washed with 0.1% TFA three times. Purified peptides were eluted with a sequential wash of 3 × 2 ml solvent B (40% ACN, 0.1% TFA). 50 µg purified digested peptides were collected in an Eppendorf tube followed by drying them in a SpeedVac. Samples were stored at -80°C until proteome analysis. The remaining eluents were collected in a pear-shaped flask followed by freezing it in cold ethanol with dry ice. The frozen peptide solutions were then lyophilized and stored at -80°C until use. Lyophilized peptides were then used for ubiquitinated peptides enrichment using the PTMScan Ubiquitin Remnant Motif (K-ε-GG) kit (refers to section 3.2.3.9). In order to improve the identification of ubiquitination, enriched ubiquitinated peptides were separated into three fractions by SCX chromatography (refers to section 3.2.3.12). Briefly the three eluents obtained from SCX were evaporated to remove ACN followed by a C18 stage-tip desalting. C18 desalting was performed as follows: 4~6 pieces of Empore C18 Extraction Disks were packed into a 200 µL yellow pipette tip to form a C18 microtip column. C18 material was equilibrated by 100 µL solvent C (50% ACN, 0.1% TFA) once, 100 µL solvent A (0.1% TFA) twice sequentially, followed by sample loading twice. The C18 microtip column containing bound peptides was washed twice with 100 µL solvent A (0.1% TFA) and enriched peptides were eluted with 2 × 50 ml solvent B (40% ACN, 0.1% TFA). The eluate was dried in a SpeedVac and stored at -80°C until MS analysis.

3.2.3.2 MS Sample preparation for pYome analysis

Cultured Cell lines for pYome analysis were also lysed in urea lysis buffer. All the sample preparation steps are the same as ubiquitinome analysis except that no SCX prefractionation was used for enriched tyrosine-phosphorylated peptides (refer to section 3.2.3.1).

3.2.3.3 MS Sample preparation for global phosphoproteome analysis

Cells for GPome analysis were lysed in NP40 lysis buffer. The lysates were sonicated followed by centrifugation to obtain a clear supernatant for protein concentration measurement by BCA assay. Equal protein amounts of each SILAC state were mixed. Except for the WSU-FSCCL cell lines which were

triple-labelled, all the other cell lines were double labelled. In general, 5 mg mixed protein lysates were used for digestion and phosphopeptide enrichment. All the proteins were reduced and alkylated with 10 mM TCEP and 20 mM IAM at 37 °C for 60 min in the dark. Then the three fold sample volume of cold (-20 °C) acetone was added and the samples were stored at -20 °C for at least 3 hours or overnight to precipitate the proteins. The samples were then centrifuged for 10 min at 4,500 rpm. The supernatants were removed properly without dislodging the protein pellets followed by washing the pellets with 70% cold (-20 °C) acetone three times. The supernatants were decanted and pellets were air dried for 10 min at RT in the fume hood. The protein pellets were dissolved in UA lysis buffer (8 M urea in 50 mM ABC) by repeated pipetting or sonicating. The dissolved protein solutions were diluted using 50 mM ABC to reduce the final concentration of urea to less than 1 M followed by trypsin digestion at an enzyme to protein ratio of 1:50 at 37°C overnight. The digested sample were acidified by addition of TFA and peptides purification was performed using Waters Oasis HLB column (1cc, 6 cc) immediately (refers to section 3.2.3.1). 50 µg purified digested peptides were stored for proteome analysis, the left peptides were used for phosphopeptide enrichment using TiO_2 (refers to 3.2.3.7). For enriched phosphopeptides basic reverse phase prefractionation was performed to improve the identification (refers to 3.2.3.11). All the fractions were dried in a SpeedVac and stored at -80°C until MS analysis.

3.2.3.4 MS Sample preparation for quantitative N-glycoproteomics

In the SugarQuant platform (TMT-based quantitative site-specific N-glycoproteomics) establishment part, DG75, Daudi cell lysates and IgM digests were used to evaluate and optimize the experimental details. Preparation of peptides from cells using conventional methods was performed as follows. DG75 and Daudi cell pellets were lysed with either one of the four buffers: Buffer 1 consisted of 4% SDS (w/v), 50 mM HEPES, pH 8.0; Buffer 2 consisted of 8 M urea in 50 mM TEAB; Buffer 3 consisted of 0.1 % (w/v) RapiGest SF surfactant (Waters) in 50 mM TEAB; Buffer 4 consisted of 1 % (w/v) RapiGest in 50 mM TEAB. All the above four buffers were supplemented with 1 × Halt Protease and Phosphatase Inhibitor Cocktail and 1 × EDTA (Thermo Scientific) s. All the cell lysates were sonicated for 15 - 30 min (30s on, 30s off) at 4 °C using a Bioruptor. After centrifugation at 14,000 × g for 15 min, the supernatants were collected and the protein concentrations were determined using the BCA Kit (Thermo Scientific). All the proteins were reduced and alkylated with 10 mM TCEP and 20 mM IAM at 37 °C for 1 h in the dark. For the samples in buffer 1, three-times sample volumes of cold (-20 °C) acetone were added followed by storage at -20 °C for at least 3 hours or overnight to precipitate the proteins. The samples were then centrifuged at 14,000 × g for 10 min and the supernatants were discarded properly without disturbing the protein pellets, then the pellets were washed with 70% cold (-20 °C) acetone three times. Discard the acetone and let the pellets stand for 10 min at RT in the fume hood to evaporate acetone. The protein pellets were redissolved in 0.1% Rapigest in 50 mM TEAB by thorough pipetting or sonicating. The proteins were digested using trypsin at an enzyme to protein ratio of 1:50 at 37 °C overnight. For the samples in buffer 2, the samples were diluted 8-fold with 50 mM ABC to a final concentration of urea less than 1M followed by overnight trypsin digestion. Then the digested peptides were desalted by Oasis HLB columns (Waters). For samples in buffer 3, trypsin was added into the reduced and alkylated protein solutions directly for overnight digestion. For samples in buffer 4, the protein solutions were diluted 10-fold with 50 mM TEAB before digestion. RapiGest in the digested peptide solutions from the samples in buffer 1, 3 and 4 would be decomposed after acidifying and then formed the pellets. The supernatant was transferred into a fresh tube after centrifugation at 14,000 g for 5 min. The purified peptide solution were then dried in a SpeedVac and stored in freeze until glycopeptide enrichment.

The PAC-based workflow was chosen as the final strategy for glycoproteome analysis, which includes SDS-assisted cell lysis and protein solubilization, PAC-based protein clean-up, on-beads proteolysis, TMT labeling, ZIC-HILIC-enabled glycopeptide enrichment, and basic reverse-phase (bRP) pre-fractionation of N-linked glycopeptides. Detailed procedures refer to each corresponding section.

Preparation of IgM peptides was described briefly. IgM was purified from normal human serum by precipitate and gel filtration techniques, stored at the concentration of 1 mg/ml in a solution of 0.05 M Tris-HCl, 0.2 M sodium chloride and 15 mM sodium azide, pH 8.0. The proteins were heated at 95 °C for 10 min. After cooling to RT, a final concentration of 10 mM TCEP and 20 mM IAM were added to the samples and incubated at 37 °C for 60 min in the dark. Trypsin was added at a trypsin to protein ratio of 1:50 for overnight digestion at 37 °C. The samples were acidified by adding 20% TFA to a final concentration of 1%. After centrifugation at 14,000 × g for 10 min, the supernatant was transferred to a fresh tube followed by desalting using Waters Oasis HBL columns. The purified peptides were dried in a SpeedVac concentrator and stored at -20 °C waiting for direct TMT labeling or MS analysis.

The IgM and yeast peptide interference model was performed as follows. Yeast protein prepared from S. cerevisiae cells were purchased from Promega (Cat. V7341). The proteins were solubilized in 6.5 M urea/50 mM Tris-HCl (pH 8) at a protein concentration of 10 mg/ml, and supplied as a frozen solution. After thawing the yeast proteins on ice, a final concentration of 10 mM TCEP and 20 mM IAM were added to the samples followed by incubation at 37 °C for 60 min in the dark. The reduced and alkylated protein solutions were diluted by the addition of 50 mM Tris-HCl (pH 8) to reduce the concentration of urea to less than 1 M. Trypsin was added at a trypsin to protein ratio of 1:50. After overnight incubation at 37 °C, the samples were acidified by adding 20% TFA to a final concentration of 1%. The insoluble particles in the samples were discarded by centrifugation at 14,000 × g for 10 min. The samples were purified with Waters Oasis HBL columns. The resulted peptides were dried in a SpeedVac concentrator.

We labelled six different protein amounts of IgM digests with six TMT6 channels reagents separately and pooled them together afterward with a ratio of 10:4:1:1:4:10 (20 μg : 8 μg : 2 μg : 2 μg : 8 μg : 20 μg). In contrast, 20 μg aliquot of yeast peptides were only labelled with three channels of TMT6 reagents (126, 127, and 128) and mixed equally. We then spiked the pooled IgM peptides into the yeast peptide mixture with an equal amount. The mixed samples were then cleaned up with Waters Oasis HBL columns followed by analysis with either Glyo-SPS-MS3 or standard MS2 methods.

3.2.3.5 Protein aggregation capture (PAC) method for protein clean-up

Three different types of magnetic beads, including Sera-Mag SpeedBeads with a hydrophilic surface (GE Healthcare, cat.no. 45152101010250), Sera-Mag SpeedBeads with a hydrophobic surface (GE Healthcare, cat.no. 65152105050250) and MagReSyn HILIC (ReSynBio, cat.no. MR-HLC005) with a mixed-mode functional surface were compared in this study. The magnetic beads were rinsed twice with water on a magnetic rack prior to use. The beads were added to the reduced and alkylated protein lysates at the working ratio of 10:1 (wt/wt, beads to proteins). In order to make sure the majority of aggregated proteins bind to the surface of beads, bead concentration should be more than 0.5 μg/μL. We then added acetonitrile to the mixture of protein lysates and beads to a final percentage of 70% (vol/vol). The samples were allowed to stay off the rack for 10 min at RT, followed by resting on the magnetic rack for 2 min at RT. The supernatant were discarded and the beads were then washed for three times with 95% (vol/vol) acetonitrile. Beads were resuspended in 50 mM TEAB containing sequencing grade modified trypsin (1:20~50 of enzyme to protein concentration) and incubated at 37 °C for 4 h or overnight in a ThermoMixer at 800 rpm. After digestion, we placed the tubes on a magnetic

rack for few minutes and collected the supernatants in fresh tubes. The last step was repeated one more to make sure that there were no residual beads in the peptide eluent.

3.2.3.6 Tandem Mass Tag (TMT) Labeling

TMT labeling procedure was performed according to the manufacturer's instruction (Thermo Scientific). Briefly, warm up the TMT reagents to RT immediately before use. 41 µL of acetonitrile was added to per 0.8 mg of the labeling reagent. Each obtained peptide solutions (100 µg ~ 400 µg) were mixed with one vial of labeling reagent (0.8 mg) and incubated for one hour at RT. The 8 µL 5% hydroxylamine were added to quench the reaction for 15 min at RT. Thereafter, samples labelled with different TMT-channels were combined and 10 % TFA was added to reach the final concentration of 0.5% TFA. Finally, samples were dried in a vacuum concentrator.

3.2.3.7 Phosphopeptides enrichment using TiO_2 beads

In brief, the dried samples was resuspended in incubation buffer (5% (vol/vol) glycerol in 80% ACN, 5% TFA). TiO_2 beads (10 µm; GL Science) at a 1:8 of peptide to bead ratio were weighed and washed three time using 1 ml of 60% ACN, 0.1% TFA, 1 ml of 80% ACN, 5% TFA and 1 ml incubation buffer sequentially. The samples were incubated with TiO_2 beads at RT for 20 min with end-over-end rotation. The peptide concentration was maintained at about 2 mg/ml during the incubation. After incubation, TiO_2 beads were loaded onto an empty spin column (5 µm frit; Hoefer Inc.) and washed three times each with incubation buffer, 80% ACN, 5% TFA and 60% ACN, 0.1% TFA. Phosphopeptides were then eluted with 1~5% (vol/vol) NH_4OH (pH ≥10.5) and acidified immediately with 10% TFA to adjust pH to less than 3. The eluate was desalted on a C18 spin column before fractionation.

3.2.3.8 Antibody-based phosphopeptides enrichment

Antibody-based enrichment for tyrosine-phosphorylated peptides was performed using the PTMScan Phospho-Tyrosine Rabbit mAB (P-Tyr-1000) Kit (CST) as described in the manufacturer's instructions. Briefly, lyophilized peptides were re-suspended in 1.4 ml 1 x IAP buffer. The pH of the peptide solution was checked by pH indicator paper to make sure that the pH was close to neutral. Then the peptide solution was cleared by centrifugation at 10,000 g at 4°C for 5 min. The antibody-bead slurry was washed four times with 1 ml 1 x PBS followed by incubation with the purified peptides for 2 h at 4 °C on a rotator. After incubation, the supernatant was saved as flow-through by centrifugation at 2,000 g for 30 seconds (sec). Beads were washed two times with 1x IAP buffer and three times with chilled HPLC water. All the centrifugation in wash steps was performed at 2,000 g for 30 sec at 4 °C. Add 0.2% TFA to the beads and incubate it for 10 min at RT, mixing gently every 2~3 min. The elution step was repeated and both eluents were combined for further use.

3.2.3.9 Antibody-based ubiquitinated peptides enrichment

Ubiquitinated peptides were enriched by the PTMScan Ubiquitin Remnant Motif (K-ε-GG) kit based on the manufacturer's protocol. All the steps are the same as for the tyrosine-phosphorylated peptides enrichment by the PTMScan Phospho-Tyrosine Rabbit mAB (P-Tyr-1000) Kit (refers to section 3.2.3.8).

3.2.3.10 Glycopeptides enrichment

The dried peptides were redissolved in loading buffer (80% ACN, 1% TFA) to maintain the peptide concentration at around 2~5 mg/ml. The peptide solution by centrifugation at 14,000 g for 5 min was

cleared at RT. Meanwhile, we weighted out the ZIC-HILIC beads (5 μm, Welch) depending on the peptide to bead ratio of 1:50, and loaded all beads in 80% ACN onto a 200 μL yellow pipette tip pre-packed with coffee filter followed by washing the beads three times using 200 μL loading buffer each time. Samples were loaded to pass through the HILIC column in the tip by centrifugation at 3000 rpm for 4~5 min. Reload the flow-through and repeat the loading step five times in total. The flow-through was collected for other analysis, such as phosphopeptides enrichment. Then the column was washed three times using 200 μL loading buffer separately. All the washing steps above were performed by centrifugation at 5000 rpm for 2~3 min at RT. The retained glycopeptides were eluted with 100 μL 0.1% TFA twice. The collected eluates were dried in a SpeedVac concentrator. For deglycoproteomics analysis, enriched glycopeptides were redissolved in 50 mM TEAB with 1,000 units PNGase F followed by incubating at 37 °C for at least 3 hours to overnight. Then the peptides were dried again and stored at -20 °C waiting for bRP prefractionation.

3.2.3.11 Basic reverse phase fractionation

Basic reverse phase analysis was performed on an Agilent 1100 series HPLC system. Enriched phosphopeptide and glycopeptide mixtures were dissolved in 50 μL mobile phase A (10 mM ammonium hydroxide in water, pH 10) separately. Elution was performed at a flow rate of 60 μL/min using mobile phase A and B (10 mM ammonium hydroxide in 80% ACN, pH 10) with a Waters XBridge C18 column (3.5 μm particles, 1.0 mm×150 mm). The gradient was 2% B for 5 min, to 34% B in 37 min, to 60% B in 8 min, to 90% B in 1 min, held at 90% B for 5.5 min, to 2% B in 0.5 min, then held at 2% B for 7 min (64 min total runtime). Peptides were detected at 214 nm and 58 fractions were collected along with the LC-separation in a time-based mode from 6 to 64 min. Fractions were finally grouped into 8 and 12 for enriched glycopeptide samples and enriched phosphopeptide samples separately. The merged fractions were dried down in a vacuum concentrator and were then stored at -80°C before use.

3.2.3.12 Strong Cation exchange (SCX) fractionation for ubiquitinated peptides

Firstly, we made a SCX microtip column by packing 4~6 pieces of Empore Cation 47 mm Extraction Disks into a 200 μL yellow pipette tip. The SCX microtip column was equilibrated by 200 μL washing buffer (50% ACN, 0.1% TFA), elution buffer 1 (pH 4.5), 2 (pH 5.5) and 3 (pH 8.5), washing buffer sequentially. Then loaded enriched ubiquitinated peptides onto SCX microtip column for twice followed by an additional washing. Peptides were eluted by stepwise 100 μL aliquots of elution buffer 1 (pH 4.5), 2 (pH 5.5) and 3 (pH 8.5). All the centrifugation in each step was performed at 5000 rpm for 2 min at RT.

3.2.4 LC-MS/MS analysis

LC-MS/MS analysis of the above mentioned samples was performed on the MS instruments including Q Exactive HF, Orbitrap Fusion and Orbitrap Fusion Lumos (Thermo Scientific), which were operated by the Bioanalytical Mass Spectrometry Group at the Max Planck Institute for Biophysical Chemistry. The generalized, most important instrument operating parameters are described in the following sections.

3.2.4.1 Q Exactive HF

Samples for ubiquitinome, pYome and GPome analysis were measured on a Q Exactive HF Hybrid Quadruple-Orbitrap Mass Spectrometer (ThermoFisher Scientific), coupled to a Dionex UltiMate 3000 UHPLC system (ThermoFisher Scientific) equipped with an in-house-packed C18 column (ReproSil-

Pur 120 C18-AQ, 1.9 μm, 75 μm × 30 cm, Dr. Maisch GmbH). Dried samples were resuspended in 5% ACN, 0.1% FA, then centrifuged for 14,000 g for 10 min and the supernatants were transferred to the sample vials. The samples were preconcentrated and desalted on a trap column (5 mm length, 30 μm inner diameter; Thermo Scientific) at 10 μL/min in loading buffer (2% ACN, 0.1% FA). Peptides were separated on a self-made capillary column (1.9 μm, 300 × 0.075 mm, ReproSil-Pur 120 C18-AQ, Dr. Maisch GmbH) using a different linear gradient at a constant flow rate of 300 nL/min. Mobile phase A consisted of 0.1% FA and mobile phase B was 80% ACN, 0.08% FA. The 118 min gradient used for ubiquitinome, pYome and proteome samples was as following: 2-8% B, 0-3 min; 8-36% B, 3-93 min; 36-60% B, 93-106 min; 60-90% B, 106-106.1 min; 90-90% B, 106.1-112 min; 90-2% B, 112-112.1 min; 2-2% B, 112.1-118 min. The 90 min gradient for GPome samples analysis was as following: 2-10% B, 0-3 min; 10-36% B, 3-72 min; 36-60% B, 72-76 min; 60-90% B, 76-76.1 min; 90-90% B, 76.1-82 min; 90-2% B, 82-82.1 min; 2-2% B, 82.1-90 min. The Q Exactive was operated in a data-dependent acquisition mode (DDA) using the top 20 most abundant precursors for higher energy collisional dissociation (HCD) in the collision cell with an isolation width of 1.6 m/z and an NCE setting of 28%. Full MS scan range was from m/z 350–1600. MS and MS/MS resolution settings were 120,000 and 30,000 FWHM at m/z 200. AGC target values and maximum injection time (IT) for MS and MS/MS were set to $1 \times e^6$ in 40 ms and $1 \times e^5$ in 128 ms, respectively. Fixed first mass and dynamic exclusion were set to 110.0 m/z and 20 s separately. All above MS parameters are commonly used for ubiquitinome, pYome and GPome analysis. Several exceptions were listed here: MS range of 300-1200 m/z was set for ubiquitinome. Isolation window of 1.4 m/z and dynamic exclusion of 25 s were used for GPome analysis.

Most of the MS parameters for proteome analysis are the same as described above except that MS/MS resolution setting, isolation window and maximum IT for MS/MS were 15,000 (m/z 200), 1.4 m/z and 32 ms separately.

3.2.4.2 Orbitrap Fusion and Orbitrap Fusion Lumos

Glycopeptides were analyzed on an Orbitrap Fusion Tribrid or a Lumos Mass Spectrometer (Thermo Scientific), coupled to a Dionex UltiMate 3000 UHPLC system (Thermo Scientific) equipped with in-house-packed C18 columns (ReproSil-Pur 120 C18-AQ, 1.9 μm pore size, 75 μm inner diameter, Dr. Maisch GmbH) with 30 cm length, or 50 cm length. The Orbitrap Fusion and Lumos were operated in data-dependent mode for both the standard MS2 method and the Glyco-SPS-MS3 method. Dried samples were redissolved in 5% ACN, 0.1% FA. Samples were centrifuged for 14,000 g for 10 min and the supernatants were transferred to sample vials. The flow rate was set to 300 nl/min. Mobile phase A consisted of 0.1% FA and mobile phase B was 80% ACN, 0.08% FA. The trap column (5 mm length, 30 μm inner diameter; Thermo Scientific) was not used in all experiments.

The 60 min gradient used for IgM samples was as following: The gradient started at 5% B and increased to 50% B in 42 min and further increased to 70% B in 4 min. The gradient then kept B constant at 90% for 6 min, followed by re-equilibration of the column with 5% B for 8 min.

The 118 min gradient for GPome analysis of DG75 cell samples upon fucosylation inhibition was as following: 5-8% B, 0-3 min; 8-40% B, 3-93 min; 40-60% B, 93-106 min; 60-90% B, 106-106.1 min; 90-90% B, 106.1-112 min; 90-5% B, 112-112.1 min; 5-5% B, 112.1-118 min.

The 118 min gradient for proteomics analyses of DG75 and DLBCL cell samples and deglycoproteomics analysis of DLBCL cell samples were as follows: two settings of analysis with trap column and without trap column were used. The gradient increased to 38% B from 5% B in 90 min,

and then to 60% B in 13 min. B was kept constant at 90% for 6 min, followed by re-equilibration of the column with 5% B. In no trap column setting, 50 min sample loading time at 2% B needed.

The 3 h gradient for glycoproteome analysis of DG75 and DLBCL cell samples was as follows. No trap column was used and sample loading time is 50 min at 2% B. The gradient kept at 5% B for 3 min, then increased to 34% B in 147 min, and to 60% B in 16 min. Kept B constant at 90% for 6 min, followed by re-equilibration of the column with 5% B for 8 min.

MS parameter optimization for the Glyco-SPS-MS3 method was performed including HCD NCE in MS2 and MS3, injection time allocation in MS2 and MS3, number of notches. Moreover, the detector type and the fragmentation type including Orbitrap_HCD, Ion trap_HCD, Orbitrap_CID and Ion trap_CID were compared. The details about the MS parameters optimization were be described in the results part.

The MS instrument settings for the optimized Glyco-SPS-MS3 method based on TMT6 labeling were described briefly in the following. MS1 settings: Detector Type-Orbitrap, Orbitrap Resolution-120 k, Mass Range (m/z)-350-2000, Maximum IT-50 ms, AGC target-$5e^5$, RF Lens-60%, Data Type-Profile, Precursor selection range (m/z)-700-2000; MS2 settings: Isolation mode-Quadrupole, Isolation window-2 m/z, Scan range mode-Auto normal, First mass-132, Activation type-HCD, Collision energy-25, Detector type-Orbitrap, Orbitrap resolution-15K, Maximum IT-150 ms, AGC target-$5e^5$, Data type-Profile; MS3 settings: Precursor selection range-700-2000, Number of Notches-10, Isolation mode-Quadrupole, Isolation window (m/z)-2, MS2 isolation window (m/z)-2, First mass (m/z)-120, Scan range mode-Auto normal, Activation type-HCD, Collision energy (%)-35, Detector type-Orbitrap, Orbitrap resolution-15 K, Maximum IT-350 ms, AGC target-$5e^5$, Number of Dependent Scans-10. In Glyco-SPS-MS3 method based on TMT10 labeling, resolution settings for MS2 and MS3 were 30,000 and 60,000 FWHM at m/z 200 separately.

Quantitative proteomics analyses based on TMT6 or TMT10 labeling in the glycoproteomics part were performed by both MS2 and SPS-MS3 methods. The parameters of MS2 method for TMT6 labeling were described as follows. MS1 settings: Detector Type-Orbitrap, Orbitrap Resolution-120000, Mass Range (m/z)-300-1600, RF Lens (%)-40, AGC Target-$1.0e^6$, Maximum IT (ms)-50, Data Type-Profile; MS2 settings: Isolation Mode-Quadrupole, Isolation Window (m/z)-1.6, Activation Type-HCD, HCD Collision Energy (%)-40, Detector Type-Orbitrap, Orbitrap Resolution-15000, First Mass (m/z)-110, AGC Target-$5.0e^4$, Maximum IT (ms)-22, Data Type-Centroid. The parameters of the SPS-MS3 method for TMT6 labeling were chosen as follows. MS1 settings: Detector Type-Orbitrap, Orbitrap Resolution-120000, Mass Range (m/z)-300-1600, RF Lens (%)-40, AGC Target-$5.0e^5$, Maximum IT (ms)-50, Data Type-Profile; MS2 settings: Isolation Mode-Quadrupole, Isolation Window (m/z)-1.2, Activation Type-HCD, HCD Collision Energy (%)-32, Detector Type-Orbitrap, Orbitrap Resolution-15000, First Mass (m/z)-132, AGC Target-$5.0e^4$, Maximum IT (ms)-60, Data Type-Centroid. MS3 settings: Number of Notches-10, Isolation mode-Quadrupole, Isolation window (m/z)-2, MS2 isolation window (m/z)-3, First mass (m/z)-120, Scan range mode-Auto m/z normal, Activation type-HCD, Collision energy (%)-65, Detector type-Orbitrap, Orbitrap resolution-15 K, Maximum IT-22 ms, AGC target-$5e^4$, Number of Dependent Scans-10. When samples were TMT10 labelled, MS2 resolution or MS3 resolution was adjusted to 60,000 FWHM at m/z 200 for MS2 method or SPS-MS3 method.

The GPome from the DG75 fucosylation inhibition model was measured on an Orbitrap Fusion using SPS-MS3 method. The MS instrument settings were as follows. MS1 settings: Detector Type-Orbitrap, Orbitrap Resolution-120 K, Scan Range-350-1550, Maximum IT-50 ms, AGC target-$4e^5$, RF Lens-60 %, Data Type-Profile; MS2 method settings: Isolation mode-Quadrupole, Isolation window-1.2 m/z,

Scan range mode-Auto normal, First mass-132, Activation type-HCD, Collision energy (%)-32, Neutral loss mass-97.9673, Detector type-Orbitrap, Orbitrap resolution-15 K, Maximum IT-60 ms, AGC target-$5e^4$, Data type-Profile; MS3 settings: Number of Notches-10, MS Isolation window (m/z)-2, MS2 isolation window (m/z)-3, First mass-120, Scan range mode-Auto m/z normal, Activation type-HCD, Collision energy-65, Detector type-Orbitrap, Orbitrap resolution-60 K, Maximum IT-105 ms, AGC target-$1e^5$, Number of Dependent Scans-10, Data type-Profile.

The Deglycoproteome from DLBCL cells by TMT10 labeling were measured on an Orbitrap Fusion Lumos using SPS-MS3 method. The MS instrument settings were as follows. MS1 settings: Detector Type-Orbitrap, Orbitrap Resolution-120 K, Scan Range-30-1600, Maximum IT-50 ms, AGC target-$4e^5$, RF Lens-40 %, Data Type-Profile; MS2 method settings: Isolation mode-Quadrupole, Isolation window-1.2 m/z, Scan range mode-Auto normal, First mass-132, Activation type-HCD, Collision energy (%)-32, Detector type-Orbitrap, Orbitrap resolution-15 K, Maximum IT-22 ms, AGC target-$5e^4$, Data type-centroid; MS3 settings: Number of Notches-10, MS Isolation window (m/z)-2, MS2 isolation window (m/z)-3, First mass-120, Scan range mode-Auto m/z normal, Activation type-HCD, Collision energy-65, Detector type-Orbitrap, Orbitrap resolution-60 K, Maximum IT-118 ms, AGC target-$5e^4$, Number of Dependent Scans-10, Data type-Profile.

3.2.5 MS raw file processing

In this study, we used Proteome Discoverer (PD, Thermo Scientific), MaxQuant [208, 209], Byonic [169], pGlyco [152, 154] and Glycobinder (which was available on GitHub (https://github.com/IvanSilbern/GlycoBinder)) for processing of all obtained MS raw files processing from the different projects.

3.2.5.1 Proteome analysis

All raw files for proteome analysis in the SILAC based quantitative ubiquitinome and phosphoproteome part were processed using MaxQuant software (v1.5.5.1, MPI for Biochemistry). MS/MS spectra were searched against a UniProtKB human database containing 92,954 protein entries (downloaded on Feb 2017) supplemented with 245 frequently observed contaminants via the Andromeda search engine. Raw files of different technical and biological replicates were defined in experimental groups and different multiplicity of SILAC labeling was set in parameter groups. Precursor and fragment ion mass tolerances were set to 6 and 20 ppm after initial recalibration, respectively. Protein N-terminal acetylation, and methionine oxidation were allowed as variable modifications. Cysteine carbamidomethylation was set as a fixed modification. Enzyme specificity was set to trypsin allowing N-terminal cleavage to proline. Minimal peptide length was set to seven amino acids, with a maximum of two missed cleavages. The false discovery rate (FDR) was set to 1% on peptide spectrum match (PSM) and protein level using a forward-and-reverse concatenated decoy database approach. For SILAC quantitation, multiplicity was set to double or triple (duplex labeling: Lys+0/Arg+0, Lys+8/Arg+10; one additional of Lys+4/Arg+6 for triplex labeling) labeling. At least two ratio counts were required for peptide quantitation. The “re-quantify” option of MaxQuant was enabled. Moreover, the “match between runs” option used in the default setting of a match time window of 0.7 min and an alignment time window of 20 min.

All raw files for proteome analysis in the glycoproteome part were processed using Proteome Discoverer (v2.1 and v2.2, Thermo Scientific). MS/MS spectra were searched against a UniProtKB/Swiss-Prot human database containing 20,315 protein entries (downloaded on March 2018) based on the decoy database searching approach. For non-TMT labelled samples, Spectrum files, spectrum selector, Sequest HT and Percolator were included in the processing workflow. Precursor

mass range was from 350 Da to 5000 Da. Trypsin was set for full digestion. Max missed cleavage sites were two and six was set for min peptide length. Precursor and fragment mass tolerance were set 10 ppm and 0.02 Da separately. Protein N-terminal acetylation, and methionine oxidation were allowed as variable modifications. Cysteine carbamidomethylation was set as a fixed modification. In the consensus workflow seven nodes were used including MSF files, PSM grouper, peptide validator, peptide and protein filter, protein scorer, protein FDR validator and protein grouping. For the validation, protein, PSMs and peptides FDR were set with strict FDR of 0.01 and relaxed FDR of 0.05. For TMT labelled samples, the node of reporter ions quantifier was added into processing workflow. The option of MS order was set depending on the MS methods used for TMT quantitation, such as MS2 or MS3. In Sequest HT node, TMT6plex or TMT10plex at any N-terminus and lysine was set as the static modification. For the labeling efficiency analysis, TMT was set as the dynamic modification.

3.2.5.2 Phosphoproteome analysis

All raw files for phosphoproteome analysis including GPome and pYome in DLBCL cell lines based on SILAC labeling were processed using MaxQuant software (v1.5.5.1, MPI for Biochemistry) against a UniProtKB human database containing 92,954 protein entries (downloaded on Feb 2017) supplemented with 245 frequently observed contaminants via the Andromeda search engine. All raw files were searched in one MaxQuant window. Experimental group, fraction part and parameter groups were defined by the replicates of GPome and pYome, fractions from each replicate and the multiplicity of SILAC labeling respectively. Most of the parameters settings are the same with those in the proteome raw files processing by MaxQuant (refers to section 3.2.5.1). In addition, phosphorylation on Serine (S), Threonine (T) and Tyrosine (Y) was set as the variable modification.

All raw files for phosphoproteome analysis in the DG75 fucosylation inhibition model based on TMT labeling were processed using MaxQuant (v1.6.5.0, MPI for Biochemistry) against a UniProtKB/Swiss-Prot human database containing 20,315 protein entries (downloaded on March 2018). Parameters were set as follows. Fully specific trypsin digestion allowing N-terminal cleavage to proline with maximal two missed cleavage was set and mass tolerance for precursors and fragment ions were 10 and 20 ppm, respectively. Protein N-terminal acetylation, and methionine oxidation were allowed as variable modifications. Cysteine carbamidomethylation was set as a fixed modification. Minimum peptide length was seven amino acids. At least two ratio counts were required for peptide quantitation. The “re-quantify” option of MaxQuant was enabled. Moreover, the “match between runs” option used in the default setting of a match time window of 0.7 min and an alignment time window of 20 min. At least two ratio counts were required for peptide quantitation. The “re-quantify” option of MaxQuant was enabled. Moreover, the “match between runs” option used in the default setting of a match time window of 0.7 min and an alignment time window of 20 min. For TMT quantitation, type of reporter ion MS3 for 6plex TMT was set. Correction factors were set depending on the information on the product data sheet (Thermo Fisher).

3.2.5.3 Ubiquitinome analysis

All raw files for ubiquitinome analysis in DLBCL cell lines based on SILAC labeling were processed using MaxQuant (v1.5.5.1, MPI for Biochemistry) against a UniProtKB human database containing 92,954 protein entries (downloaded on Feb 2017) supplemented with 245 frequently observed contaminants via the Andromeda search engine. Most of the parameter settings were identical to the ones used for the phosphoproteome analysis in DLBCL cell lines based on SILAC labeling. All raw files from different cell lines and treatments were searched in one MaxQuant window with defined experiment, fraction and parameter settings. Di-glycine on lysine (KGG) was set as the variable

modification instead of STY phosphorylation. For SILAC quantitation, multiplicity was set to two for double (Lys+0, Lys+6) labeling. At least two ratio counts were required for peptide quantitation. Both the "match between runs" and "re-quantify" options of MaxQuant were enabled.

3.2.5.4 Glycoproteome analysis

All raw files for intact glycopeptide identification in DG75 and Daudi samples and IgM were processed by pGlyco 2. For quantitative glycoproteome analysis in TMT6plex or TMT10plex-labelled DG75 and DLBCL cell lines, all raw files were processed by GlycoBinder (MPI for Biophysical Chemistry), which is available on GitHub (https://github.com/IvanSilbern/GlycoBinder). The main parameters for pGlyco 2 were set as follows: fully specific trypsin digestion with maximal two missed cleavage, mass tolerance for precursor ions and fragment ions of 10 and 20 ppm, respectively, cysteine carbamidomethylation and TMT0/6/10-plex on peptide N-terminal and lysine residues as fixed modifications and methionine oxidation as a variable modification. All the parameters for GlycoBinder searching are available on the GitHub page. The human protein database (Human-Specific) was downloaded from Swiss-Prot (March 2018, human, 20,303 entries). In addition, in the SugarQuant part, three other specific databases were used. The "Glycoprotein-reviewed" (4824 entries, March 2018) database was downloaded from Uniprot using the keyword "glycoprotein". The database "Random-1000" was generated by random selection of 1000 protein sequences from the "Human-reviewed" database using an R base function. The "B-cell-specific" database was built using the glycoproteins identified from about 250 runs for DG75 or Daudi cells in our lab.

For the identification of IgM glycopeptides, IGHM_HUMAN and IGJ_HUMAN in the FASTA file were used as the database. For IgM identifications, we considered only GPSMs with PepScore > 7 and GlyScore > 8 reported by pGlyco 2 for further analysis. For the complex samples of different cell lines, we used the total FDR < 2% for both the first and second database search. The reported "TotalScore", "PepScore", "GlyScore", "GlyIonRatio" and "PepIonRatio" were for the evaluation of identification confidence. PepScore means the scores for peptide sequences based on b/y ion. GlyScore means the scores for glycan composition based on Y ions, and TotalScore means the scores for intact glycopeptide based on the weighted sum of the PepScore and GlyScore according to the pGlyco soring algorithm. GlyIonRatio and PepIonRatio means the ratios = #matched Y ions/#theoretical Y ions and #matched peptide ions/#theoretical peptide ions, respectively.

For processing Glyco-SPS-MS3 results using PD, the nodes of "Spectrum selector" and "Spectrum grouper" were used for converting and combining spectra from MS2 and MS3, respectively. The parameters for spectrum grouper node were listed as follows. Precursor mass criterions: same singly charged mass; Precursor mass tolerance: 0.1 ppm; Max. RT difference: 0.04 min; Allow mass analyzer mismatch: False; Allow MS order mismatch: True. The resulting mgf files were used for pGlyco 2 searches.

For the comparison of the detector type and fragmentation type in MS2 methods, Proteome Discoverer was used to convert the raw files to mgf and separate the MS2 scan depending on their separate settings. Then the resulting mgf files were searched using Byonic software against IGHM_HUMAN and IGJ_HUMAN database. The parameters included a 10 ppm of precursor ion tolerance and 20 ppm of fragment ion tolerance for Orbitrap and 0.5 Da of fragment ion tolerance for Iontrap, full trypsin specificity on both termini, up to two missed cleavages, carbamidomethylation of cysteine residues (+57.02146 Da) as static modifications, oxidation of methionine residues (+ 15.99492 Da) as a variable modification.

For the IgM-Yeast interference model, the raw MS files were processed using GlycoBinder and searched against a protein database including sequences from IGHM_HUMAN and IGJ_HUMAN followed by sequences of proteins encoded by all known S.cerevisiae ORFs. FDR < 2% for both the first and second database search was used.

For deglycoproteomics analysis of TMT10 labelled DLBC samples, all raw files were processed using MaxQuant (v1.6.5.0, MPI for Biochemistry) against a UniProtKB/Swiss-Prot human database containing 20,315 protein entries (downloaded on March 2018). Parameters were set as follows. Fully specific trypsin digestion allowing N-terminal cleavages to proline with maximal two missed cleavage was set and mass tolerance for precursors and fragment ions were 10 and 20 ppm, respectively. Protein N-terminal acetylation, asparagine deamidation and methionine oxidation were allowed as variable modifications. Cysteine carbamidomethylation was set as a fixed modification. Minimum peptide length was seven amino acids. At least two ratio counts were required for peptide quantitation. The "re-quantify" option of MaxQuant was enabled. Moreover, the "match between runs" option used in the default setting of a match time window of 0.7 min and an alignment time window of 20 min. At least two ratio counts were required for peptide quantitation. For TMT quantitation, type of reporter ion MS3 for 10plex TMT was set. Correction factors were set depending on the information on the product data sheet (Thermo Fisher). N-linked glycosite-containing peptides were required to contain N-!P-S/T/C motifs.

3.2.6 Data analysis

3.2.6.1 N-glycoproteome analysis in SugarQuant

Unique glycoforms, glycosites, and glycan compositions which contain the TMT reporter ions intensities were output in separate text files. To account for quantitative errors introduced before TMT-labeling, TMT-ratios of glycoforms determined in each glycoproteomics experiment on DG75 cells were normalized by the median TMT-ratio of respective proteomics analysis. Significant regulation is defined using Z score >2 or <-2. Unless mentioned elsewise, all figures were made using Excel 2016.

For the sunburst plot, we investigated and annotated the protein functions of 2FF-affected glycoproteins manually via surveying literature relevant to the respective gene and protein names (including alternative names) and keywords of "cancer", "lymphoma" or "B cell". We summarized the information about the biological function(s) and subcellular localization of all proteins and classified them accordingly into multiple categories. For instance, the first inner layer contains the groups of lymphoma-related, drug target in other cancer, enzymes and others. The number of proteins in each category is proportional to the size of the corresponding categorical area in the sunburst plot. The interaction networks of proteins were done by STRING (V 11.0) [210].

Boxplots, violin plots and density plots were created using Origin 2020. In the boxplots, centerlines and squares in plotted boxes indicate the median and mean, respectively. The upper and lower ends of the boxes show the 75th and 25th percentiles. The extreme line shows 1.5× the interquartile range. In violin plots, white circles in black boxes indicate the median. The upper and lower ends of the black boxes show the 75th and 25th percentiles. The extreme line shows 1.5× the interquartile range.

3.2.6.2 Phosphoproteome analysis in 2FF-treated DG75

Phospho (STY) Sites.txt table from MaxQuant was pre-processed with Perseus (V 1.6.5.0) [211, 212]. After the removal of 'Reverse' and 'contaminant' entries, the class I phosphosites with localization probability ≥0.75 were filtered. The function of "Expand site table" was used to convert the

phosphosites into phospho-events (p-events). Two-sample t-test in the volcano plot was performed to filter the significantly regulated p-events using the log2 transformed ratios of total p-events with the FDR less than 0.5% and S0 of 0.1. KEGG pathway analysis was performed by STRING (V 11.0) [210] with the medium confident required interaction score of 0.4.

3.2.6.3 Glycoproteome analysis in DLBCL

Histogram analysis for protein expression, Hierarchical clustering in Heatmap, Principal component analysis (PCA), Volcano plot and the Pearson correlation analysis were performed in Perseus (V 1.6.10.43). The truncated graphs were made in Origin 2020. The overlap analysis among different replicates was made on the website of VENNY 2.1(https://bioinfogp.cnb.csic.es/tools/venny/). All the other figures were made using Excel 2016. The details of special data filtering could be found in the corresponding result part.

3.2.6.4 Phosphoproteome analysis in DLBCL

Phospho (STY) Sites.txt table from MaxQuant was pre-processed with Perseus (V 1.6.10.43). After the removal of 'Reverse' and 'contaminant' entries, the class I phosphosites with localization probability ≥0.75 were filtered. The function of "Expand site table" was used to convert the phosphosites into phospho-events (p-events). Histogram analysis for protein expression, Hierarchical clustering in Heatmap, Principal component analysis (PCA) and the Pearson correlation analysis were performed in Perseus (V 1.6.10.43). The truncated graphs were made in Origin 2020. CV profiling was made in Excel 2016. All the left analyses were performed using R Markdown which was created by Dr. Federico Comoglio (enGene Statistics, Zurich, Switzerland). Pathway enrichment analysis was performed using Entrez gene identifiers (converted from UNIPROT identifiers) and REACTOME database terms.

3.2.6.5 Ubiquitinome analysis in DLBCL

GlyGly (K) Sites.txt table from MaxQuant was used to obtain the ubi-events using the class I sites after the removal of contaminant and reverse entries in Perseus (V 1.6.10.43). Similar to phosphoproteome analysis workflow, histogram analysis for protein expression, hierarchical clustering in heatmap, principal component analysis (PCA) and the pearson correlation analysis were performed in Perseus (1.6.10.43). The truncated graphs were made in Origin 2020. CV profiling was made in Excel 2016. All the other analyses were performed using R Markdown which was created by Dr. Federico Comoglio.

4. Results

Post-translation modifications (PTMs) are very important biological mechanisms for regulating various signaling pathways and controlling cell fates. MS-based quantitative proteomics was used for systematic investigation of different PTMs including glycosylation, phosphorylation and ubiquitination in diffuse large B-cell lymphoma and Burkitt's lymphoma under several treatment conditions on a large-scale.

4.1 The establishment of quantitative N-glycoproteomics platform of SugarQuant

Previous MS based-studies for glycoproteomics mainly focused on the released glycan chains (glycomics) or the deglycosylated peptides (deglycoproteomics) separately [148]. However, the determination of the linkage between protein and glycan chain and the characterization of the micro-heterogeneity of glycosylation in a site-specific manner are necessary by the analysis of intact N-glycopeptides. Until now, a complete pipeline involving accurate quantification of intact glycopeptides in a large scale and automatic data analysis is not mature. With the unique advantages of chemical labeling strategy in intact glycopeptide analysis (refers to section 1.3.1.5), both iTRAQ and TMT reagents have been used for large-scale quantitative glycoproteomics studies [151, 153]. However, a systematic optimization of MS acquisition parameters specific for labelled glycopeptides is still missing. Moreover, MS2 based isobaric chemical labeling leads to less reliable identification and quantification of intact glycopeptides due to co-isolated interference. Recently developed synchronous precursor selection (SPS)-MS3 technology has improved the quantitative accuracy by reducing co-isolated precursor interference [126, 183], but its application in quantitative N-glycoproteomics is still not surveyed. Finally, the additional chemical-labeling reaction makes the conventional glycopeptide preparation procedure even more laborious and troublesome. All the technical obstacles plus the lack of a simple streamlined workflow for processing and interpreting the data have hampered the development of quantitative glycoproteomics. Accurate quantitation of intact N-glycopeptides in systematic level has remained technically challenging, and software tools for processing quantitative glycoproteomics MS data are still lacking.

We developed a multiplexed quantitative N-glycoproteomics platform named SugarQuant, which is an integrated workflow including fast and efficient sample preparation, optimized multi-notch SP3 (Glyco-SPS-MS3) acquisition method and streamlined data-processing tool (GlycoBinder) for the more accurate identification and quantitation of intact N-glycopeptides on a large scale.

4.1.1 The general workflow of SugarQuant

SugarQuant includes three parts: 1. Protein aggregation capture (PAC) based sample preparation; 2. Glyco-SPS-MS3 analysis of multiplexed intact glycopeptides; 3. R-based script of GlycoBinder (available on GitHub (https://github.com/IvanSilbern/GlycoBinder)) for automatically qualitative and quantitative data processing (**Figure 4.1**).

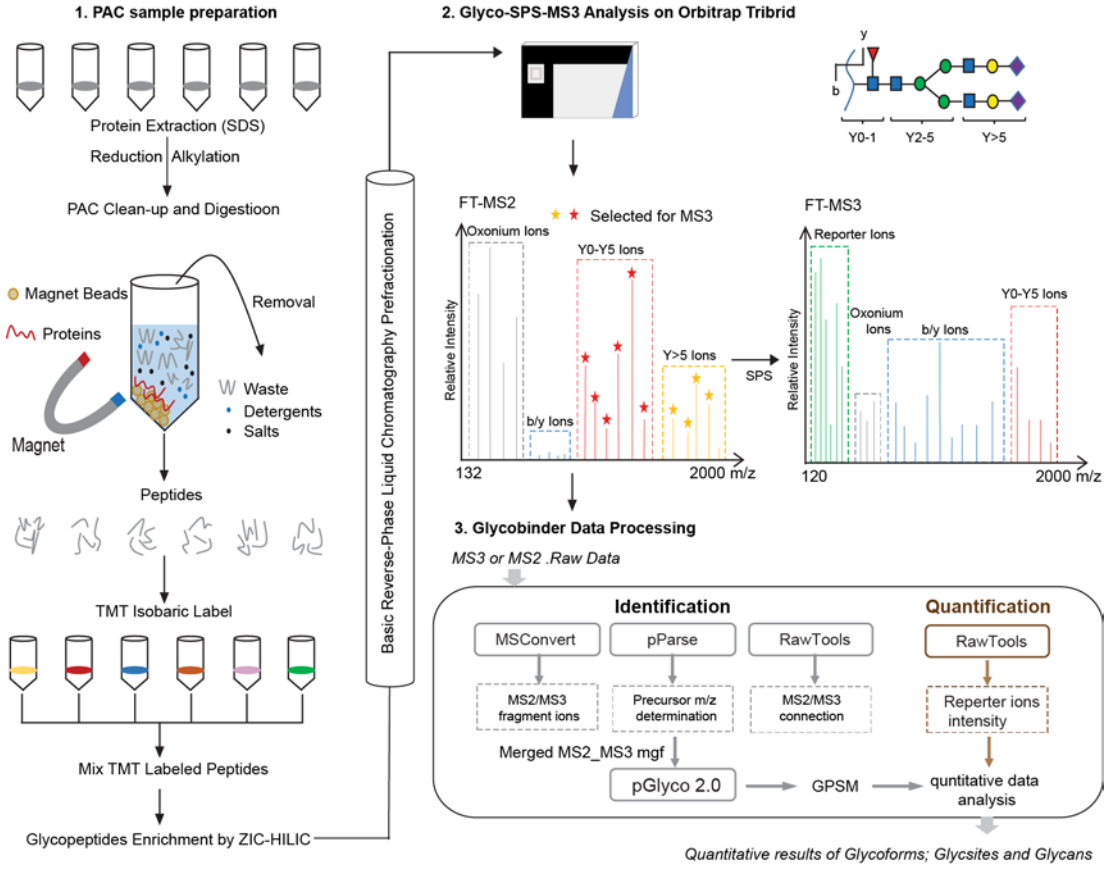


Figure 4.1 Schematic of SugarQuant for multiplex quantitative N-glycoproteomics.

The workflow of SugarQuant includes PAC-based method for fast sample preparation of TMT labelled glycopeptides, Glyco-SPS-MS3 acquisition method for confident identification and accurate quantification of intact N-glycopeptides and the R-based script of GlycoBinder (available on GitHub (https://github.com/IvanSilbern/GlycoBinder)) for automatic data processing.

In the first part, cell pellets were lysed in the 4% SDS lysis buffer for protein extraction. Detergents clean-up and endoproteolytic digestion was performed by PAC [213-215]. The digested peptides were multiplex TMT labelled followed by ZIC-HILIC glycopeptide enrichment. In order to improve the throughput of identification, enriched TMT labelled glycopeptides were fractionated by basic reverse phase (BRP) based prefractionation for MS analysis. In the second part, a modified SPS-MS3 method, namely Glyco-SPS-MS3 was used for multiplex intact N-glycopeptides quantitation with fine-tuned MS acquisition parameters. Compared to previous MS/MS approaches, our Glyco-SPS-MS3 showed increased glycopeptide fragment ions information on spectrum and less co-isolated precursor interference, which improved the quality of identification and the accuracy of quantification, separately. In the third part, we developed a script using R programming language, namely GlycoBinder, which integrates seamlessly several existing software packages including MSConvert [216], pParse [217], RawTools [218] and pGlyco2.0 [154]. It enables to combine MS2 and MS3 fragments ions for N-glycopeptide identification and extract TMT reporter-ion intensities from MS3 scans for each N-glycopeptide-spectrum-match (GPSM). GlycoBinder unifies redundant GPSMs with their quantitative values and reports different tables for unique glycoforms, glycosylated sites and glycan compositions on a multi-dimensional quantitation level. Finally, we demonstrated the capability of SugarQuant to accurately determine the global and site-specific changes of protein glycosylation in human Burkitt's lymphoma cells upon fucosylation inhibition.

4.1.2 The PAC-based method enables highly efficient preparation of TMT-labelled glycopeptides

In this part, we evaluated and optimized currently available experimental workflows specific for TMT-labelled N-glycopeptides (refers to section 5.1.1). As a result, we simplified the laborious sample preparation of intact N-glycopeptides from complex samples. PAC-based sample preparation includes sodium dodecyl sulfate (SDS)-assisted cell lysis and protein extraction, MagHILIC-based protein clean-up, tryptic digestion on beads, TMT labeling, N-glycopeptide enrichment by ZIC-HILIC, and bRP prefractionation (**Figure 4.1**).

This workflow has outstanding benefits in terms of much shorter sample treating time, reduced sample loss, more efficient membrane proteins extraction and perfect compatibility with TMT labeling. We used 4% SDS to solubilize largely membrane-associated glycoproteins and showed that lysis with SDS recovered 9% more unique proteins and 20% more unique glycoforms from human Burkitt's lymphoma DG75 cells than 1% RapiGest, respectively (**Figure 4.2a**). After protein extraction with SDS, we removed detergents and unnecessary substances followed by tryptic digestion in a PAC-based method (refers to section 3.2.3.5). Then we compared the protein retaining capacities of three magnetic beads bearing various surface functional groups. The results showed that all beads functioned undistinguishably and resulted in similar numbers of protein and unique glycoforms identifications (**Figure 4.2b**) [214]. In order to decrease the extent of beads sticking to the tube wall during overnight digestion, the proteins were digested on beads only for four hours. The number of identified proteins using PAC method with 4h digestion is comparable to the overnight digestion in acetone precipitation (AP) method while the percentages of identified glycoforms increased by 13.8% in spite of the expected rised missed cleavage rate (**Figure 4.2c**). Subsequently, the digested peptides were transferred to a new tube and labelled by multiplex TMT reagents. Then we enriched the TMT labelled glycopeptides by ZIC-HILIC beads from the mixed TMT labelled samples. This showed that TMT-labeling did not affect the ZIC-HILIC-based glycopeptide enrichment efficiency and specificity when enough ion pair (TFA) reagents were added to acidify the labelled samples before enrichment (**Figure 4.2d**).

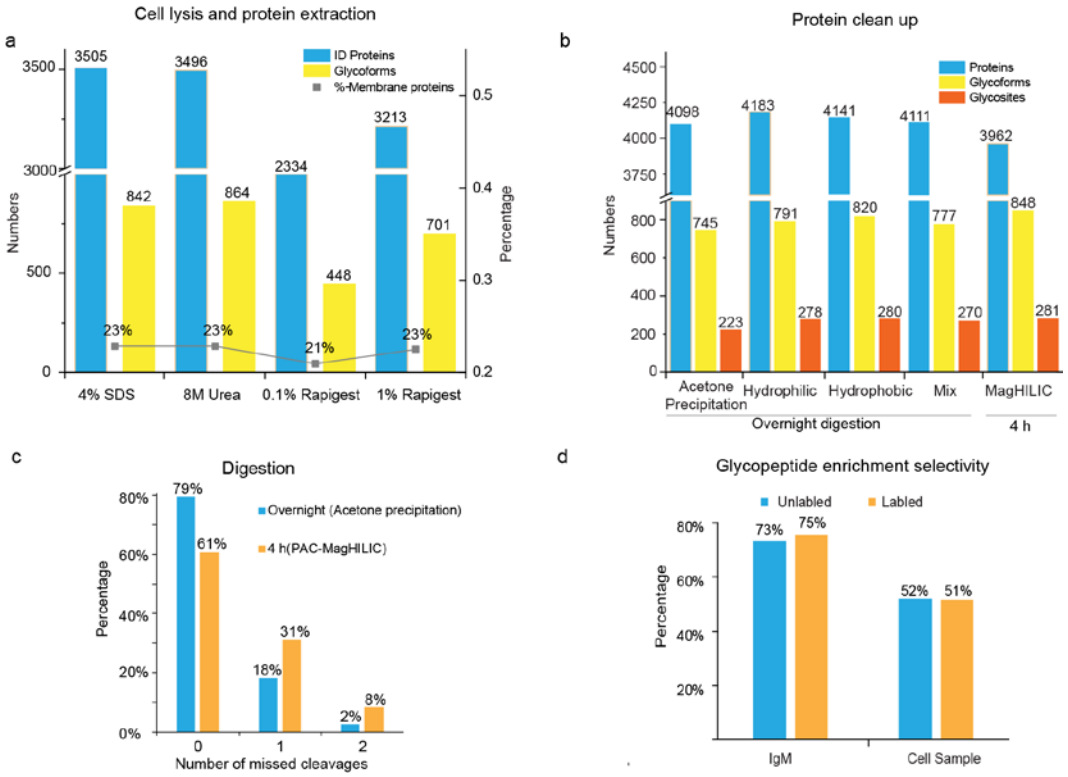


Figure 4.2 Optimization of experimental parameters for sample preparation.

(a) Four different lysis buffers were used for protein extraction and solubilization from DG75 cells. The sample preparation methods are described in section 3.2.3.4. The number of identified proteins and glycoforms are shown in blue and yellow, respectively. Among the identified proteins, percentages of membrane-associated proteins are also shown. (b) We compared the conventional acetone precipitation with the PAC methods using various types of beads (refers to section 3.2.3.5) for protein clean-up. The numbers of identified protein, glycoforms, and glycosites are shown in blue, yellow, and orange, respectively. Note that the PAC method using MagHILIC was performed for 4 h digestion while others were digested overnight. (c) The distribution of number of missed cleavages among the AP-overnight digestion (blue) and the PAC-MagHILIC methods (orange). (d) Comparison of glycopeptide enrichment specificity between unlabelled (blue) and labelled samples (orange) from IgM or cell samples. The enrichment efficiency was defined by the ratio of glycan-oxonium-ion-containing spectrum to all MS2 scans in a LC-MS/MS run. The results for IgM and cell samples were derived from triplicate and duplicate sample preparations respectively.

Off-line bRP for peptides prefractionation was demonstrated to significantly improve the depth of proteome and phosphoproteome analyses [219, 220]. Similarly, it brings its benefits to deep N-glycoproteomics analysis. Based on an equal initial protein amount and total MS measuring time (20 hours), bRP prefractionation led to a 53.5 % increase of glycopeptide identification compared to multiple injections with a longer LC-MS/MS gradient (**Figure 4.3a**). This result can be attributed to the high separation efficiency of bRP, where 94.1% of the N-glycopeptides were uniquely identified in less than two fractions (**Figure 4.3b**). In contrast, almost one-third (32.9%) of glycopeptides were repetitively identified in all five injection replicates, and only 26.7% of the glycopeptides were uniquely identified in a single run (**Figure 4.3c**). Furthermore, our results show that adjustment of the LC gradient is necessary to reflect the increased hydrophobicity of TMT-labelled glycopeptides in an LC-MS/MS run (**Figure 4.3d**). Finally, MS analysis without trap column on LC identified more glycopeptides than that with trap column, due to that extremely hydrophilic or hydrophobic glycopeptides are easily lost during the trapping step (**Figure 4.3e**).

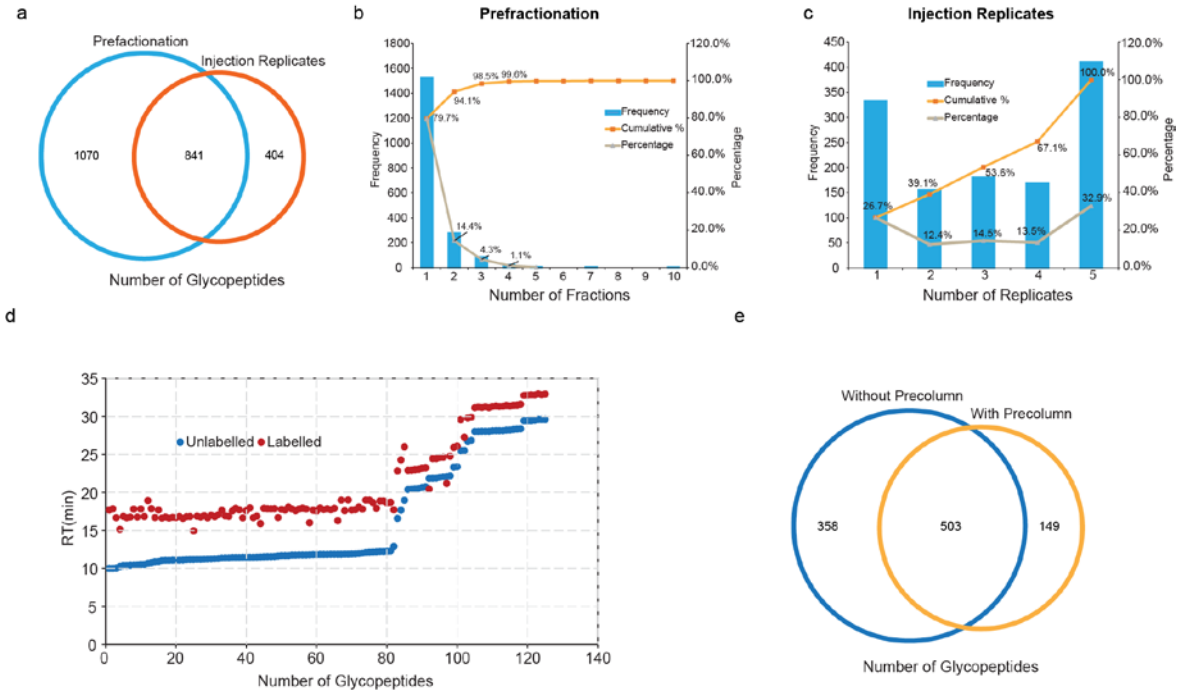


Figure 4.3 Investigation of optimal experimental conditions for sample preparation.

(a) Comparison of the glycopeptide identifications achieved by bRP prefractionation or up to five injection replicates using a 50 cm analytical column and longer gradient (4 hours) without prefractionation. (b) The distribution of identified glycopeptides in different numbers of fractions. 79.7% of total glycopeptides were identified in only one fraction, and 14.4% were identified in two fractions. (c) The distribution of identified glycopeptides in different numbers of injection replicates. 26.7% were exclusively identified in one of the injections, and 32.9% were identified in all five injection replicates. (d) Retention time of TMT labelled glycopeptides and unlabelled ones from IgM under the same LC gradient. (e) Comparison of N-glycopeptide MS analysis in the absence or presence of a trap column.

Compared to the 2~3 days sample handling time for conventional acetone-precipitation (AP) based approach for, our optimized workflow resulted in a reduction of the total handling time to one day (**Figure 4.4**). One key factor was the shortened proteolysis time to 4 hours. Despite the reduced handling time, the PAC-based method still recovered 4-14% more glycoforms and 21-26% more N-glycosites than the conventional AP-based method (**Figure 4.2b**), suggesting fewer sample loss compared to AP. Importantly, the PAC-method was for the first time applied to large-scale quantitative glycoproteomics in complex sample. In summary, the optimized workflow including cell lysis, PAC-based clean-up and proteolysis, TMT-labeling, N-glycopeptide enrichment by ZIC-HILIC and off-line pre-fractionation of N-glycopeptides prior to LC-MS/MS analysis enables large-scale multiplexed quantitative N-glycoproteomics with higher throughput and sensitivity in one day.

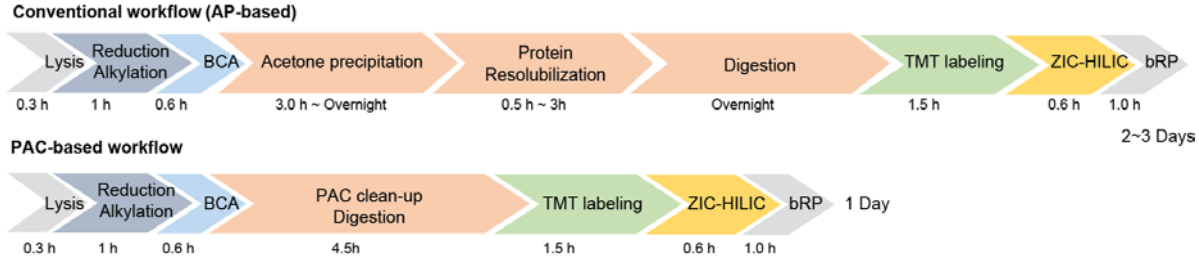


Figure 4.4 Timelines of the conventional AP-based and optimized PAC-based workflow.

It is important to note that the time required for lyophilization is variable in a sample dependent manner and thus not included here.

4.1.3 Development of a Glyco-SPS-MS3 strategy for improved identification and quantification of TMT-labelled N-glycopeptides

In the past few years, continuous developments of new MS methods such as new fragmentation types and more efficient searching engines have improved the characterization of intact N-glycopeptides significantly [147]. However, such improvements don't survey the characteristics of the quantitative TMT-labelled counterparts. Therefore, we used purified IgM digests as a model to evaluate the MS characters of TMT-labelled and unlabelled N-glycopeptides on Orbitrap Tribid mass spectrometers in order to obtain the optimal MS acquisition settings, which are applicable to the multiplexed isobaric labelling quantitation of N-glycoproteomics on a large scale.

Our results showed that different normalized collision energies (NCEs) were needed in MS2 analysis using HCD for the optimal fragmentation of unlabelled and TMT labelled N-glycopeptides from IgM. 30-35% NCE led to the largest numbers of identified labelled N-glycopeptides, while 25-30% were the best settings for non-labelled glycopeptides (**Figure 4.5a**). Moreover, a higher NCE was required for TMT-labelled N-glycopeptides to obtain better glycan scores (**Figure 4.5b**).

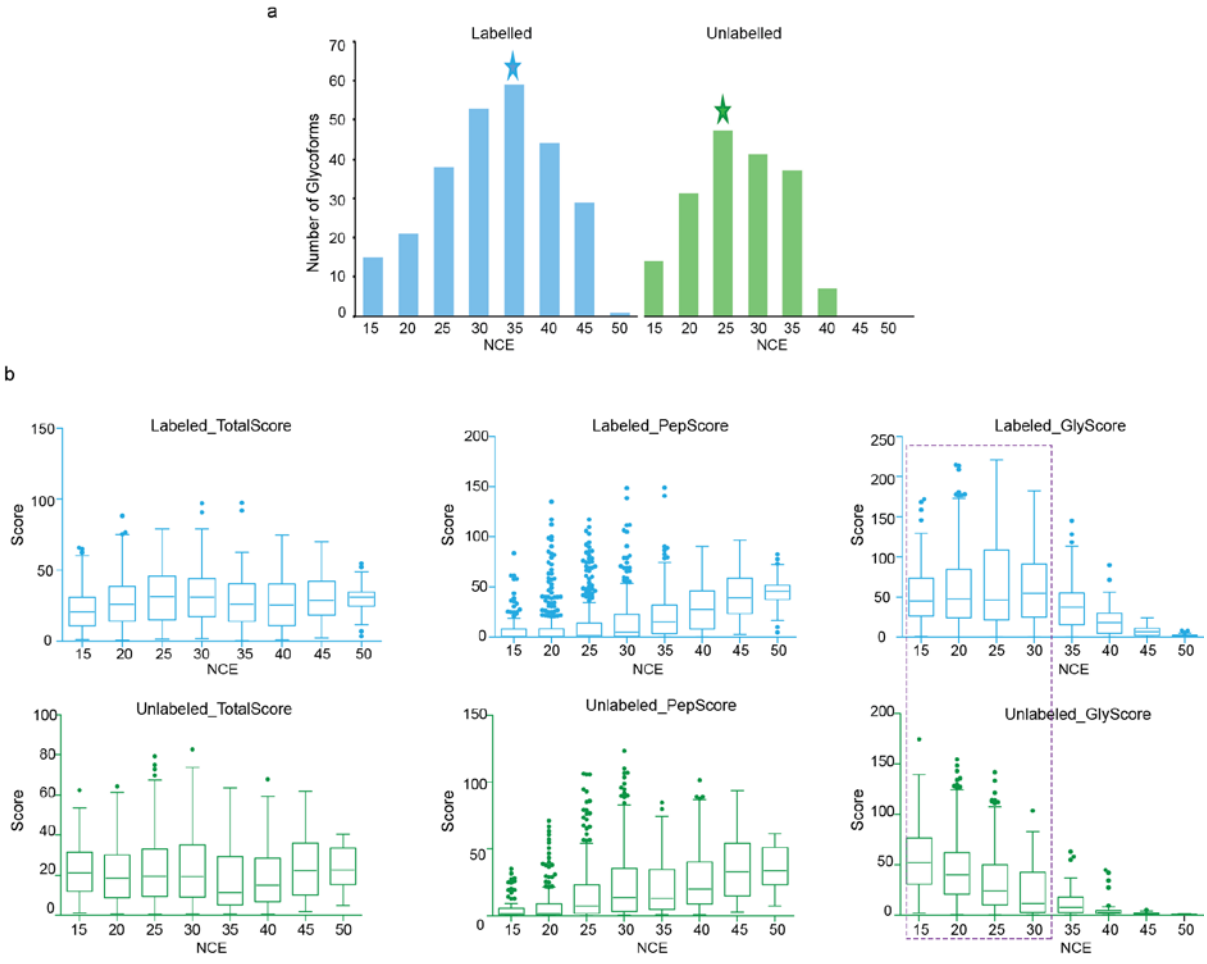


Figure 4.5 Optimized NCE settings for TMT-labelled glycopeptides in MS2 analysis using HCD.

(a) Comparison of the identified glycoforms from labelled and unlabelled IgM samples under different normalized collision energies (NCEs). Each condition includes two technical replicates. (b) Box plots show the distributions of total scores (left), peptide scores (PepScore, middle) and glycan scores (GlyScore, right) of TMT labelled (upper pannels) and non-labelled (lower panels) glycopeptides from IgM identified under various HCD NCEs in LC-MS/MS runs. The purple dashed line highlights the significantly different glycan score distribution caused by TMT-labeling.

Indeed, we carefully evaluated the NCE settings for all potential fragment ions from twelve labelled IgM glycopeptides (**Figure 4.6 a**) in MS/MS analysis, in order to obtain a better understanding of the fragmentation behaviors of TMT-labelled glycopeptides. All potential fragment ions were categorized into five types based on the HCD fragmentation characteristics for TMT-labelled N-glycopeptides, including Y>5, Y0-1, Y2-5, b/y, and TMT reporter (**Figure 4.6b**). The definition of each type was listed as following: Y0-1, intact peptide backbone attached with zero to 1 monosaccharide, including Y0, $^{1,2}X_0$ ions (cross-ring fragmentation) and Y1; Y2-5, intact peptide backbone attached with N-glycan core structure, including Y2, Y3, Y4, Y5 and Y1f, Y2f, Y3f, Y4f and Y5f (f is fucose); Y>5, intact peptide backbone with N-glycans extending from the N-glycan core structure; b/y ions, peptide fragments without glycan attached. We found that the favored NCEs for generating different types of fragment ions varied significantly. 15-20% and 30-35% NCE were suitable for the generation of bigger (Y>5) and smaller (Y0-1 and Y2-5) glycan product ions respectively, 40-45% NCE for peptide fragment (b/y) ions and above 45% NCE for reporter ions (**Figure 4.6 c, d**).

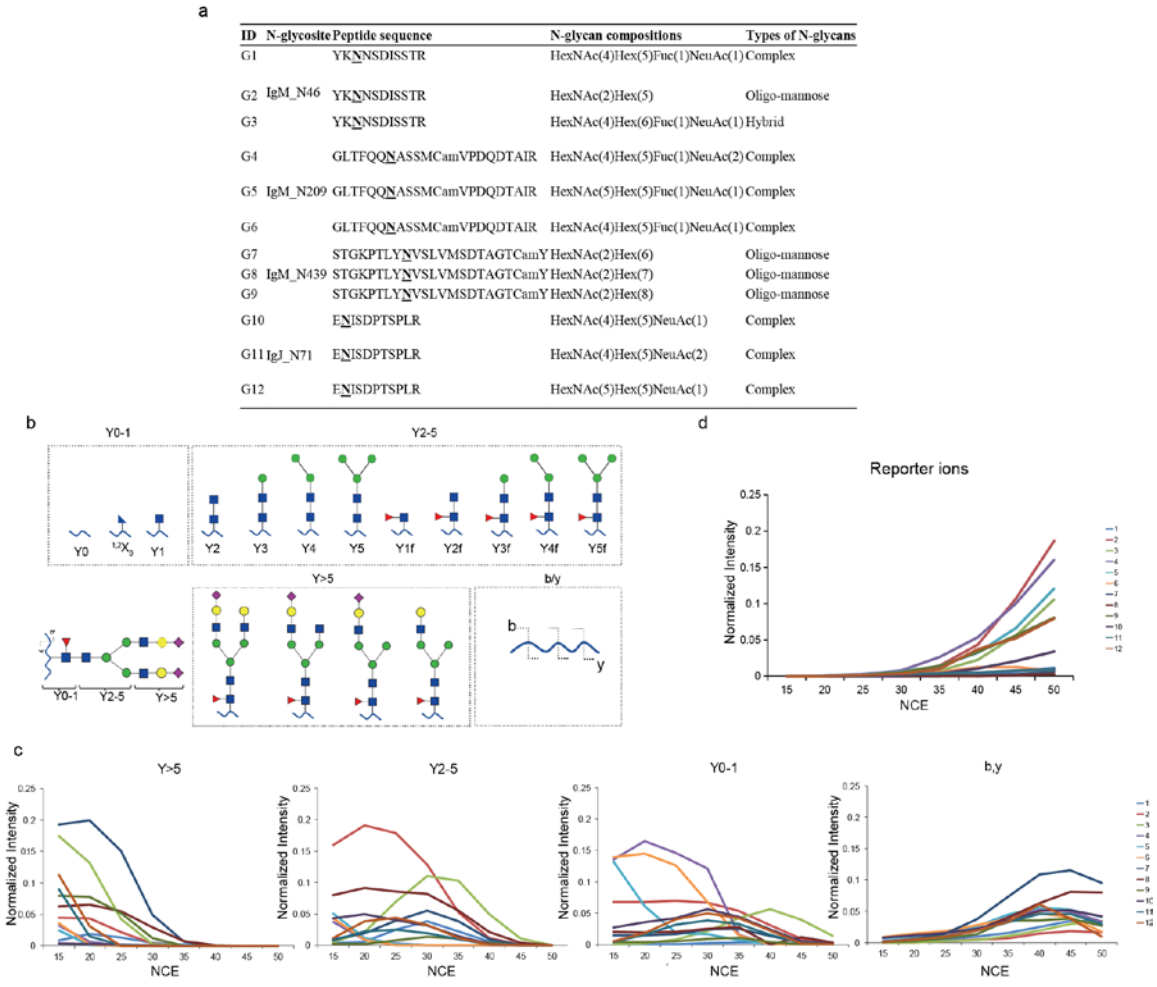


Figure 4.6 Fragmentation patterns of TMT labelled N-glycopeptides from IgM under various NCEs in HCD-MS2 analysis.

(a) Twelve TMT-labelled N-glycopeptides from IgM were selected to monitor their fragmentation patterns under various NCEs (15%, 20%, 25%, 30%, 35%, 40%, 45%, and 50%). (b) All potential product ions of an N-glycopeptide were classified into four types based on fragmentation characteristics under HCD. The definition of Y>5, Y0-1, Y2-5, b/y was described in the main text. (c, d) The intensities of fragment ions (c) and reporter ions (d) of TMT-labelled glycopeptides detected under different NCEs in eight consecutive spectra were extracted and normalized to the total ion current of the respective spectra. The code-colors of twelve different glycopeptides were shown on the right.

Although HCD with stepped collision energy (sNCE) helps to improve the unlabelled intact glycopeptide identification [154], such a wide range of favorable NCEs for different fragment ion types made it challenging to obtain better fragments for confident identification of TMT labelled N-glycopeptides without compromising the generation of TMT reporter ions in conventional HCD-MS2 analysis. Furthermore, from the previously published datasets we see that either TMT- or iTRAQ-labelled glycopeptides generated nearly one-order less intense TMT/iTRAQ reporter ions than non-glycosylated peptides in HCD-MS2 analysis (**Figure 4.7b**), even with 45% NCE (**Figure 4.7a**). One possible reason was that the breakage of the more labile glycosidic bonds was prone to occur and took most of the collisional energy so that the generation of reporter ions was limited. The lower reporter intensity was adverse to quantification sensitivity and accuracy. Furthermore, multiplex TMT/iTRAQ-labelled quantification with standard MS2 methods often suffers from impaired accuracy due to the co-isolation interference. There is an urgent need to develop novel techniques specifically for multiplexed N-glycoproteome quantification with high sensitivity and high accuracy.

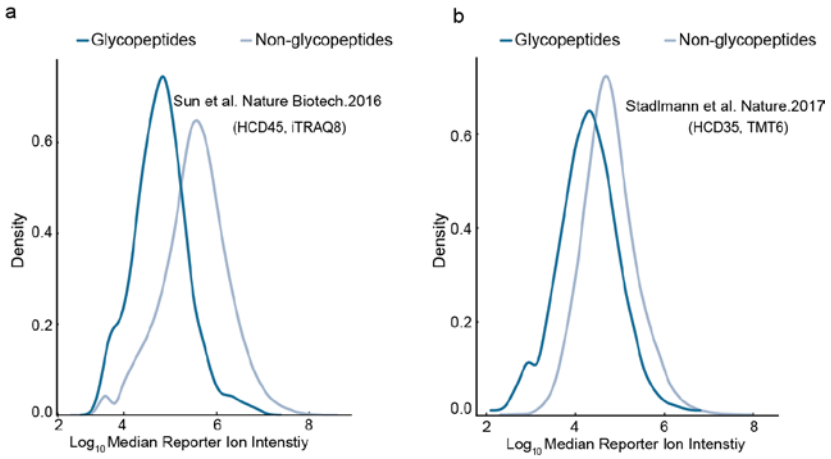


Figure 4.7 Distributions of detected reporter ion intensities from glycopeptides or non-glycopeptides.

The original publications for the used dataset, the used chemical labeling reagents and applied NCEs were noted in the figure [151, 153]. The median of all reporter ions intensities on each MS2 spectrum was log10-transformed for use. Reporter ions from glycopeptides and non-glycopeptides were labelled in dark and light blue, respectively.

Based on the published MS3 strategies for intact N-glycopeptide identification [152, 163] and the SPS-MS3 method for more accurate quantitation of proteome data, we proposed a new MS acquisition workflow for TMT-labelled N-glycopeptides analysis, namely Glyco-SPS-MS3 (**Figure 4.8a**). It applied HCD with low NCE (around 25%) on N-glycopeptide precursors to produce intense glycan-product Y ions bearing the entire peptide sequences. The top ten abundant fragment ions (mainly Y ions) were then co-selected (700-2000 m/z) and further co-fragmented into peptide b-/y-ions as well as TMT reporter ions by applying higher NCE (35-40%) HCD in MS3. Unlike the original SPS-MS3 that utilizes a fast-scanning ion trap and parallel CID fragmentation for high-speed MS2 peptide sequencing and co-selection, co-fragmentation of multiple MS2 fragment ions with high NCE (HCD65) in MS3 to produce abundant reporter ions, our Glyco-SPS-MS3 used HCD and Orbitrap detection in both MS2 and MS3. Our results showed that HCD fragmentation followed by high-resolution and high-accuracy Orbitrap detection provided more high-scoring N-glycopeptide identifications from IgM digests in standard MS2 analysis (**Figure 4.8b,c**). It also allowed us to combine all the fragment ions on MS2 and MS3 spectrum derived from the same N-glycopeptide precursor via a post-acquisition process (refers to the GlycoBinder part) in order to obtain confident and reliable intact N-glycopeptide identification. Synchronous selection of multiple MS2 fragments for MS3 analysis further boosted the overall sensitivity and reduced co-isolation interference for quantitation.

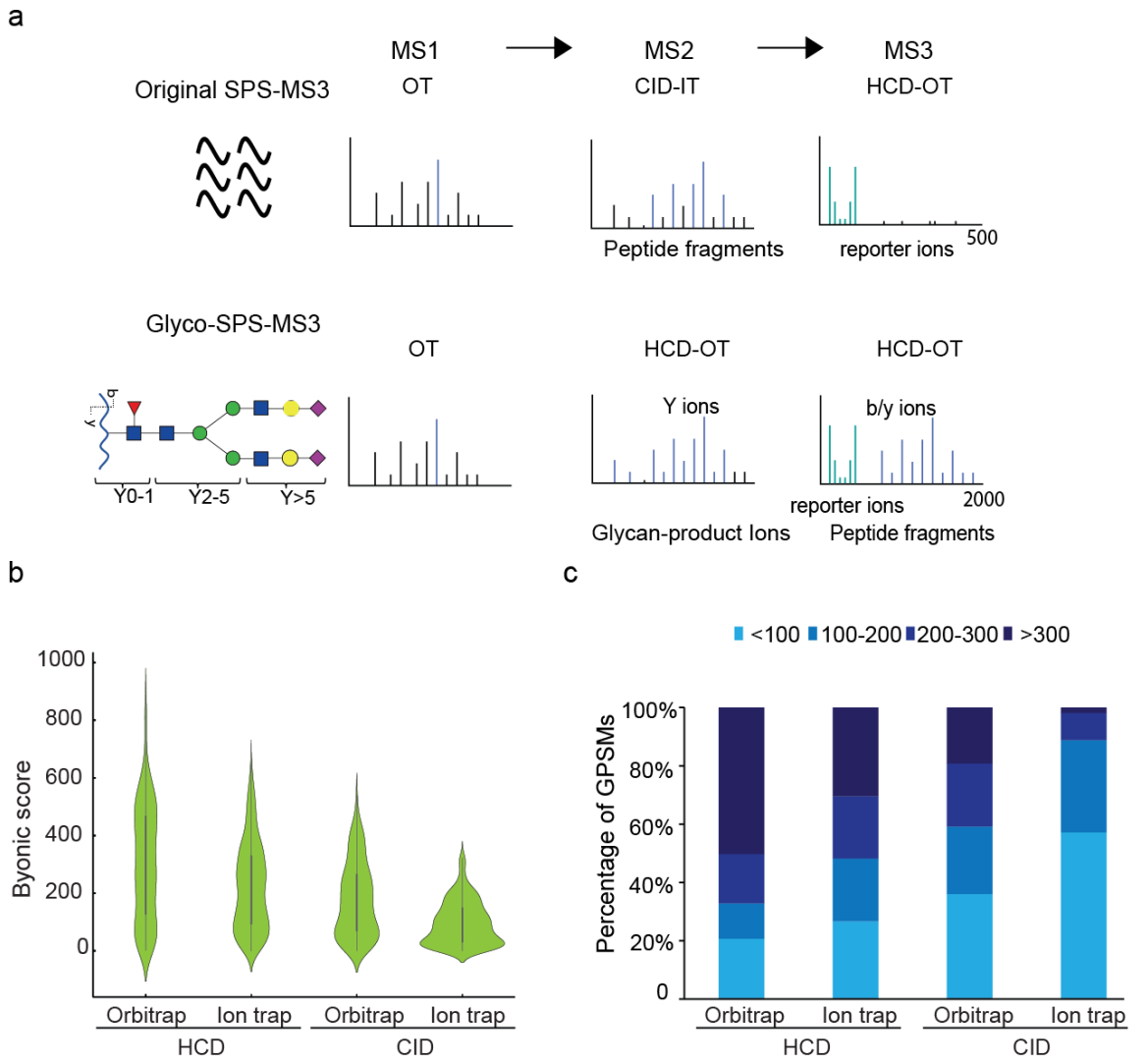


Figure 4.8 Schematic diagram of the Glyco-SPS-MS3 workflow and the selection of MS detectors and fragmentation modes for Glyco-SPS-MS3.

(a) The original SPS-MS3 and the Glyco-SPS-MS3 workflow. The original SPS-MS3 used CID and ion trap in MS2 for peptide sequencing, and then co-selected and co-fragmented the top N most abundant fragment ions in MS2 (in blue) using HCD followed by Orbitrap detection for MS3, which sets a narrow scan range of 100-500 m/z for record of reporter ions. Glyco-SPS-MS3 used low NCE in MS2 and high NCE in MS3 under HCD fragmentation and by Orbitrap detection. (b) The distribution of Byonic scores of identified IgM glycopeptides obtained from four different combinations (Orbitrap_HCD, Ion trap_HCD, Orbitrap_CID and Ion trap_CID) in MS2 analysis. (c) The distribution of identified Glycopeptide-spectrum matches (GPSMs) from the above mentioned four combinations based on their Byonic scores is shown in a stacked bar chart. Byonic score above 300 was suggested as the threshold for confident identification [169].

In a first attempt, we surveyed various combinations of NCE settings in the Glyco-SPS-MS3 method and either single or stepped NCE (sNCE) in standard MS2 methods for the comparison of resulting identified GPSMs and glycoforms, scores of GPSMs including total, peptide and glycan scores from pGlyco 2.0 output and the signal-to-noise (S/N) ratio of the TMT reporter ions (**Figure 4.9**). Our results show that the overall higher scored GPSMs in the database search were obtained for Glyco-SPS-MS3 over the MS2 methods (**Figure 4.9 a,b,c**), which was especially reflected in the glycan score (**Figure 4.9c**). However, regarding the absolute number, more GPSMs were obtained from the database search in MS2 with sNCE than compared to the Glyco-SPS-MS3 method, while the number of unique

glycoforms was similar, suggesting that many of the GPSMs in MS2 are low scored and Glyco-SPS-MS3 is necessary for more reliable identification (**Figure 4.9d**). Importantly, Glyco-SPS-MS3 also improved the overall median S/N ratios of reporter ions by 2-3 folds as compared to MS2 methods (**Figure 4.9e**). The TMT reporter ions S/N continued to increase their intensity along with higher NCEs used in MS3 (40% vs 35% NCE in MS3, **Figure 4.9e**), however, an NCE of 65% resulted in the identification of GPSMs with low score (**Figure 4.9g**). Among other NCE settings (**Figure 4.9f,g**), we concluded that the combination of 25%-30% NCEs for MS2 and 35%-40% for MS3 is the best choice due to the higher identification numbers and GPSMs scores. Notably, the application of sNCE on either MS2 or MS3 within Glyco-SPS-MS3, such as the combination of an NCE of 25% in MS2 with NCEs of 25%/35%/45% in MS3, did not improve the identification of glycopeptides by the database search (**Figure 4.9d**).

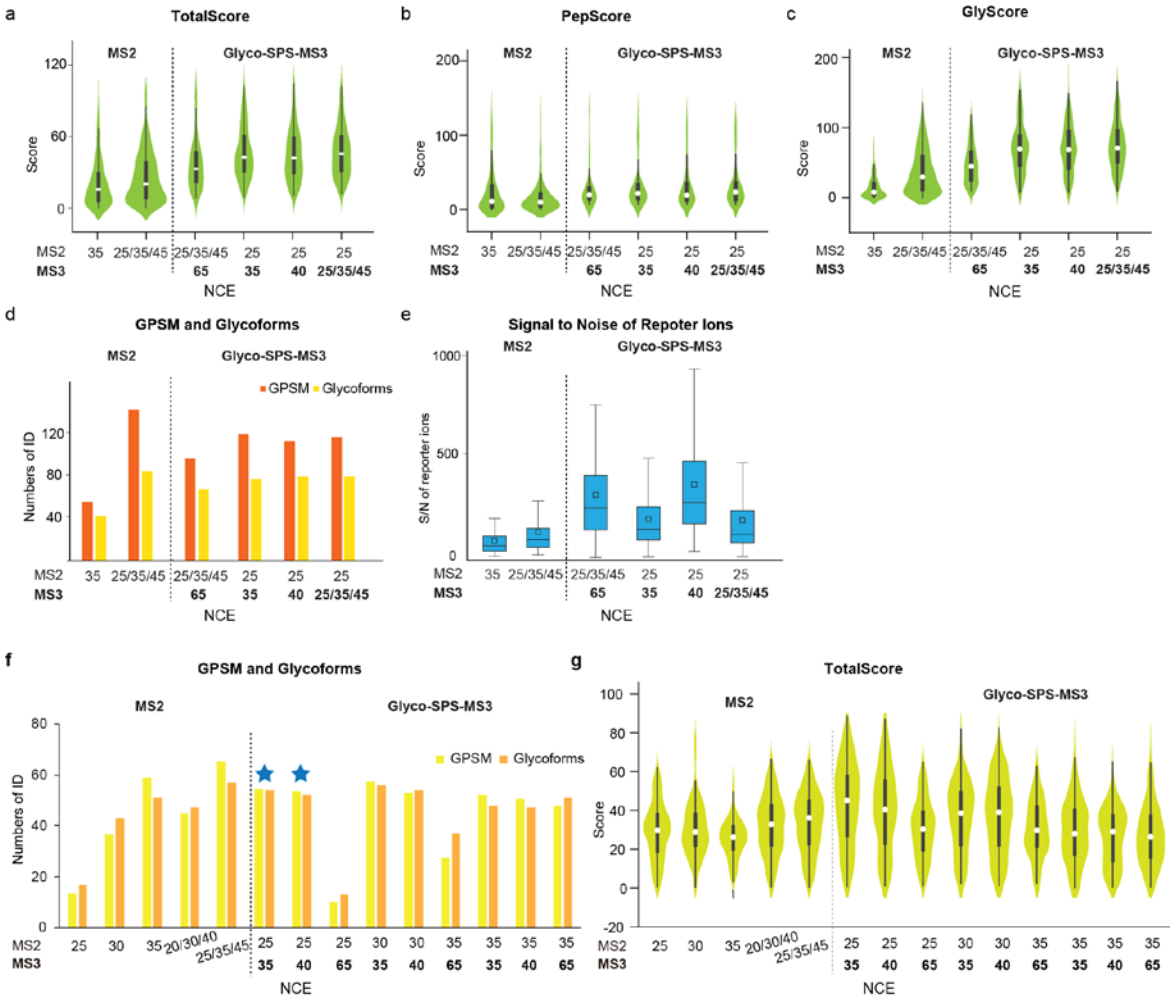


Figure 4.9 Evaluation of various single or stepped NCE (sNCE) settings in Glyco-SPS-MS3 versus standard MS2 methods.

The total score (a, g), peptide score (b) and glycan sore (c) of identified GPSMs in Glyco-SPS-MS3 and MS2 methods with various NCE settings were shown in violin plots; (d, f) Number of identified GPSMs and unique glycoforms from two replicates using different acquisition methods. (e) The distributions of S/N ratios of TMT-reporter ions using MS2 or Glyco-SPS-MS3 methods with various NCE settings were shown in box plot.

It is commonly established in the field that any MS3-based method has the drawback of a longer duty cycle. However, our results show that N-glycopeptide identification in a standard MS2 analysis was improved by increasing the automatic gain control (AGC) and prolonging the ion injection time (IT)

(**Figure 4.10a,b**). When using an injection time of 100 ms, 81% of MS2 scans reached the set AGC target value of $5e^4$, 70% reached e^5 and only 5% reached $5e^5$. Simultaneously the GPSMs that were identified by pGlyco 2.0 search when the AGC was set to $5e^5$ exceeded the number of GPSMs when the AGC was set to $5e^4$ and e^5 by almost three-fold. The results indicated that the amount of precursor ions is vital for the generation of enough fragment ions for confident identification. Furthermore, as the IT was prolonged to 250 ms and 500 ms, more MS2 scans reached the set's AGC target value of $5e^5$ and more GPSMs were identified. By applying an AGC of $5e^5$ and an IT of 500 ms, the most high-scoring GPSMs were obtained no matter pGlyco searching or Byonic searching was used. To maintain a comparable cycle time with standard MS2 methods, we allocated the total 500 ms of IT in different proportions to the MS2 and MS3 scans in the Glyco-SPS-MS3 method (**Figure 4.10c**). The results showed no significant differences in the calculated ratios for the identified GPSMs and glycoforms for different IT distribution to MS2 and MS3 (**Figure 4.10c**). Interestingly, a longer IT in MS3, or a shorter IT in MS2, led to an increased peptide score along with a slightly decreased glycan score (**Figure 4.10 d, e, f**). This observation is consistent with our suggestion that low-NCE MS2 provides more glycan Y ions, while high-NCE MS3 generates more peptide b/y ions. The evidence of 30% fewer triggered precursor ions using the Glyco-SPS-MS3 method compared to the MS2 methods showed a significant decrease in the acquisition speed (**Figure 4.11**). Worthwhile, Glyco-SPS-MS3 showed overall 2.5-3.6 and 1.1-1.3 fold GPSMs identification rates respectively compared to the MS2 methods with single energy and MS2 methods with sNCE using the same cycle time, probably owing to improved spectrum qualities through the combination of MS2 and MS3 scans for database searching. Furthermore, as the numbers of notches (the number of fragment ions in MS2 spectrum selected for MS3 analysis) increased, higher numbers of GPSMs and glycoforms were identified mainly based on the increased peptide ion ratio of identified peptide fragment ions to all theoretical peptide fragment ions (**Figure 4.12 a,b,c**). In summary, we have established the Glyco-SPS-MS method for sequencing of glycosylated peptides by tweaking/ systematically adjusting several MS settings, including NCE, AGC, notches and IT allocation in MS2 and MS3 (**Table 4.1**).

Table 4.1 Optimization of MS parameters in the Glyco-SPS-MS3 method.

Parameters	**Settings**	**Optimal**
Detector in MS2	Orbitrap, Ion trap	Orbitrap
Fragmentation in MS2	HCD, CID	HCD
HCD NCE in MS2	Single energy: 25, 30, 35 sNCE : 20\30\40, 25\35\45	sNCE-25\35\45
HCD NCE in MS2 and MS3	1. MS2 (single energy)_MS3 (single energy): 25_35, 25_40, 25_65, 30_35, 30_40, 30_65, 35_35, 35_40, 35_65; 2. MS2 (sNCE)_MS3 (single energy): 25\35\45_65; 3. MS2 (single energy)_MS3 (sNCE): 25_25\35\45;	25_35; 25_40;
Injection time in MS2	100 ms; 250 ms; 500 ms	500 ms
Injection time allocation in MS2 and MS3 (ms)	MS2_MS3: 150_350; 200_300; 250_250; 300_200; 350_150	150_350
AGC	$5e^4$; e^5; $5e^5$	$5e^5$
Notches	3; 5; 10	10

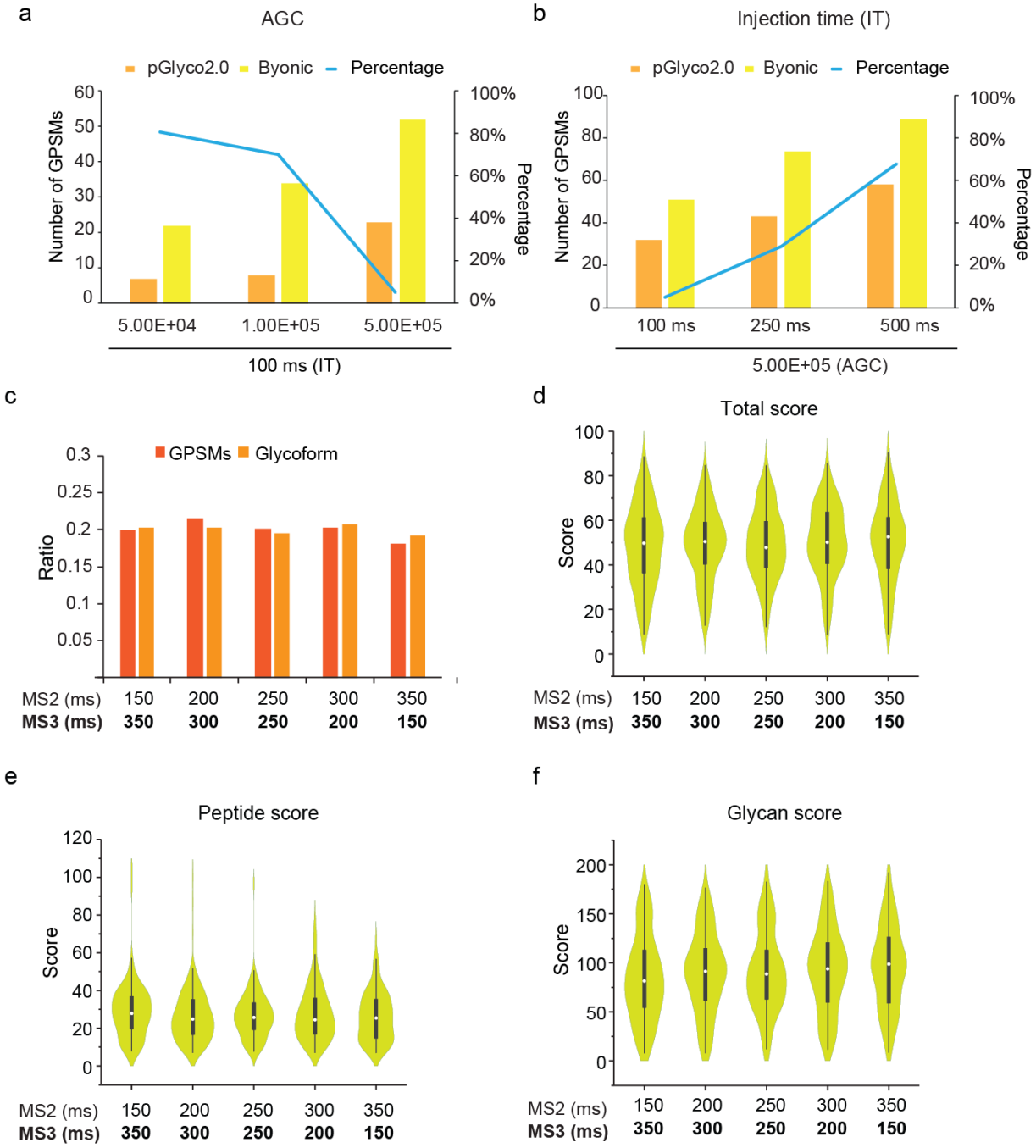


Figure 4.10. Optimization of AGC target values and IT.

(a,b) Numbers of GPSMs identified by either pGlyco2.0 or Byonic using different settings for the AGC target values and the IT in standard MS2 analysis. The percentages of MS/MS scans reaching pre-defined AGC are showed by broken line. All identification required stringent criteria: pGlyco 2.0: PepScore >7 and GlycScore>8; and Byonic: score >300. (c) The ratios of numbers of identified GPSMs and unique glycoforms for individual IT allocations to MS2 and MS3 in Glyco-SPS-MS3 divided by the total identification numbers among all the settings' results are shown. The allocation of ITs to MS2 and MS3 are marked on the x axis and AGC $5e^5$ was used for both MS2 and MS3. Distributions of total score (d), peptide score (e), and glycan score (f) from each IT allocation setting are shown separately.

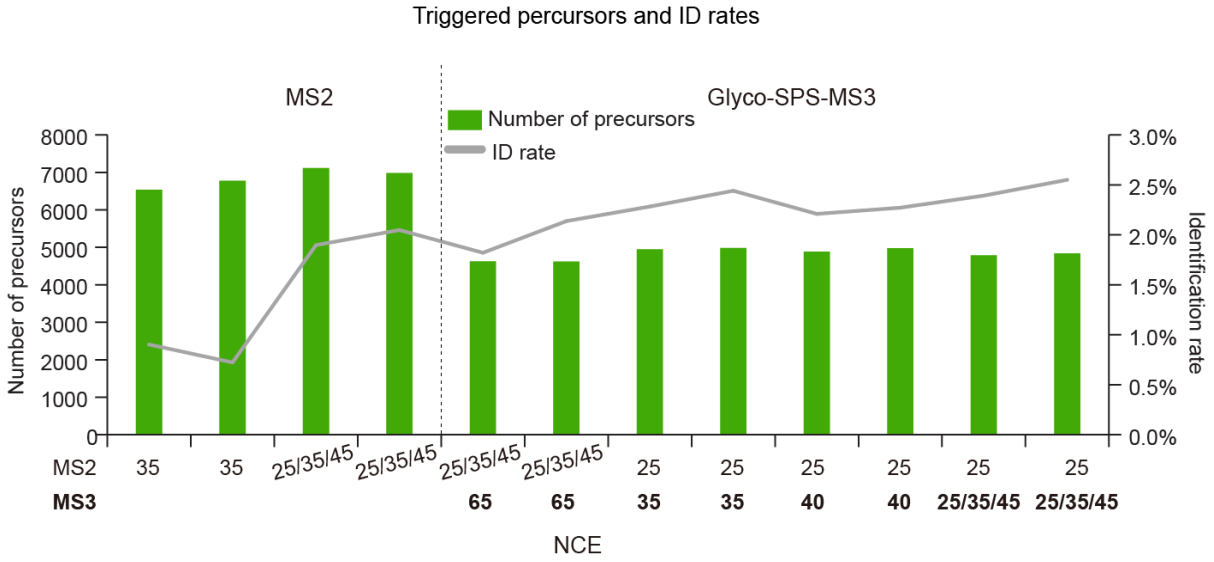


Figure 4.11 Number of total triggered precursor ions (MS2 scans) and the GPSMs identification rates in standard MS2 and the Glyco-SPS-MS3 method with different NCEs settings.

Analysis of each NCE setting includes two replicates.

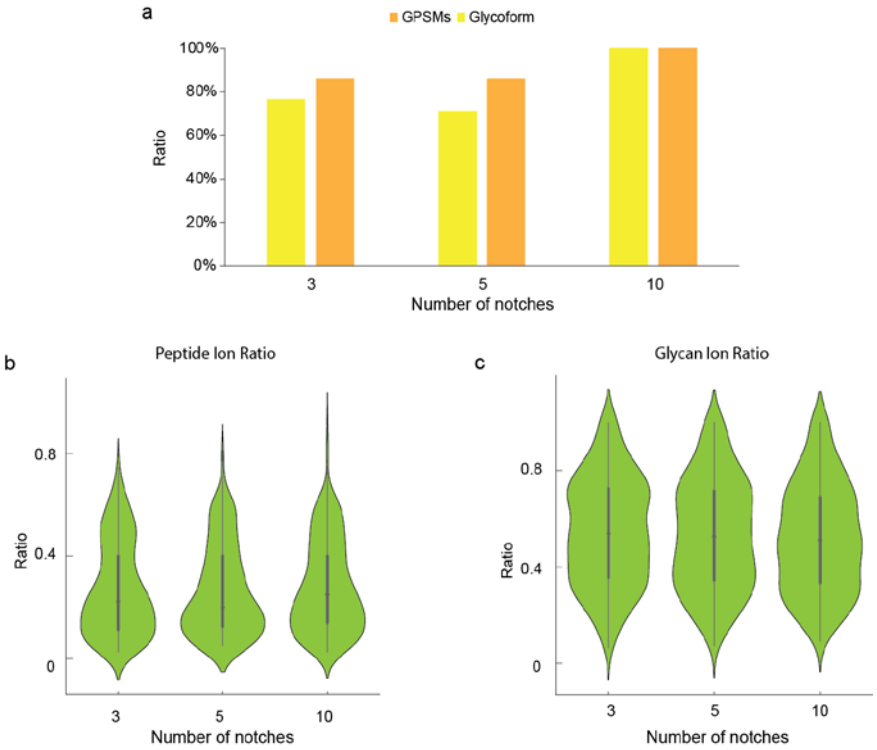


Table 4.12 Optimization of the number of notches selected for MS3 analysis.

(a) Identification of GPSMs and unique glycoforms using different notches for MS3. (b, c) Violin plots showing the distribution of peptide ion ratio, and glycan ion ratio in different notches for MS3. Ratio represents the percentage of the number of individual fragment ions (peptide or glycan) in all fragment ions in one spectra.

Subsequently, we designed an IgM-yeast mixture model to evaluate the quantitation accuracy of our Glyco-SPS-MS3 method. We labelled IgM digests separately using TMT6-plex reagents for six channels and pooled them together according to a ratio of 10:4:1:1:4:10. Moreover, digested peptides from yeast were labelled with only the first three channels of TMT6 reagents (126, 127 and 128) and mixed equally. Then we spiked the mixed IgM peptides into the yeast peptides mixture at an equal

amount (**Figure 4.13a**). Glyco-SPS-MS3 and MS2 methods were used to analyze the quantitation of IgM-yeast mixture. Ideally, IgM glycopeptides should be quantified with the ratio of 10:4:1:1:4:10 in the mixture model. Taking in account of the occurrence of co-isolation, the co-selected yeast peptides contributed extra amounts of reporter ions that distorted the ratios of the interfered channels (i.e., 126, 127, and 128) while leaving the remaining ones (129, 130, and 131) unchanged. Indeed, we accurately determined the pre-defined 2.5-, 4-, and 10-fold changes in the not-interfered channels with both MS2 and Glyco-SPS-MS3 methods (**Figure 4.13b**). Among the interfered channels, skewed median fold changes of 2.2, 1.9, and 3.9 were detected in the MS2 method, while reporter ion ratios (ie, 2.9, 3.8 and 11.2) were recorded more accurately in the Glyco-SPS-MS3 method(**Figure 4.13b**).

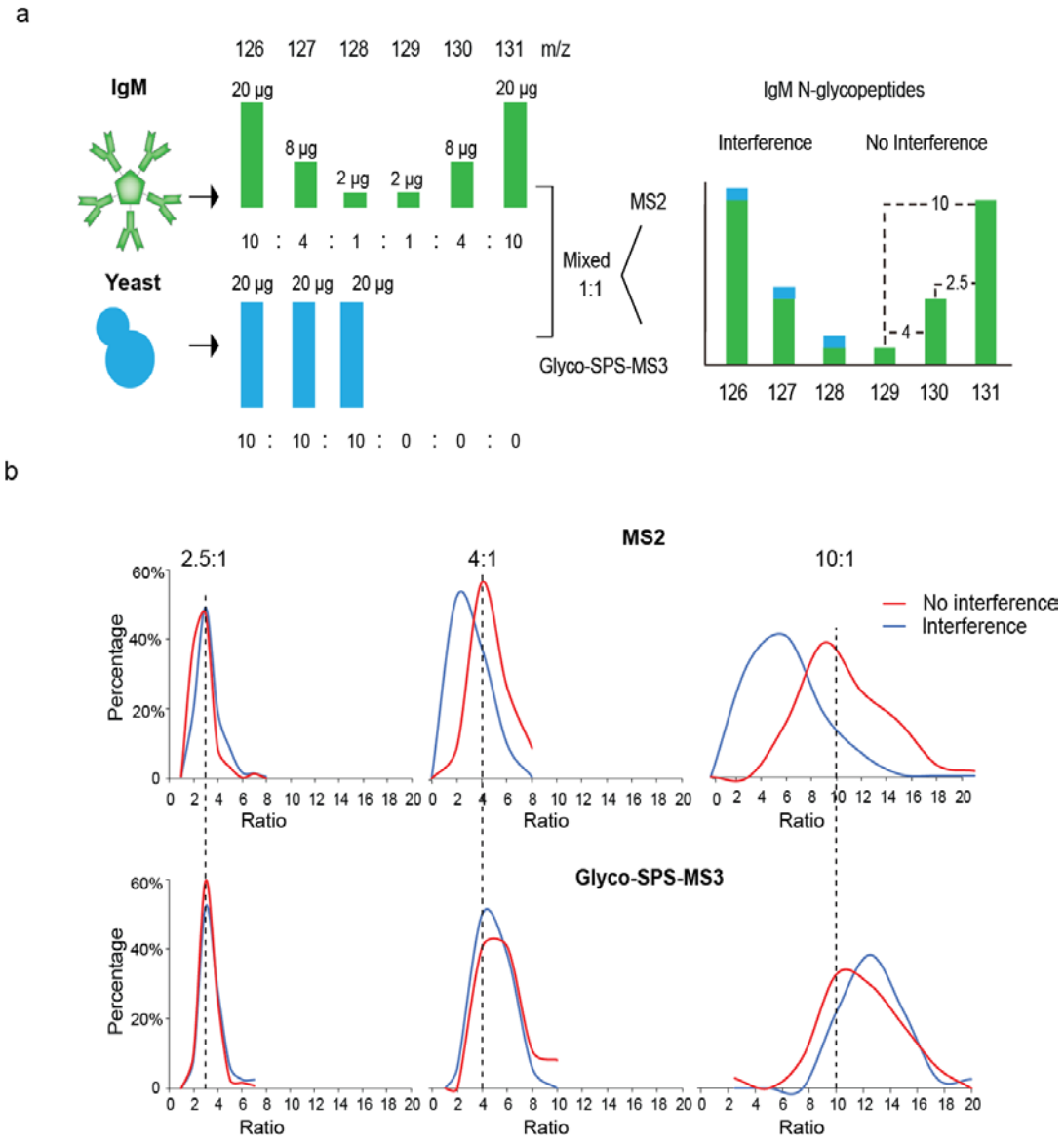


Figure 4.13 Evaluation of quantitation accuracy of the Glyco-SPS-MS3 and MS2 method using IgM-yeast mixture model.

(a) Design of the IgM-yeast mixture model. (b) Distribution of reporter ion ratios of IgM glycopeptides from the IgM-yeast mixture using the MS2 or the Glyco-SPS-MS3 method. Channels with and without yeast interference are shown in blue and red, respectively.

4.1.4 GlycoBinder for one step quantitative N-glycoproteomics data processing

To make full use of the unique advantages of Glyco-SPS-MS3, we further developed a workflow for processing of the quantitative SPS-MS3 data and implemented it in a R-based script, namely GlycoBinder (available on GitHub (https://github.com/IvanSilbern/GlycoBinder)). Although there is increasing number of open software tools proposing specialized scoring systems to achieve reliable intact N-glycopeptide identification [147], for example, Byonic and SugarQB which achieve the multiplex intact N-glycopeptide quantitation based on the MS2 analysis, however, none of them supports processing of MS3 data and multiplex quantitation. Here, we decided to use the free pGlyco 2.0 search engine, which can be executed in the command line and allows for precise and confident intact N-glycopeptides identification with strict quality FDR control [154]. Thus we developed GlycoBinder, an R-based tool that integrates several published computational tools to automatically extract and combine MS2 and MS3 fragment ions into a pseudo-spectrum, and conduct intact N-glycopeptide identification by database search using the pGlyco 2 algorithm (**Figure 4.14a**). Specifically, GlycoBinder uses MSConvert [216] to convert MS2 and MS3 spectra into mgf format, and further uses RawTools [218] to report a table listing the respective scan numbers of all MS3 scans and their parent MS2 scans. GlycoBinder then merges all MS2 and MS3 fragment ions accordingly using a pre-defined mass tolerance (1 ppm by default). GlycoBinder executes pParse [217] to re-assign the monoisotopic peak of each precursor. Subsequently, GlycoBinder searches the merged pseudo-spectra with corrected precursor m/z-values for glycopeptide identification using pGlyco 2 algorithm. Finally, GlycoBinder utilizes RawTools to extract TMT reporter ions intensities from MS raw files and then supplement each GPSM with corresponding quantitative values. **Figure 4.14b** showed the process of propagation of quantitative information from GPSMs to unique glycoforms, glycosites and glycans. For one N-glycosylation site (N) on a protein (P), multiple glycans (G) may exist on that site, resulting in multiple glycoforms (represented as P-N-G). Quantification of unique glycoforms is achieved by summing TMT-reporter ion intensities of all involved GPSMs resulting from miscleavages, modifications, and different charge states. Quantified glycoforms are further combined for glycosite (P-N) quantification. As for a unique glycan, the quantification values of all glycoforms with this particular glycan in the sample are combined.

We evaluated the performance of GlycoBinder for the analysis of TMT-labelled N-linked glycopeptides in a human database search derived from DG75 cells, which were prepared and measured with the pipeline SugarQuant as outlined above. Three strategies including merging spectrum of MS2 and MS3 scans by Spectrum Grouper node in PD, merging the spectra from MS2/MS3 scans extracted by MSConvert without precursor mass redetermination by pParse (GlycoBinder without pParse) and with precursor mass correction by pParse (GlycoBinder with pParse) were compared (**Figure 4.14c**). Glycobinder with pParse identified 6% and 15% more GPSMs when compared to GlycoBinder without pParse and the PD merging strategy, respectively. This improvement can be attributed to the precursor mass correction by pParse. While MSConvert and PD directly output the original precursor mass as reported on raw files, pParse re-determines the monoisotopic peaks of precursors and de-convolute co-isolated precursors. It was demonstrated previously that accurate precursor mass is vital for accurate N-glycopeptide assignment [153, 154]. Subsequently, GlycoBinder concatenates each GPSM with corresponding reporter ion intensities resulting from the same spectra. We also confirmed that the reporter ions intensities extracted by RawTools were consistent with the values reported by PD (**Figure 4.14d**).

Next we investigated whether the GlycoBinder search will be affected by the size and specificity of the protein database. In previous publications, researchers often used sample-specific glycan and protein

databases from the corresponding glycomic and deglycoproteomics analysis of the same sample for intact glycopeptide analysis in order to decrease the false positive identification rate [153]. The complexity of experimental manipulation increased to a large extent and the limited depth of deglycoproteomics decreased the representativeness of constructed protein databases for intact glycopeptide analysis. In our study, four different protein databases were used (refers to section 3.2.5.4). It included: (1) reviewed human proteome database from Swiss-Prot ("Human-reviewed", 20303 entries, downloaded in March 2018), (2) a curated glycoprotein database from Swiss-Prot (Glycoprotein-reviewed, 4824 entries, downloaded in March 2018), (3) an in-house built "B-cell-Specific glycoprotein" database with 974 glycoproteins that were identified in our previous studies in DG75 and Daudi cell lines, (4) a non-related database consisting of a random selection of 1000 proteins (Random-1000) from the Human-reviewed database. All the used protein database can be found in **Supplementary Data S1** (**Figure 4.14e**). When analyzing TMT-labelled N-glycopeptides enriched from DG75 cells, similar glycoform identifications (4381 vs 4335) at 2% FDR with 80% overlap (**Figure 4.14f**) were obtained by a GlycoBinder search using the Human-reviewed and the Glycoprotein-reviewed databases separately, although the size of the Glycoprotein-reviewed database was only one fourth of the Human-reviewed database. A GlycoBinder search using the B-cell-specific glycoprotein database resulted in 10% higher numbers of (4782 vs 4335) of glycoform identifications than using the Glycoprotein-reviewed database despite the fact that the B-cell-specific database contains only 974 proteins, which corresponds to a 95% decrease in database size. In contrast, the search against the Random-1000 database resulted in the identification of only 243 unique glycoforms, suggesting that using a smaller database does not spontaneously cause overfitting of the database search. The specificity of the database rather than its size is more critical for the improvement of identification [221]. In order to avoid the use of specific experimentally constructed databases from deglycoproteomics, we conducted a faster and more convenient two-step database search. Briefly, we search the raw files against the Human-reviewed database firstly. A second search for the same raw data against the identified glycoproteins from the first search was performed then. The two-step search achieved the identification of 5367 glycoforms and 855 glycosites, representing a separate 22.5% and 8.5% increase as compared to the results from the single search using the Human-reviewed database. More than 90% of the identified glycosites were overlapping in the first single and two-step search (**Figure 4.14g**). Moreover, the two-step search helped to rescue many low-scored glycoforms at a fixed 2% FDR which couldn't be identified in the first search (**Figure 4.14h**).

In summary, GlycoBinder not only supports the standard database search with any user-defined FASTA protein sequences but also allows an automated two-step search for glycopeptide identification with considerably higher sensitivity. In addition, GlycoBinder uses the annotated results from pGlyco 2 software and propagates the corresponding GPSMs to different levels of quantification, including unique glycosylation sites, unique glycoforms, and unique glycan compositions. The propagation of quantitative values of GPSMs is performed in the same way as in quantitative proteomics, where peptide quantifications are propagated to the protein level [222]. These tables could be used directly for the subsequent data analysis. Importantly, GlycoBinder is performed with one single command execution to achieve the streamlined data processing of multiplexed intact N-glycopeptide quantitative mass spectrometry data. To maximize the utility's value, GlycoBinder also supports raw files acquired with common MS2 methods for those who have no access to the MS3-capable instruments. GlycoBinder is available on GitHub (https://github.com/IvanSilbern/GlycoBinder).

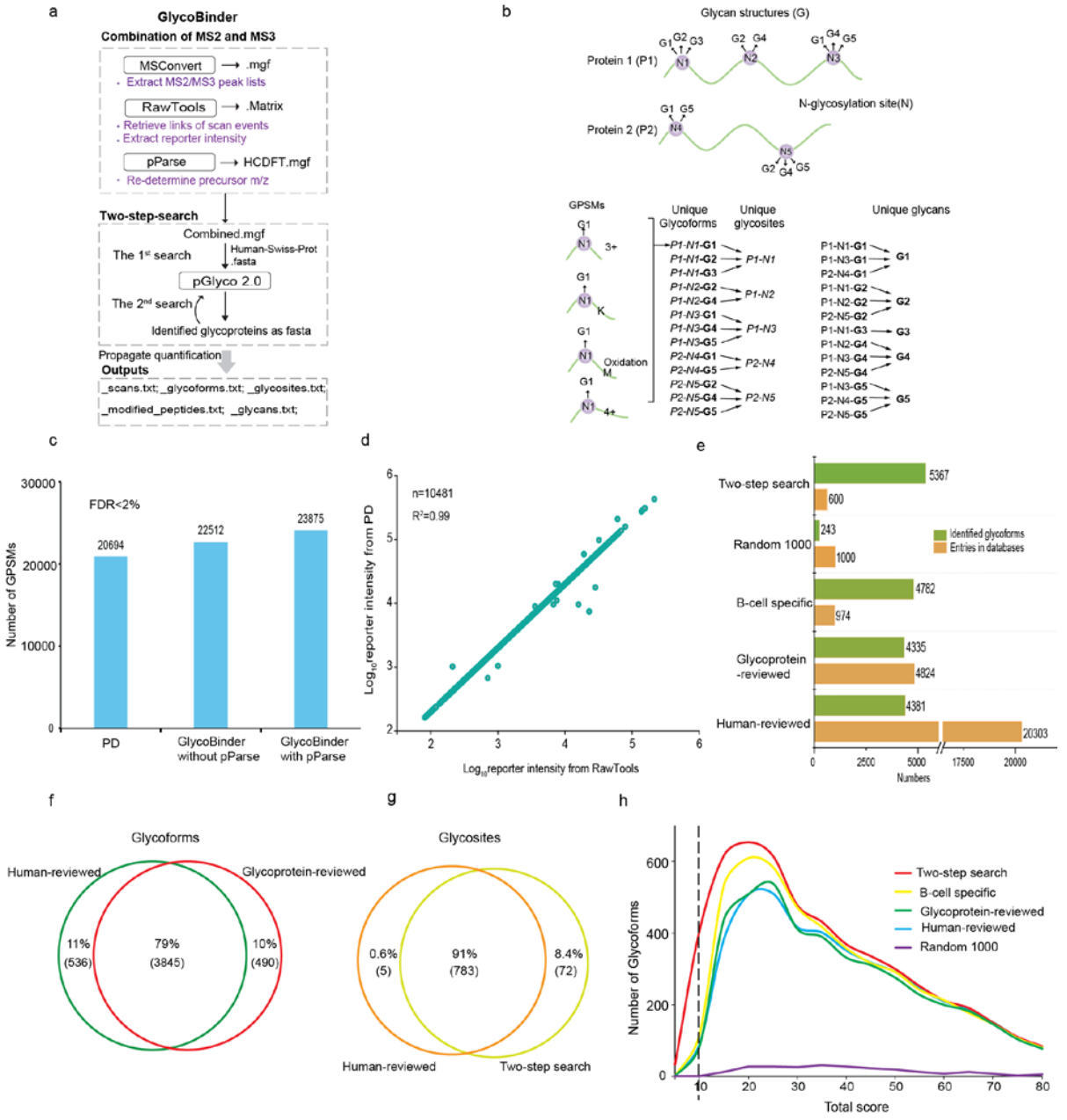


Figure 4.14 GlycoBinder enables streamlined MS3 data processing for intact N-glycopeptide identification and accurate multiplex quantitation.

(a) The schematic workflow of GlycoBinder. (b) Schematic explanation of the propagation of quantitative information from GPSMs to unique glycoforms, glycosites and glycans in this study. N-glycosylation site (N), protein (P), glycans (G), unique glycoforms (P-N-G), unique glycosite (P-N). (c) Investigation of spectra-merging workflows including PD, GlycoBinder without pParse and GlycoBinder with pParse by comparing numbers of identified GPSMs. All combined mgf files were searched using pGlyco 2.0. (d) Correlation of reporter ion intensities extracted by RawTools and PD. Each circle represents the extracted reporter ion intensity from one scan. A total of 10481 reporter ions were used. (e) Comparison of the two-step search with single searches using various databases. Sizes of the databases (brown) and the resulting numbers of identified glycoforms (light green) are shown. The overlap of identified glycoforms (f) and glycosites (g) obtained from GlycoBinder searches using different protein databases in Venn diagrams. (h) Aligned score distributions of the identified glycoforms by GlycoBinder searches using different protein databases.

4.1.5 Multiplex quantitation of N-glycoproteome and phosphoproteome analysis on fucosylation-inhibited human Burkitt's lymphoma cells

Fucose is a natural deoxyhexose sugar which is lack of a hydroxyl group on carbon 6 compared to glucose. It reported that inhibition of fucosylation using fluorinated fucose analog 2-deoxy-2-fluoro-L-fucose (2FF) suppressed cell proliferation and migration in human liver cancer cells [192]. However, it is still unclear whether 2FF affects any protein- or site-specific glycosylation regulations in cancer cells, although the global expression of fucosylated glycan is indeed inhibited by 2FF [191]. In order to gain more understanding about the mechanisms of 2FF-treatment and 2FF-sensitive glycosylation sites, we applied our developed SugarQuant platform to analyze the quantitative N-glycoproteome in 2FF-treated human Burkitt's lymphoma cells (DG75). DG75 cells were treated with a series of concentrations (60-600 µM) of 2FF for three days. We performed the PAC based-sample preparation for the treated cells which includes protein extraction in 4% SDS, detergents clean-up by PAC and digestion, TMT labeling, glycopeptides enrichment by ZIC-HILIC and bRP. The resulted labelled N-glycopeptides were analyzed using the Glyco-SPS-MS3 (each fraction for three injection replicates) or MS2 method (each fraction for one injection replicate) on an Orbitrap Tribid mass spectrometer. The obtained MS raw files were processed using GlycoBinder. In addition, quantitative proteomics and deep phosphoproteomics analyses were performed separately using one aliquot of TMT-labelled peptides before glycopeptide enrichment and the flow-through (unbound parts to HILIC beads) after enrichment (**Figure 4.15a**). Before implementing the MS analysis, the decreased expression of global fucosylation was confirmed by lectin blotting against a biotinylated fucose-specific lectin (*Aleuria Aurantia* lectin, AAL, **Figure 4.15b**)

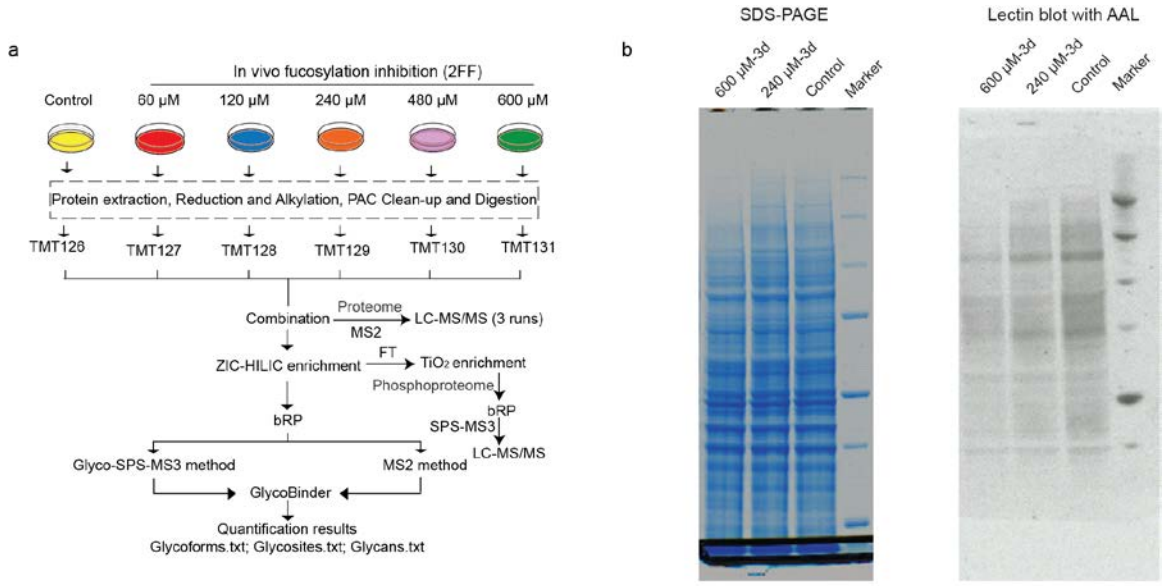


Figure 4.15 Quantitative N-glycoproteome and phosphoproteome analysis in 2FF-treated DG75 cells.

(a) The workflow of N-glycoproteome analysis used SugarQuant and global phosphoproteome analysis for 2FF-treated DG75 cells. (b) Validation of fucosylation expression in 2FF-treated DG75 cells by lectin blotting. DG75 cells were treated with 2FF at the final concentration of 240 µM or 600 µM for 3 days or with DMSO as control. Equal amounts of proteins extracted from 3 treatments were loaded on SDS-PAGE with 2 replicates. One was stained with coomassie brilliant blue (left) and the other one was blotted against AAL lectin (right).

4.1.5.1 Quantitative N-glycoproteome analysis using SugarQuant in 2FF-treated DG75 cells

In total, we identified 5367 unique glycoforms corresponding to 414 glycan structures on 855 glycosites from 528 glycoproteins, which is the first quantitative N-glycoproteomics analysis in human B cell

lymphoma (The list of identified glycoforms, glycosites and glycans can be found in **Supplementary Data S2**). Our results showed a great microheterogenetiy of the site-specific glycosylation in DG75 cells, from the evidence that more than 27% of identified glycosites bore more than five different glycan chains with an average of 6.27 glycoforms on each site (**Figure 4.16a**). Especially, more than 100 different glycoforms were identified on four glycosites, for example, 140 glycoforms on the 424th position of asparagine (Asn) on Slc3a2, 115 glycoforms on the 46th position of Asn on IGHM, 105 glycoforms on Asn 249 of Lamp1, and 104 glycoforms on Asn 380 of PTPRC. Among 528 identified glycoproteins, about 66.5% of them were identified to carry one glycosite and only 13% carry more than two glycosites (**Figure 4.16b**). The most glycan chains on glycoforms included 8-14 monosaccharides (**Figure 4.16c)** with a mass around 2 kDa (**Figure 4.16d**) and the average number of monosaccharides on identified glycan chains of the total glycoforms was about 11.

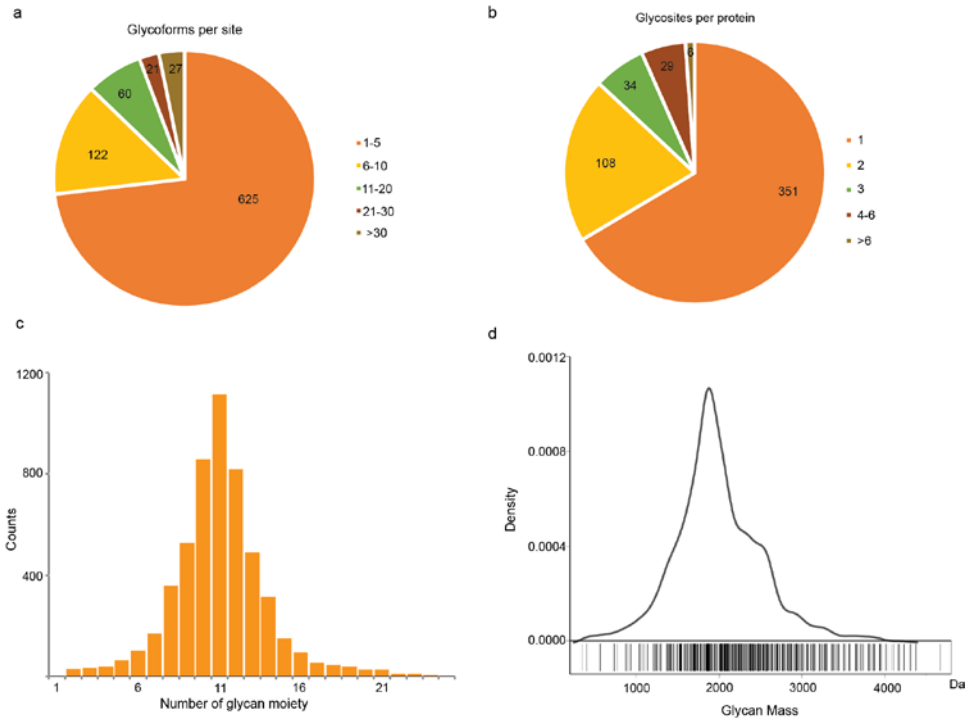


Figure 4.16 Microheterogenetiy of site-specific N-glycosylation in DG75 cells.

(a) The distribution of glycoforms identified on each glycosite. (b) The distribution of glycosites identified on each glycoprotein. (c) The distribution of the number of monosaccharides on glycan moieties of all identified glycoforms. (d) The distribution of glycan mass.

We performed a direct comparison of Glyco-SPS-MS3 and the optimized MS2 method for identifying TMT-labelled N-glycopeptides using the same LC settings and equal sample amounts (detailed parameters for both methods can be found in **Supplementary Table S1** in the appendices). The results show that the Glyco-SPS-MS3 method identified slightly fewer glycoforms (-6.4%), glycosites (-2.6%) and glycans (-2%) than the MS2 method (**Figure 4.17a**). In conventional proteome or phosphoproteome analysis using SPS-MS3 method instead of standard MS2 method, 13%-30% of identification rates decreased [126, 184]. However, SugarQuant achieved a comparable numbers of identification in N-glycoproteomics analysis as compared even with the optimized MS2 method for TMT-labelled N-glycopeptides analysis regardless of the slower acquisition cycle time. Presumably our fine-tuned parameters in the Glyco-SPS-MS3 method and efficient strategy for the combination of MS2 and MS3 information in GlycoBinder improved the N-glycopeptide analysis substantially. In addition, SugarQuant showed higher reporter ion intensities in N-glycoproteomics analysis (**Figure 4.17b**).

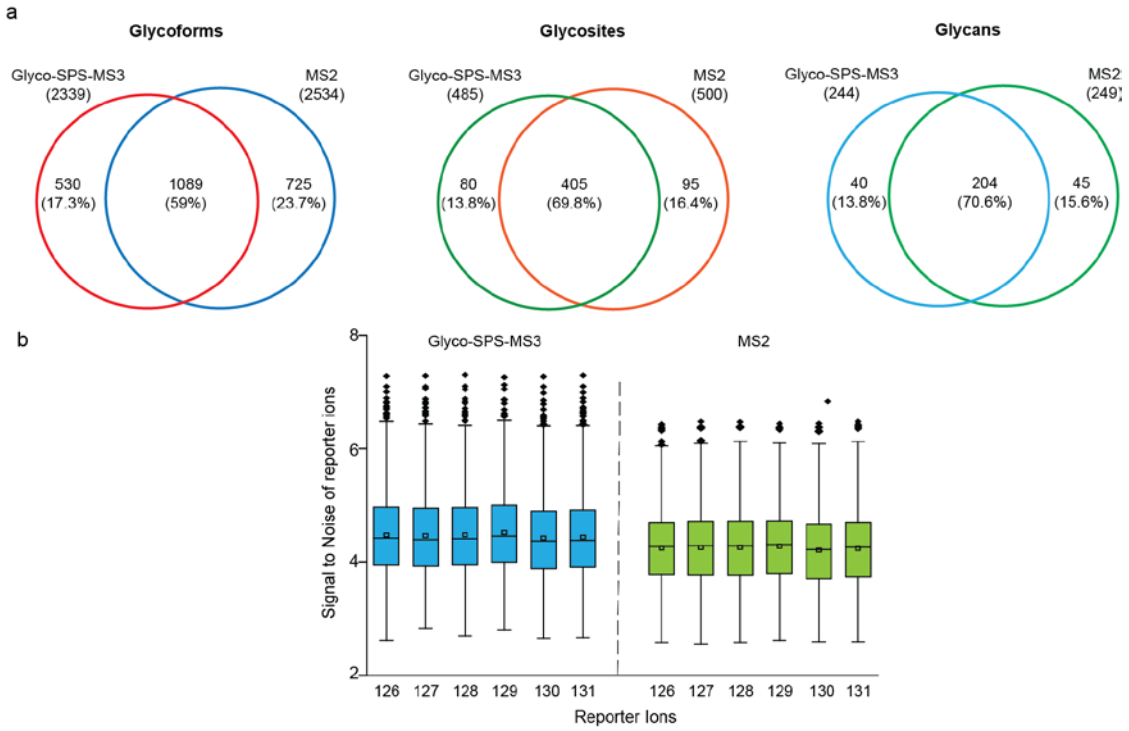


Figure 4.17 Comparison of identifications from MS analyses using Glyco-SPS-MS3 and the optimized MS2 methods in DG75 cells.

(a)The overlaps of identified glycoforms (left), glycosites (middle) and glycans (right) using MS2 or Glyco-SPS-MS3 methods shown in Venn diagrams. (b) The reporter ions signal to noise (S/N) obtained from Glyco-SPS-MS3 or MS2 methods.

As for quantitation, we compared the N-glycosylation expression resulting from the Glyco-SPS-MS3 and MS2 methods on the glycoform- , glycosite- and glycan-centric levels in 2FF-treated DG75 cells. In order to exclude that the observed altered glycosylation was not caused by a change in substrate protein expression, we performed the single-shot proteome quantitation analysis. The results showed that the overall protein expressions in 2FF treated samples upon various concentrations were not changed (**Figure 4.18,** the single-shot proteome data can be found in **Supplementary Data S3** in the appendices). Furthermore, in order to eliminate the variability occurring in the earlier sample preparation steps namely before mixing multiplex labelled samples, the median of all the proteins' quantitative ratios in each concentration of 2FF-treated sample was used for the calibration of their corresponding glycoproteome data as normalization factor (NF) (**Figure 4.18**). The decrease of the expression of fucosylated glycoforms under 2FF treatment at all the concentrations was captured in Glyco-SPS-MS3, while the expression of non-fucosylated glycoforms remained unchanged (**Figure 4.19a, b**). However, MS2 analysis diminished the evident decrease in fucosylated glycoforms. Due to the fact that the glycoforms containing fucose in their glycan moiety were only 40% of the total quantified glycoforms, we assumed that the quantitation of fucosylated glycoforms in the MS2 method was heavily hampered by the non-fucosylated glycoforms and the reduced co-isolation interference of TMT reporter ions improved the quantitation accuracy in Glyco-SPS-MS3 as well. Our results showed a substantial decrease of fucosylated glycans, that was accompanied by a concentration-dependent increase of non-fucosylated ones, in 2FF-treated DG75 cells (**Figure 4.19c**). In addition, we also detected 2FF-induced site-specific changes in glycosylation occupancy in Glyco-SPS-MS3 (**Figure 4.19d**), which was lost in previous glycan-centric analyses [191, 192].

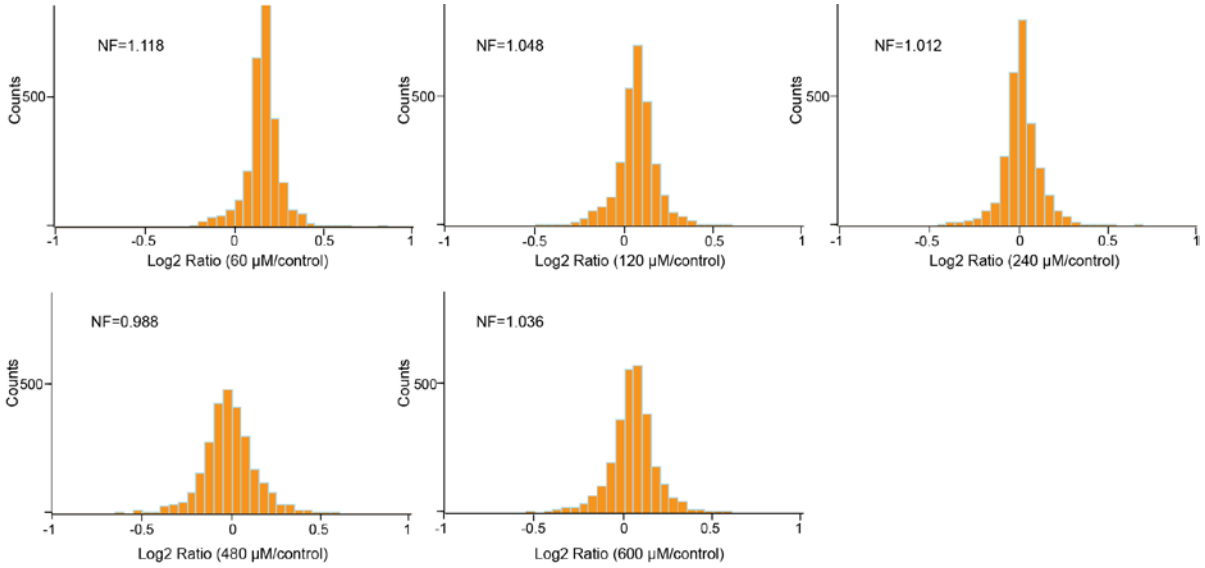


Figure 4.18 Quantitative proteome analyses of DG75 cells that were treated with various concentrations of 2FF.

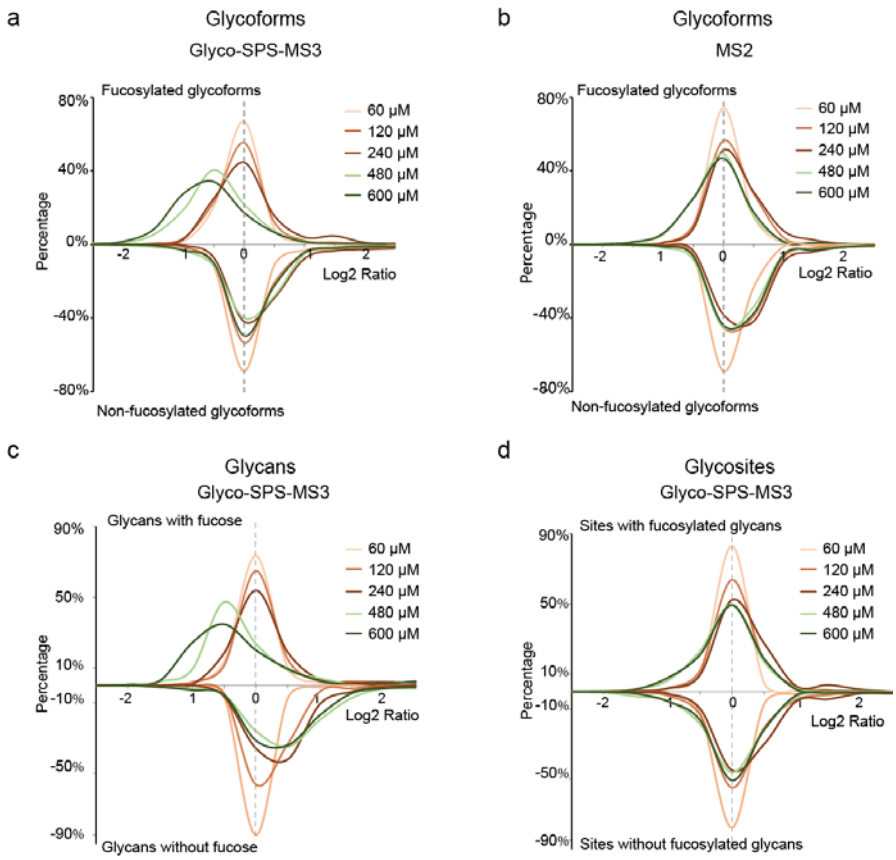


Figure 4.19 Multi-dimension quantitative N-glycoproteomics of 2FF-treated DG75 cells.

Ratio distributions of glycoforms quantified via Glyco-SPS-MS3 (a) or the MS2 method (b) upon 2FF treatment. Fucosylated glycoforms (upper panel) and non-fucosylated ones (lower panel) are shown separately. Ratio distributions of glycan (c) or glycosite (d) levels determined by Glyco-SPS-MS3 upon 2FF treatment. 2FF treatments with various concentrations are color coded as shown.

Next we investigated significantly regulated intact glycopeptide patterns (site-specific glycoform) and glycosite level from MS analysis using the Glyco-SPS-MS3 method in 2FF-treated DG75 cells. Since the expression of intact glycopeptide in the 60 µM 2FF-treated sample did not show significant decrease, the quantitative ratio corresponding to a Z score of -2 in 60 µM 2FF-treated sample was used as the cut-off to filter the significantly regulated glycoforms (-1.246 of $\log_2$ ratio) or glycosites (-1.189 of $\log_2$ ratio) in 480 µM 2FF-treated sample. We found a total of 321 glycoforms (intact glycopeptide level) with decreased expression, which occurred on 138 glycosites mapping to 105 glycoproteins. In addition, 48 glycosites on 46 proteins showed decreased site occupancy on the glycosite level. Among 138 glycosites with significantly decreased glycoforms, 42 exhibited lower glycosylation levels (All the significantly regulated glycoforms and glycosites can be found in **Supplementary Data S2** in the appendices). Our results revealed that not only fucosylated glycopeptides, but also site-specific glycoforms and glycosylation site levels were impaired by 2FF treatment. Functional analysis of those 2FF-affected glycoproteins revealed that most of them were highly related to lymphoma (**Figure 4.20a**), such as, enzymes, CD molecules, immunoglobulins on the B cell membrane, integrins, and MHC type II molecules. The glycoproteins also which showed highly confident functional connections between them in STRING database (**Figure 4.20b**). In addition, we found that some of these 2FF-affected glycoproteins including CD54, CD50, SLC3A2, ALCAM, CALRL, TFR1, PTPRJ and PPT1, mediated many vital biological processes such as cell proliferation, metastasis, migration and tumorigenesis and were used for potential drug targets. The annotations of these glycoproteins altered by 2FF can be found in **Supplementary Data S4** in the appendices.

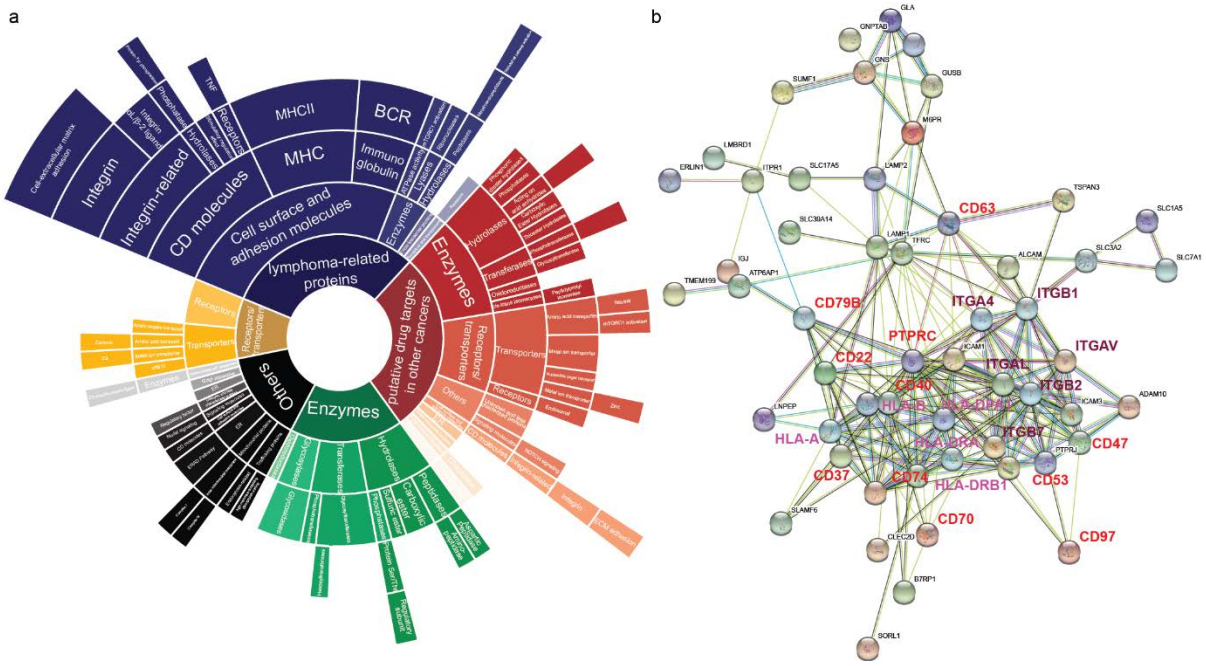


Figure 4.20 Functional annotation of the glycoproteins with regulated glycosylation upon 2FF treatment in DG75 cells.

(a) A sunburst plot of the enriched functions of the glycoproteins with significantly changed glycosylation upon 2FF treatment. Dr. Momchil Ninov performed the analysis and made this figure. (b) The STRING interaction networks of the glycoproteins with changed glycosylation upon 2FF treatment. Dr. Pan Fang made it.

4.1.5.2 Quantitative Phosphoproteome analysis in 2FF-treated DG75 cells

From the quantitative glycoproteomic analysis, we learnt that many critical effectors participating in BCR signaling were 2FF-affected glycoproteins, which implies that 2FF-induced glycosylation change could affect BCR signaling. Indeed, it has been reported that 2FF suppresses downstream signaling pathways for tumor growth, such as phospho-EGFR, -AKT, -ERK, and –FAK43 by decreasing the fucosylation expression of membrane glycoproteins such as the EGF receptor and integrin β1. However, there is no large-scale analysis of 2FF-induced phosphorylation change till now. We thus performed a global quantitative phosphoproteome analysis of 2FF-treated DG75 cells using the flow-through fraction from glycopeptide enrichment in order to explore about the underlying mechanisms of 2FF induced- phosphorylation changes in Burkitt's lymphoma cells. We enriched the phosphopeptides from the unbound fraction of glycopeptide enrichment by titanium dioxide beads followed by basic reverse phase prefractionation. Double replicates were analyzed from each fraction on Orbitrap Tribid mass spectrometer with SPS-MS3 acquisition method.

In total, 14867 unique phosphosites (p-sites) were identified, almost 60% (8901) of them were class I sites whose localization probabilities were more than 0.75. These 8901 sites corresponded to 10288 phospho-events (p-events) from 3040 proteins (The data can be found in **Supplementary Data S5** in the appendices). One, two and three of p-events corresponded to the quantified phosphopeptides with one, two and more-than two p-sites. In detail, a phosphopeptide with one p-sites meant one p-event. A phosphopeptide with two p-sites mapped to three p-events at most. Similar to N-glycoproteomics analysis, all the phosphoproteome data was calibrated using the normalization factor from the single-shot proteome data for further analysis in order to eliminate the variability occurring in the earlier sample preparation steps. Compared to 60 µM 2FF treatment, much more p-events were significantly down-regulated in high concentration 2FF-treated sample (480 and 600 µM) than in lower concentration 2FF conditions (120 and 240 µM). Moreover, the numbers of up-regulated p-events decreased as the concentration of 2FF increased in DG75 cells (**Figure 4.21**). We assumed that protein dephosphorylation was induced via the decreased fucosylated glycopeptide expression in 2FF-treated DG75 cells. Under 480 µM 2FF treatment, 482 up-regulated and 1418 down-regulated phosphor-events were detected compared to 60 µM 2FF condition from the two-sample t-test with 0.5% FDR. We noticed that these 2FF-affected phosphoproteins included important effectors involving BCR signaling, such as, NFAC1, CD19, SYK, NFKB1, PTPRC, NFATC2 and BLNK, suggesting that the suppression effects on BCR signaling concomitant with 2FF-induced de-fucosylation.

A total of 928 up-regulated and 1459 down-regulated p-events were identified in various concentrations of 2FF treatments. Interestingly, 68 up-regulated p-events in low concentration 2FF-treated sample were found to be down-regulated ones in high concentration of 2FF treatment. These phosphopeptides were mainly from Bcl2-associated transcription fator1 (BCALF1) and other transcription factors. The functional analysis of 2FF-regulated phosphoproteins with down-regulated p-events based on KEGG pathway annotation revealed that affected phosphoproteins were mainly enriched in the pathways relating to spliceosome, RNA transport, and cell cycle (**Figure 4.22**).

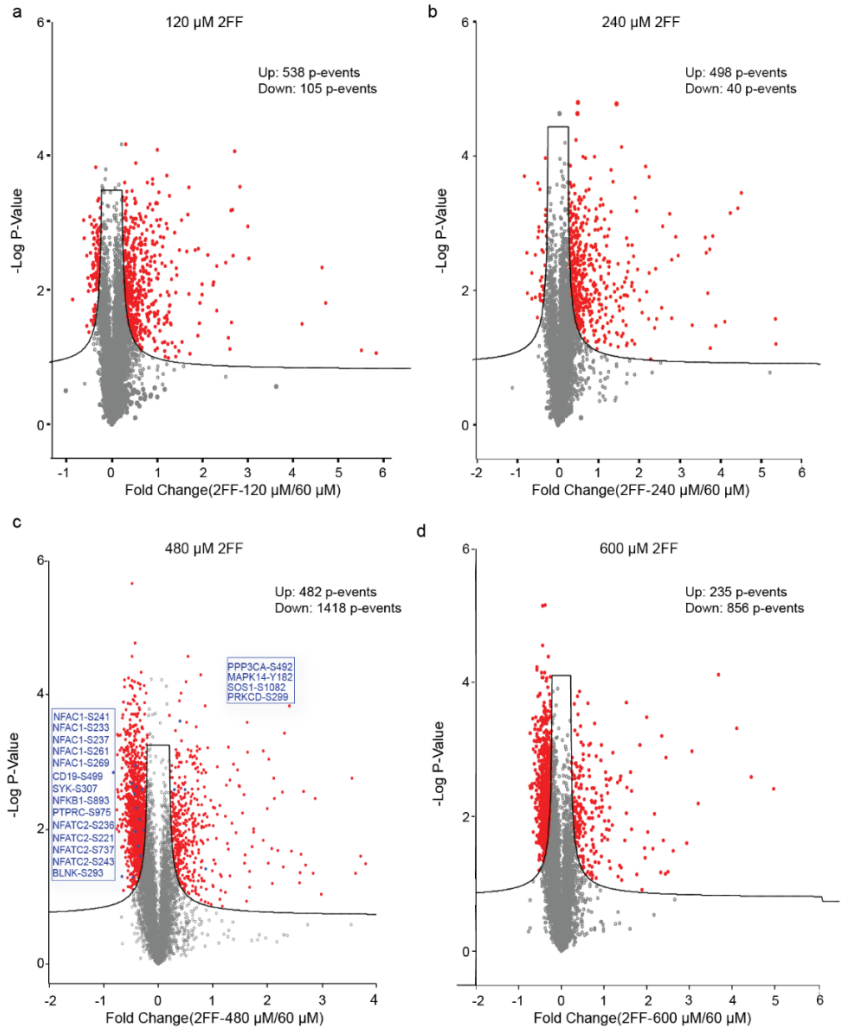


Figure 4.21 Quantitative phosphoproteome analysis in various concentrations of 2FF compared to 60 μM 2FF treatments in DG75 cells.

The significantly regulated phosphorylation in (a) 120 μM (b) 240 μM (c) 480 μM (d) 600 μM 2FF treatments compared to 60 μM 2FF treatment was shown in volcano plot. Significantly regulated p-events were marked as red. In 480 μM condition, 2FF-affected phosphoproteins with their regulated phosphosites were highlighted.

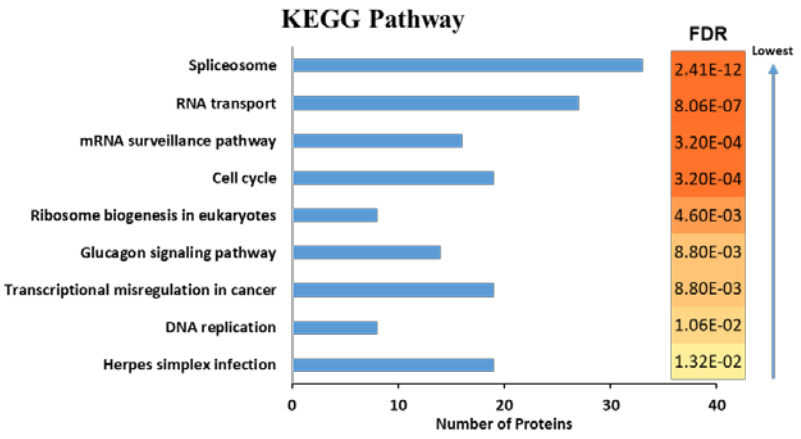


Figure 4.22 Functional enrichment analysis of the phosphoproteins with regulated phosphorylation in 2FF-treated DG75 cells.

FDR based on Benjamini-Hochberg procedure was listed.

4.2 Quantitative N-glycoproteomics analysis of diffuse large B-cell lymphoma subtypes using SugarQuant

The classification of heterogeneous cancer subtypes is helpful and important for the clinical diagnostics and treatment. A molecular classification of histologically and morphologically indistinguishable DLBCLs into three subtypes including GCB, ABC and unclassified one was achieved by gene expression profiling. It also demonstrated that the consistent classification of DLBCL into ABC and GCB subtypes based on their protein expressions using SILAC-based quantitative proteomics analyses [223]. Although more and more evidence shows the relationship between aberrant glycosylation and cancer development, the mechanisms of regulation involved by receptor glycosylation are still very difficult to interpret. This is because of the inherent complexity of the glycosylation synthesis, which is mediated by various glycosyltransferases, and the resulting micro- and macro-heterogeneity of glycans. A previous study showed that quantitative glycosylation site-occupancy profiling from deglycoproteomics analysis enabled to classify the two subtypes of ABC and GCB DLBCL [79]. Due to the limit of technology at that time, de-glycosylated peptides were used for MS analysis after releasing N-glycans enzymatically, thus resulting in the loss of the important information of glycan moieties and the linkage between glycan and protein. In this study, we investigated the site-specific quantitative N-glycoproteome in the subtypes of DLBCL using SugarQuant. Finally, it enabled us to precisely determine the intact glycopeptide expression, glycosite-occupancy among the used cell lines.

4.2.1 Strategy for site-specific quantitative N-glycoproteomics analysis in DLBCL

Five ABC cell lines including HBL1, TMD8, OCI-Ly3, OCI-Ly10 and RIVA and four GCB cell lines including DOHH2, SUDHL4, SUDHL5 and WSU-FSCCL were selected to investigate their site-specific quantitative N-glycoproteome using our developed SugarQuant platform (**Figure 4.23**). Triple biological replicates per cell line analysis were performed in total.

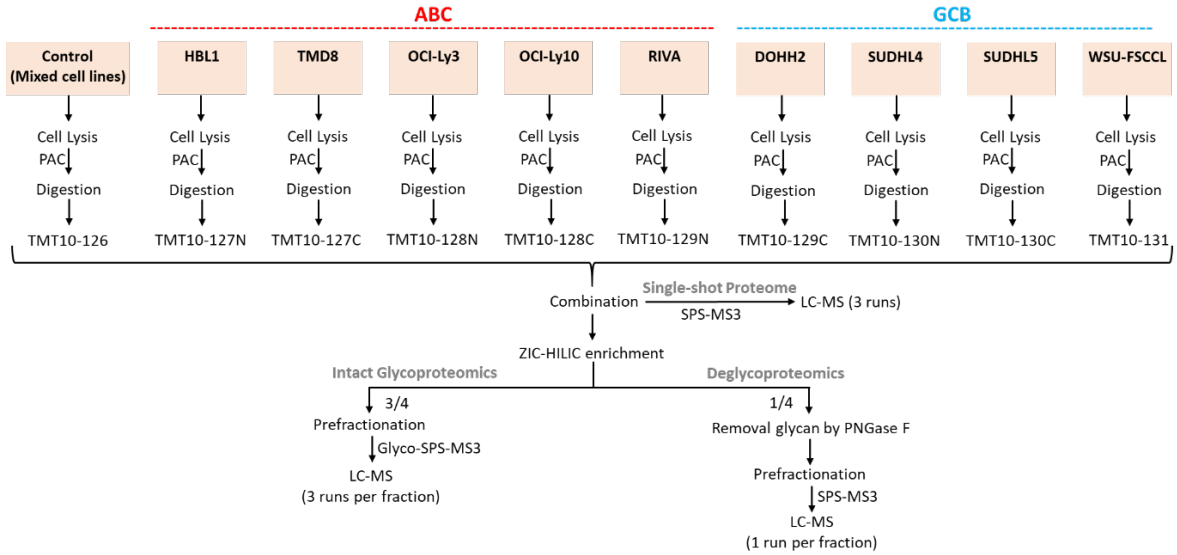


Figure 4.23 Workflow of quantitative N-glycoproteomics analysis in 5 ABC and 4 GCB DLBCL cell lines.

Briefly, we mixed the extracted proteins from nine cell lines in equal protein amount to prepare a pooled control for relative quantification. For the control and each of the cell lines, equal protein amounts were taken for following PAC-based sample preparation and TMT10 labelling. Then we mixed TMT10 multiplexed labelled digested peptides and saved one aliquot (~10 µg) for single-shot proteome analysis using conventional SPS-MS3 method. Three-fourths of enriched TMT labelled glycopeptides were

prefractionated using bRP followed by LC-MS analysis using the Glyco-SPS-MS3 method, which was regarded as site-specific or intact N-glycoproteomics analysis. From the remaining 25%, N-glycans was removed the by PNGase F to obtain the deglycosylated peptides. Then we performed bRP and collected 8 fractions for MS analysis using conventional SPS-MS3 method. We defined it as deglycoproteomics which was used to evaluate the quantitation accuracy of intact glycopeptides in site-specific N-glycoproteomics analysis using SugarQuant.

4.2.2 Single-shot proteome analysis in DLBCL

There are two goals for performing single-shot proteome analysis: (1) To eliminate the occurring variability in the protein digestion step from the parallel manipulation of multiple; (2) To guarantee that the significantly regulated glycopeptide expressions are not derived from a change in the expression level of their corresponding substrate protein.

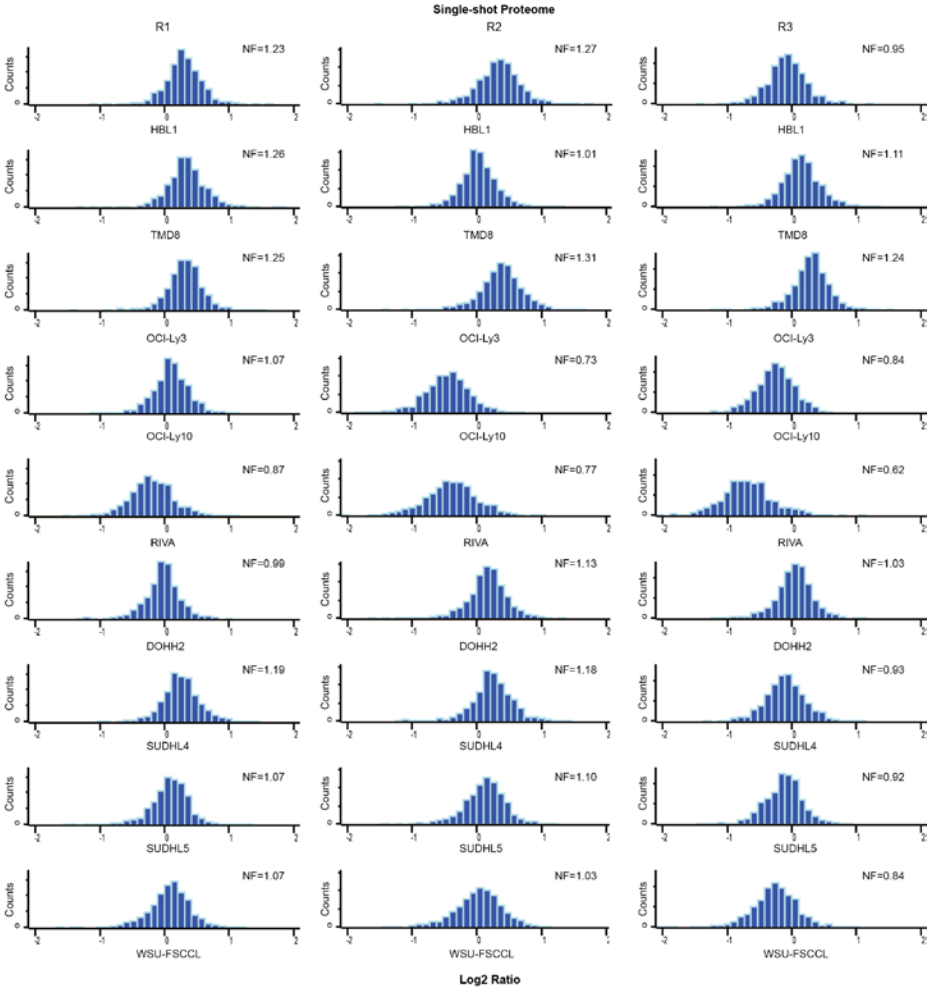


Figure 4.24 Quantitative proteome analysis in DLBCL cell lines.

The normalization factor (NF) was marked for all of them. Histogram showed the distribution of Log_2 transformed ratios of total quantified proteins in each cell line in three biological replicates.

In each replicate and each cell line, the log2-transformed ratios of all quantified proteins mainly located between -1 and +1, indicating that the majority of proteins are not regulated (**Figure 4.24**). Meanwhile, we can find some down-regulated proteins in OCI-Ly10 in R2 analysis, RIVA in R2 and R3 analysis. This could be interpreted by the incorrect 1:1 mixing from the evidence of that the distribution of total protein expression shifted to one side as a whole. To eliminate the interference from the experimental error, we calculated the median of the ratios of all quantified proteins in each cell line in each replicate

as the normalization factor (NF), which was used for the calibration of site-specific N-glycoproteome and deglycoproteome data. All the proteins and their quantitation information in each replicate can be found in **Supplementary Data S6** in the appendices.

4.2.3 Characteristics of the site-specific N-glycoproteome in DLBCL cell lines

We identified 7153 unique glycoforms, which contained 407 glycan compositions on 1052 glycosites from 586 glycoproteins (Data can be found in **Supplementary Data S7** in the appendices). To our knowledge, this is the first large-scale intact N-glycopeptides quantitative analysis in human diffuse large B-cell lymphoma cell lines.

In addition, our results revealed substantial micro-heterogeneity of the site-specific N-glycosylation in DLBCL. At least five different glycan compositions on 28% (303/1052) of all N-glycosites were identified, with an average of 6.8 glycoforms per N-glycosite (**Figure 4.25a**). Extremely, more than 90 different glycoforms were identified on 7 distinct N-glycosites, respectively. For example, 129 glycoforms on Asn249 in LAMP1, 100 glycoforms on Asn129 in CD53, 98 glycoforms on Asn380 in PTPRC, 94 glycoforms on Asn46 in IGHM. Among the 586 glycoproteins mapped to the glycoforms, 363 proteins were identified with a single glycosite. About 38% have two or more different glycosites. Of that, 13 proteins were identified with more than 6 sites (**Figure 4.25b**). For instance, the maximum of 18 sites were found on Sortilin-related receptor (SORL). Receptor-type tyrosine-protein phosphatase eta (PTPRJ) and receptor-type tyrosine-protein phosphatase C (PTPRC), as the tyrosine phosphatase involved in the regulation of B lymphocyte activation, found 9 and 7 sites respectively. 11 sites were identified on Lymphocyte antigen 75 (LY75) which regulation the proliferation of B-lymphocytes. Moreover, some other proteins of regulatory interest were also found to be heavily glycosylated. The majority of the glycoforms bore glycans containing 6–16 monosaccharide moieties with the average of 11 (**Figure 4.25c**).

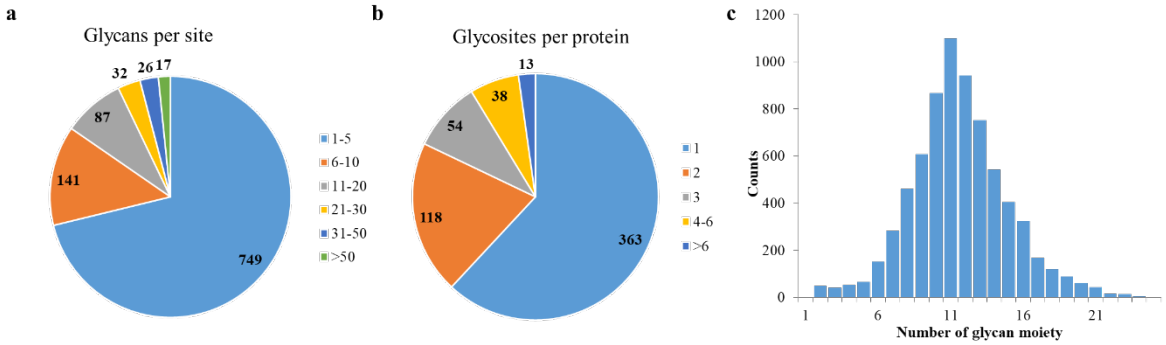


Figure 4.25 Microheterogenetiy of site-specific N-glycosylation in DLBCL.

(a) The distribution of glycoforms identified on per glycosite. (b) The distribution of glycosites identified on per glycoprotein. (c) The distribution of the number of monosaccharides on glycan chain of all identified glycoforms.

More than 6000 glycoforms, about 1020 N-glycosites and 387 glycans on average were quantified in each cell line (**Figure 4.26a, b and c**). Around 77% (5534/7152) glycoforms, 89% (940/1052) glycosites and 85% (344/407) glycans were quantified across all 9 cell lines. In addition, more than 93% glycoforms, 97% glycosites and 96% glycans were quantified in at least 6 cell lines, representing the excellent overlap among all used 9 cell lines (**Figure 4.26d, e and f**). Moreover, we calculated the Pearson correlation coefficients (r) of the ratios in quantified glycoform, glycosite and glycan levels in each cell line among triplicates. The r of 0.6~0.9 confirmed the high reproducibility and low batch effect in our quantitative strategy (**Figure 4.26g, h and i**).

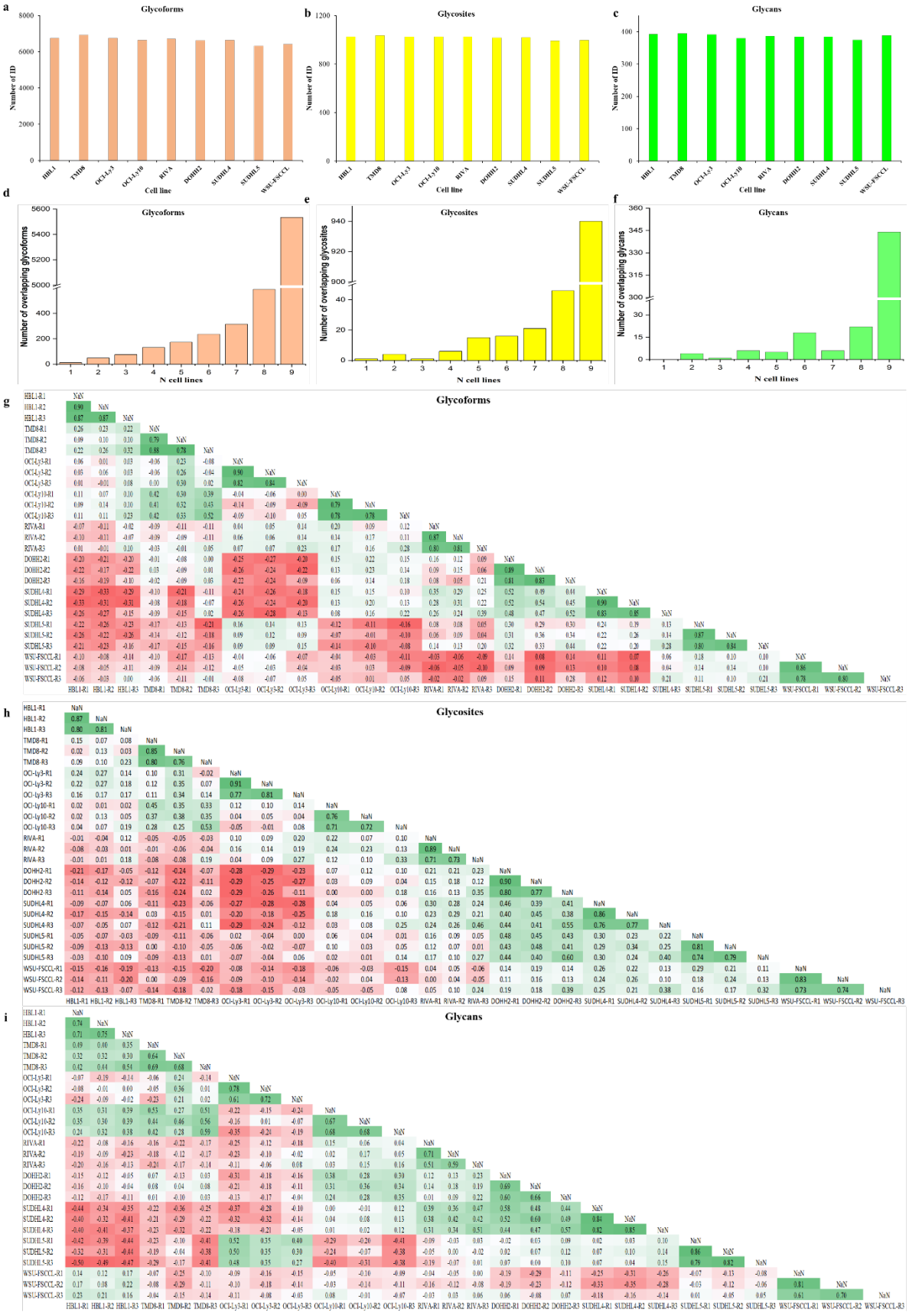


Figure 4.26 Quantified N-glycoforms, glycosites and glycans in DLBCL.

Number of glycoforms (a), glycosites (b) and glycans (c) quantified in each cell line. Overlap of glycoforms (d), glycosites (e) and glycans (f) quantified across the 9 cell lines. Pearson's correlation of glycoforms (g), glycosite (h) and glycan (i) across the three replicates.

4.2.4 Quantitative site-occupancy of N-glycosites in deglycoproteomics (DG)

In deglycoproteomics analysis (DG), we filtered for sites with the localization probability greater than 0.75 as class I sites and those which matched with the consensus sequence motif (N-X-S/T/C; X≠P) [224]. Total 1486 sites were identified and 1468 of them, which attributed to 836 proteins, were quantified at least once in three replicates. The overlapped identification was near 40% in triplicates. The average of 1412 glycosites were quantified in each of 9 cell lines (**Figure 4.27a**). In addition, about 86% (1262/1468) glycosites were quantified in all 9 cell lines, and almost 97% (1417/1468) were quantified in at least 6 cell lines (**Figure 4.27b**). Glycosites in the same cell line among triplicates were co-clustered closely with the correlation coefficients (r) more than 0.7 (**Figure 4.27c**), which demonstrated the very high reproducibility and confident quantitation in the deglycoproteomics analysis. All those commonly identified glycosites sites in the same cell line across triplicates from IG and DG analysis were co-clustered in a close manner with the high Pearson r (**Figure 4.27d**). Indeed, the deglycoproteomics analysis provided further support of the precise quantitation of intact glycopeptides using SugarQuant.

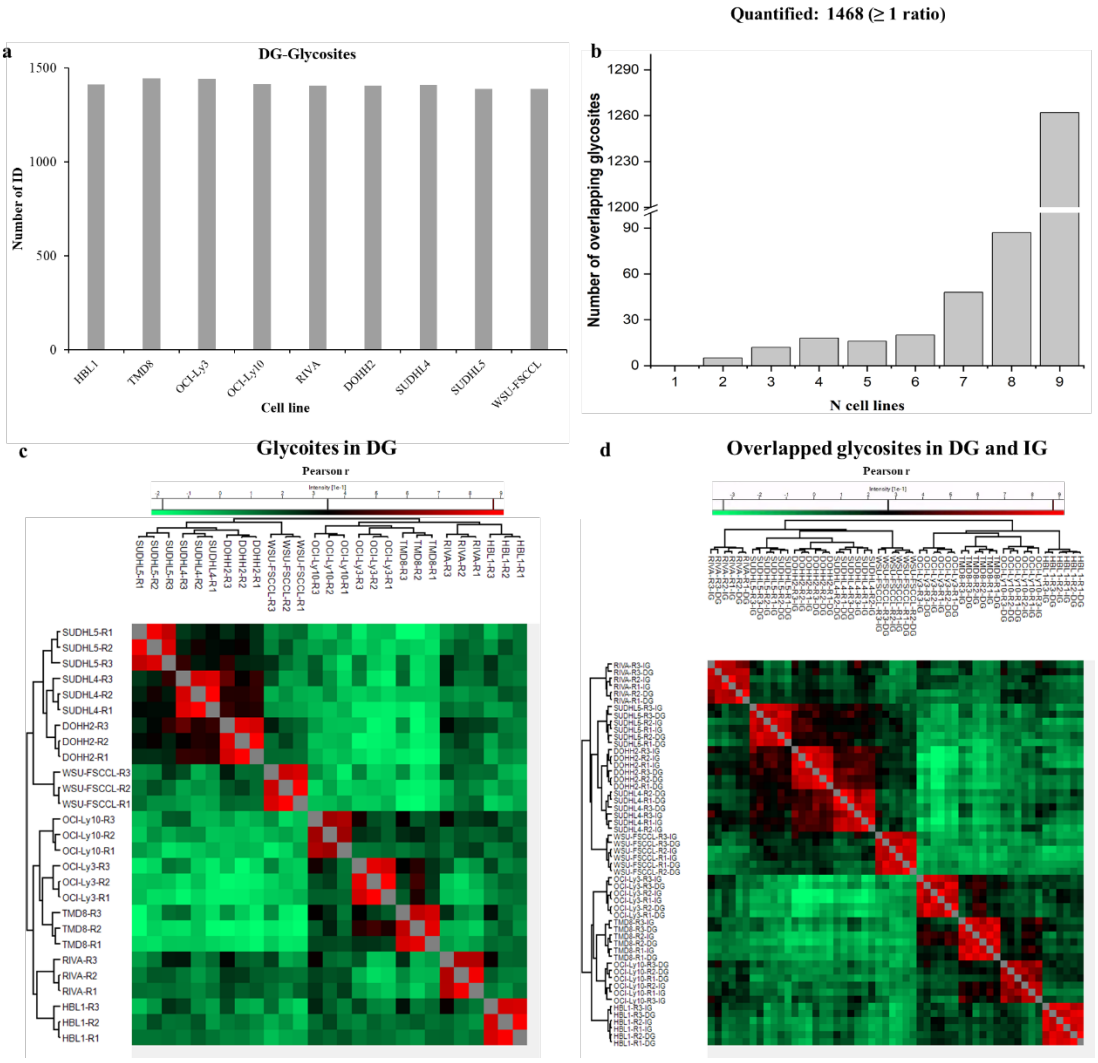


Figure 4.27 Identified and quantified N-glycosites in deglycoproteomics analysis of DLBCL.

(a) Number of glycosites quantified in each cell line. (b) Overlap of glycosites quantified across the 9 cell lines. (c) Heat map of Pearson's correlation across the triplicates in deglycoproteomics analysis. (d) Heat map of Pearson's correlation of the overlapped quantified glycosites in deglycoproteomics (DG) and intact glycoproteomics (IG) across the triplicates.

In DG, the glycopeptides with only one site, two co-existing sites or more than two sites may be quantified simultaneously. We defined them as the different glycosylation events (glyco-events). A total of 1621 quantified glyco-events within at least one sample were used for the further analysis (All the quantified glyco-events in deglycoproteomics analysis can be found in **Supplementary Data S8** in the appendices). To obtain the confident comparison across each cell line, 420 glyco-events, which were quantified in all 27 samples (9 cell lines × 3 biological replicates), were firstly chosen. Subsequently, we performed a multiple-samples t test by ANOVA (analysis of variance) using the Permutation-based FDR with 5% in order to remove the indistinctive glyco-events among the 9 cell lines, resulting in final 374 glyco-events. The unsupervised hierarchical clustering analysis showed that the same cell line within the replicated measurements were always co-clustered closely with each other. The distinct branches of the dendrogram segregated the ABC and GCB cell lines as separated groups (**Figure 4.28**). It demonstrated that these subtypes could be classified as the different entities from the distinct deglycopeptides expression patterns. It also showed no batch effect and good cell line coherency. In previous publication [79] it covered some glycoproteins for the segregation of two subtypes, such as CD40, CD80, MILR1, which were also identified in our result. Moreover, some new differentiating markers were found like PTPRC, TMUB1, ALCAM (CD166 antigen), etc.

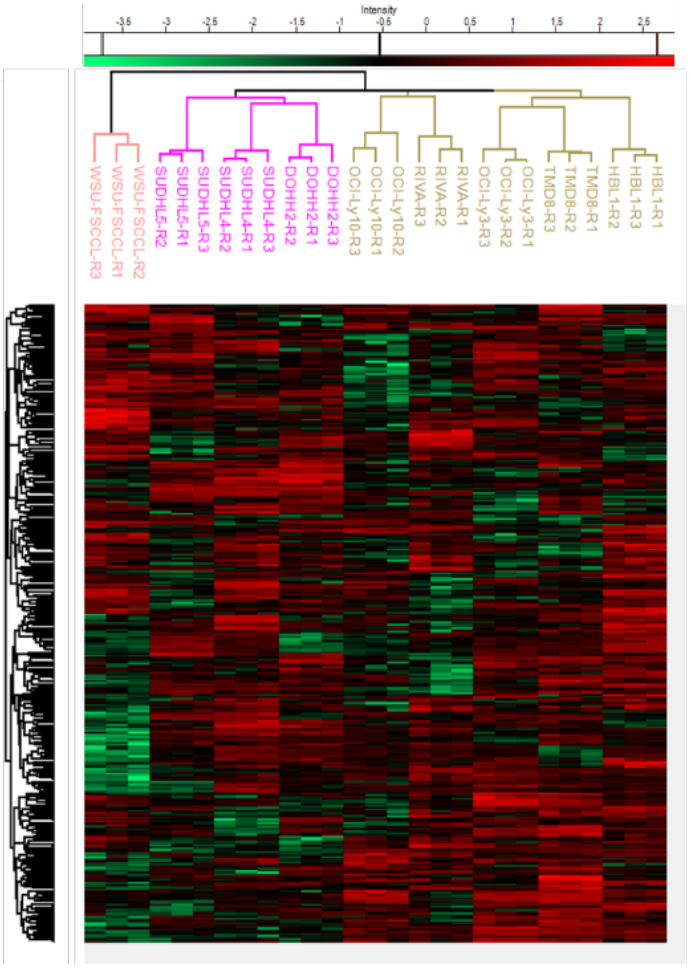


Figure 4.28 Hierarchical clustering of deglycopeptides in DLBCL.

Heat map of the hierarchical clustering (Euclidian distance) of 9 DLBCL cell lines based on their corresponding site-occupancy of N-glycosite in DG.

Subsequently, we performed a two-sample t-test analysis to look for the glycopeptide signature which represented the significant segregation of the two subtypes. 787 glyco-events which had at least two valid ratios in ABC and GCB subgroups were used. It resulted in 194 glyco-events as the signature with the settings of FDR of 0.05 and S0 of 0.1, (**Figure 4.29a**). 108 of 194 glyco-events were quantified within all 27 samples. Of that, 64 deglycopeptides mapping to 43 glycoproteins were up-regulated in the ABC DLBCL subtypes. Some of these 43 glycoproteins played very important roles in BCR signaling. For example, some glycopeptides on IgM, which worked together with CD79 as the B cell receptor to initiate the BCR signaling, were up-regulated. It supported the previous discovery, one way of the initiation of the chronic BCR signaling in ABC DLBCL is via IgM binding to itself glycans [49]. The expression of CD44 was impaired by the high expressed BCL-6 in GCB subtype, which probably resulted in the lower glycopeptide levels in GCB DLBCL compared to ABC DLBCL [225, 226]. In addition, in LY75 and ICAM1, which were important for cell proliferation and adhesion in B-lymphocytes [79], up-regulated glycopeptides were also found in ABC DLBCL. 44 deglycopeptides mapping to 39 glycoproteins were highly expressed in GCB subtypes. Moreover, some of these 39 proteins were already found to be important in GCB DLBCL, such as, MME, the known prognostic marker for GCB, was mainly mediated by NF-κB pathway [79]. This dataset represented that these glycoproteins and their glycosylation had distinct roles in ABC and GCB DLBCL. It will help the researchers in lymphoma field to investigate the distinct mechanisms of tumorigenesis in ABC and GCB DLBCL.

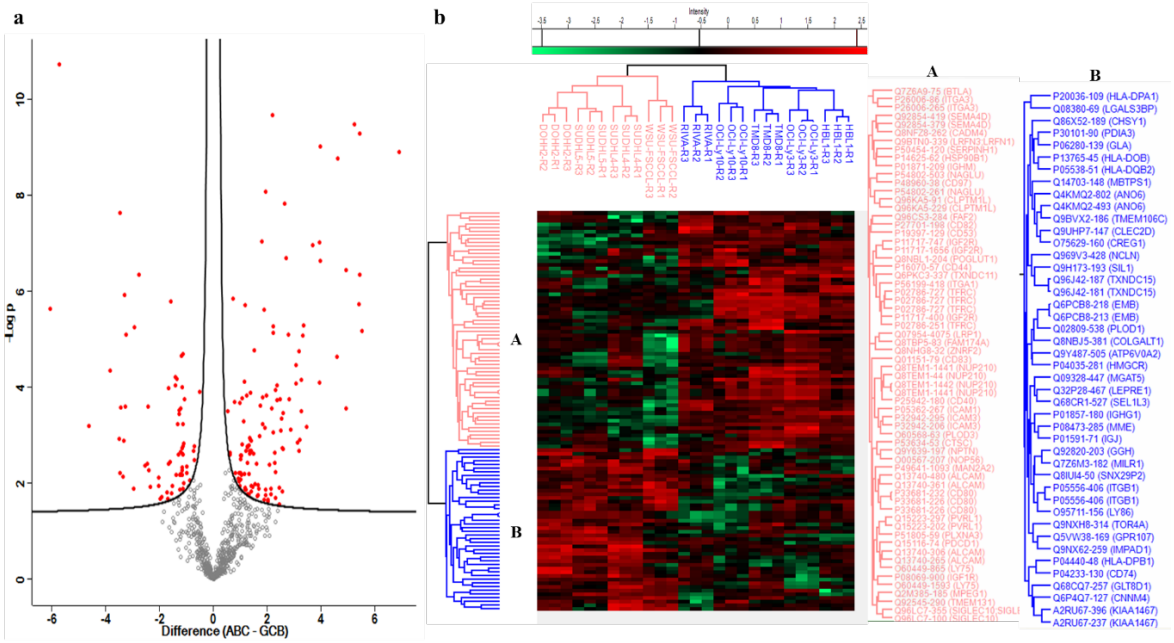


Figure 4.29 Two-sample t test analysis and hierarchical clustering of the signature deglycopeptides.

(a) Two-sample t test analysis for ABC and GCB DLBCL subtypes. The deglycopeptides which significantly segregated these two subgroups were marked in red color. (b) Heat map of 108 glyco-events which were extracted as the most significant signatures after t test analysis of the two subtypes in DLBCL.

4.2.5 Quantitative site-specific N-glycoproteome analysis of DLBCL by SugarQuant

In site-specific N-glycoproteomics analysis, 1410 glycoforms, which were quantified in all 27 samples, were used for the confident comparison among each cell line after two-sample t-test analysis with the settings of FDR of 0.05 and S0 of 0.1. From the PCA analysis, the N-glycoforms of 9 DLBCL cell lines measured in triplicates segregated into ABC-DLBCL (red) and GCB-DLBCL (blue) subtypes based on component 1 which accounts for 22.4% of variability, although the RIVA located on the boundary of

two zones in component 1 (**Figure 4.30b**). The biological replicate measurements across each cell line were always close to each other, eliminating the influence of batch effect among triple replicates. Interestingly, the unsupervised hierarchical clustering of the glycoforms revealed that OCI-Ly3 was co-clustered closely with SUDHL5 and RIVA was co-clustered closely with SUDHL4 and DOHH2. The other 3 ABC cell lines including HBL1, TMD8 and OCI-Ly10 located in one branch of the dendrogram (**Figure 4.30a**).

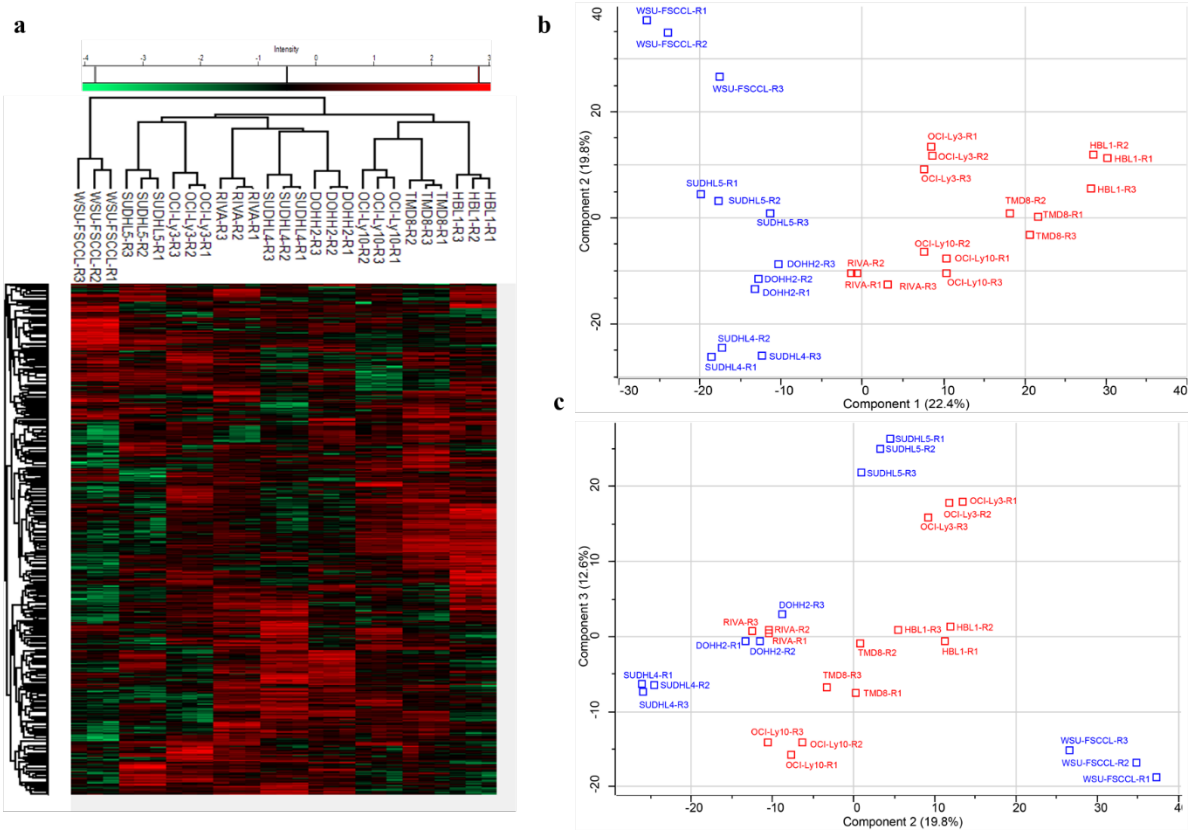


Figure 4.30 Hierarchical clustering and Principal component analysis of glycoforms in DLBCL.

(a) Heat map of the hierarchical clustering (Euclidian distance) of 9 DLBCL cell lines based on their corresponding N-glycoforms expression in IG. (b) The segregation of DLBCL subtypes using the glycoforms of 9 DLBCL cell lines measured in triplicates based on component 1 and component 2 in PCA. (c) The segregation of DLBCL subtypes using the glycoforms of 9 DLBCL cell lines measured in triplicates based on component 2 and component 3 in PCA.

The unique glycoforms expressions in RIVA and OCI-Ly3 cell lines were probably different with those in the other 3 ABC cell lines. However, the difference was not displayed based on the main component 1 (22.4 %) and component 2 (19.8 %) in the PCA analysis. Furthermore, based on the component 2 (19.8 %) and component 3 (12.6 %) in PCA we found that the glycoforms expression in OCI-Ly3 is much closer to those in SUDHL5, the glycoforms expression in RIVA is more similar with those in DOHH2 (**Figure 4.30c**). This result was unexpectedly consistent with the hierarchical clustering analysis of glycoforms in 9 DLBCL cell lines. We deduced that the expression of some specific types of glycosylation modified peptides in OCI-Ly3 and RIVA was different compared to the other three ABC DLBCL cell lines.

Subsequently, we classified the glycoforms into different types based on whether it bore the fucose or sialic acid on the glycan chain. Our results revealed that the glycoforms bore fucose or sialic acid in

OCI-Ly3 and RIVA were differentially expressed in HBL1, TMD8 and OCI-Ly10, resulting in the segregation of OCI-Ly3 and RIVA in distinct zone with the other three ABC cell lines in component 1 in PCA (**Figure 4.31a, c**). Whereas, the glycoforms without fucose or sialic acid on the glycan moiety in 9 DLBCL cell lines measured in triplicates obtained the perfect segregation of ABC and GCB subtypes based on component 1 which accounted for 21.7% or 20% variability (**Figure 4.31b, d**).

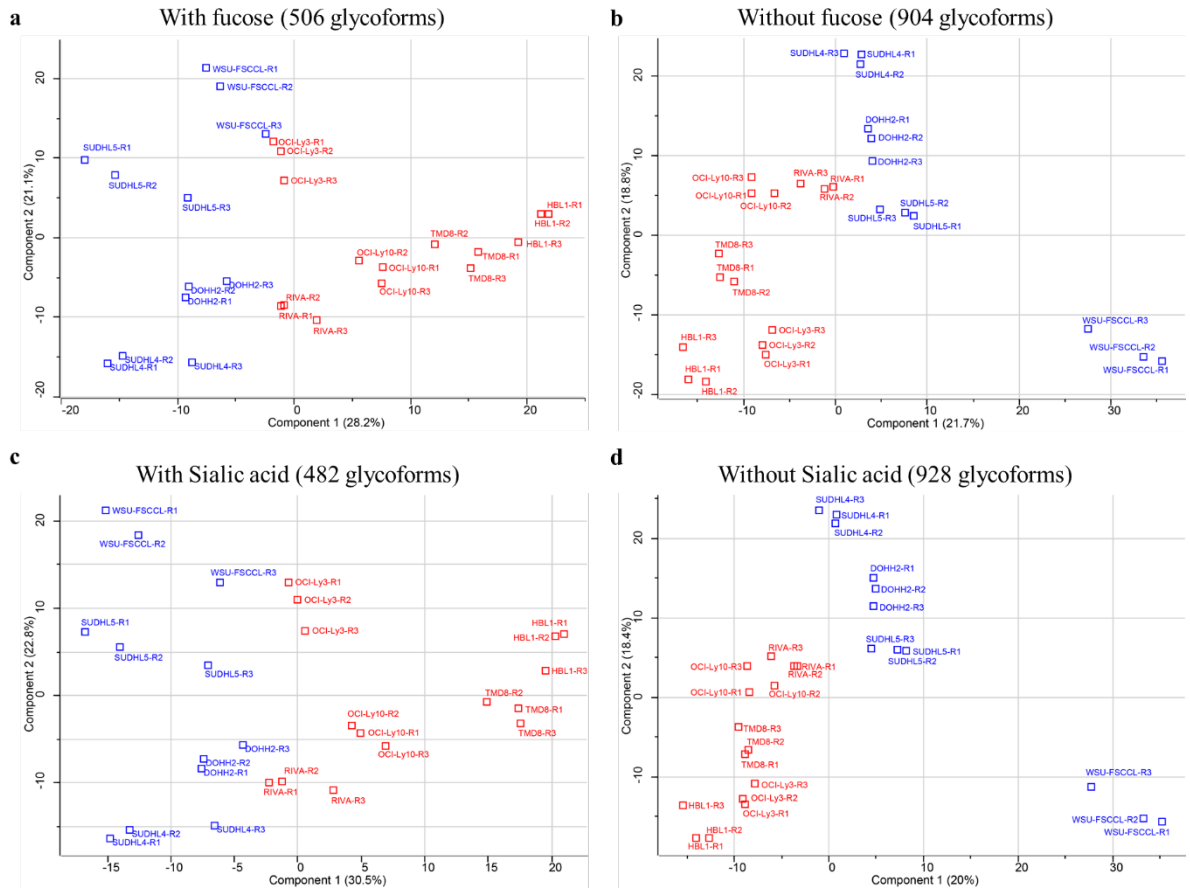


Figure 4.31 Principal component analysis of specific glycoforms bore fucose or sialic acid in DLBCL.

(a) PCA analysis of the glycoforms bore fucose among 9 DLBCL cell lines measured in triplicates. (b) PCA analysis of the glycoforms without fucose on their glycan moiety across 9 DLBCL cell lines measured in triplicates. (c) PCA analysis of the glycoforms bore sialic acid among 9 DLBCL cell lines measured in triplicates. (d) PCA analysis of the glycoforms without sialic acid on their glycan moiety across 9 DLBCL cell lines measured in triplicates.

Irrespective of the specific glycoforms in OCI-Ly3 and RIVA, we performed a two-sample t-test analysis to survey the glycoform signature which represented the significant segregation of these two subtypes. 2678 glycoforms, which had at least two valid ratios in ABC and GCB subgroups separately, were chosen for a two-sample t test analysis with the settings of FDR of 0.05 and S0 of 0.1. It resulted in 1005 glycoforms as the signature (**Figure 4.32a**). 573 glycoforms which were quantified within all the 27 samples as the signature for the segregation of ABC and GCB subtypes were showed in the heat map (**Figure 4.32b**). Of that, 371 and 202 glycoforms were up-regulated in ABC subtype GCB subtype, respectively. Interestingly, higher percentage of fucosylated and sialyated glycoforms were found in ABC DLBCL.

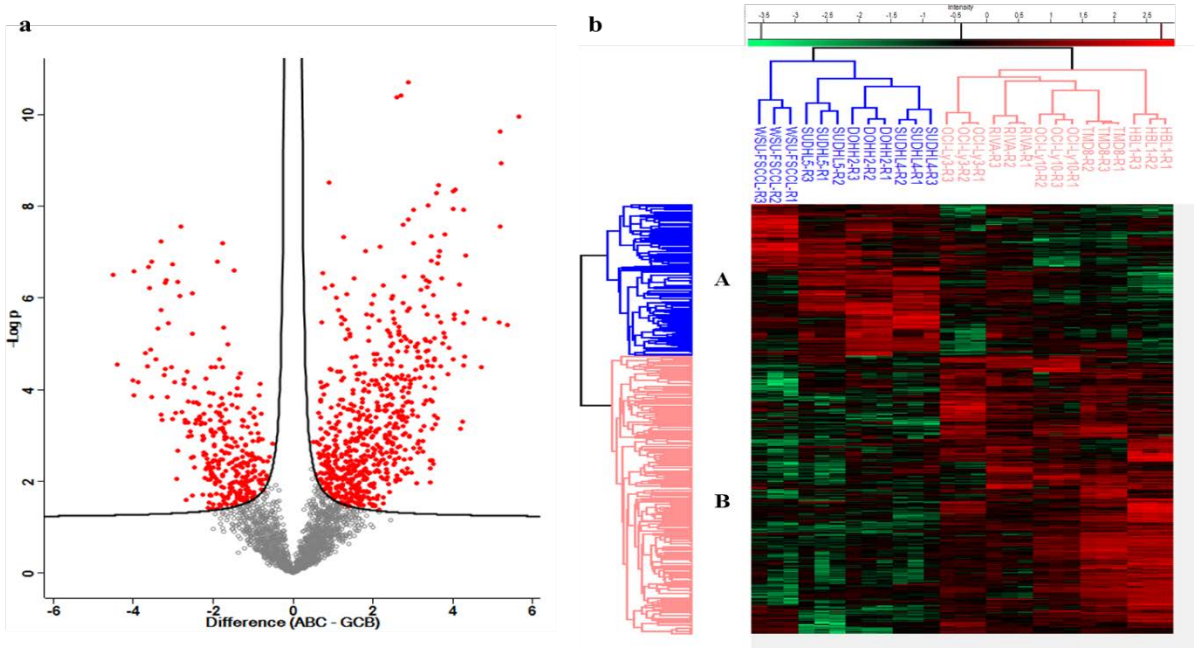


Figure 4.32 Two-sample t test analysis and Hierarchical clustering of the N-glycoforms as a signature for the segregation of DLBCL subtypes.

(a) Two-sample t test analysis for the ABC and GCB subtypes. The glycoforms which significantly segregated these two subgroups were marked in red color. (b) Heat map of the 573 glycoforms which were extracted as the most significant signatures after t test analysis of the two subtypes in DLBCL.

Lastly, site-occupancy of glycosites in IG also enabled to segregate the 9 cell lines into ABC and GCB DLBCL subtypes (**Figure 4.33b**), which was consistent with the results in DG. Furthermore, N-glycans of 9 DLBCL cell lines measured in triplicates couldn't segregate ABC and GCB subtypes (**Figure 4.33b**). We deduced that individual glycan information couldn't represent the glycosylation profiles in lymphoma cell lines. Glycosylation works as the site-specific regulation. Another possible interpretation is that the N-glycan information was only a small portions of existed glycans in lymphoma. A specific glycomic analysis may help for the interpretation.

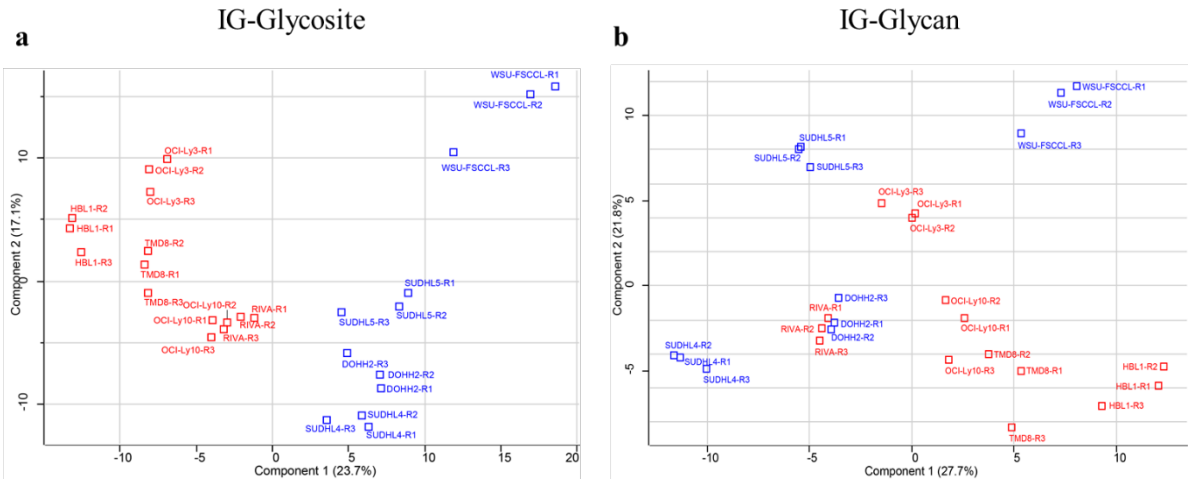


Figure 4.33 Principal component analysis of site-occupancy and glycan in IG.

(a) PCA analysis of N-glycosites among 9 DLBCL cell lines measured in triplicates segregated into ABC and GCB DLBCL subtypes. (b) PCA analysis of glycan composition across 9 DLBCL cell lines measured in triplicates.

4.3 N-glycoproteome analysis in FUT8 knock-out and 2FF-treated TMD8 cells

In this study, we inhibited the fucosylation using 2FF and impaired the core-fucosylation by knockout of FUT8 in a lymphoma cell line and investigated the altered fucosylation expression using our developed site-specific quantitative N-glycoproteomics platform. In our lab, the CRISPR dropout screen data in ABC and GCB DLBCL cell lines revealed that FUT8 is not an essential gene for the survival of TMD8 cell, so that it enabled us to survey the alteration of fucosylated glycopeptide expression. Our above mentioned study revealed that the changed fucosylation expression in Burkitt's lymphoma with 2FF treatment could be detected using the SugarQuant platform. Here, we intended to survey the effect of 2FF in TMD8 cell line and the role of FUT8 in TMD8.

4.3.1 Decrease of Fucosylation levels in 2FF-treated and FUT8 knock-out TMD8 cells

In 2FF-treated TMD8 cells, our result showed that the inhibition of fucosylation levels was in a dose- and time-dependent manner (**Figure 4.34**). Fucosylation expression was almost unchanged in TMD8 cells treated with 2FF at any concentration for 1 day. For 3 days' treatment, decreased AAL signal was detected even at the lowest concentration of 7.5 μM 2FF. Samples treated with 2FF at the concentration of 120 μM and 480 μM for 3 days' culture were chosen for site-specific quantitative glycoproteomics analysis.

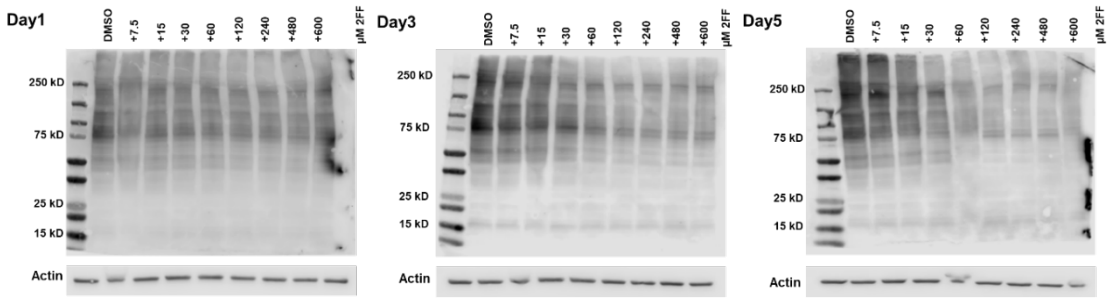


Figure 4.34 Validation of fucosylation expression in 2FF-treated TMD8 cells by lectin blotting.

TMD8 cells were cultured with 2FF at a series of concentrations (600 μM, 480 μM, 240 μM, 120 μM, 60 μM, 30 μM, 15 μM and 7.5 μM) for 1, 3 and 5 days separately. Cell lysates were used for lectin blot analysis using *Aleuria aurantia lectin* (AAL), which enables to bind core fucose in α (1, 2) linkage structure and subterminal fucose in α(1,3/4) linkage structure [185]. Signals for 1 day (left), 3 days (middle) and 5 days (right) were showed separately. Actin was used as loading control.

In FUT8 KO-treated TMD8 cells, western blot showed that all of the three sgRNAs induced the decrease of FUT8 expression (**Figure 4.35a**). In addition, the significantly decreased fucosylation expression was detected in the samples treated with sgFUT8-2 and sgFUT8-3 from lectin blot (**Figure 4.35b**). Then, the samples treated with sgFUT8-2 and sgFUT8-3 were used for site-specific N-glycoproteomics quantitative strategy.

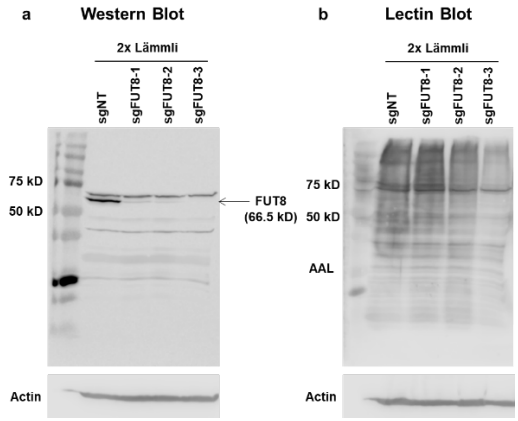


Figure 4.35 Validation of FUT8 KO effects among 3 used sgRNAs in TMD8 and the resulting fucosylation expression in each FUT8 KO-treated TMD8 cells.

2x Lämmli buffer were used to collect FUT8 KO-treated cells as cell lysates, then subjected to western blot analysis using FUT8 antibody (a) and lectin blot analysis using AAL (b). Actin was used as loading control.

4.3.2 Strategy for N-glycoproteomics analysis in 2FF-treated and FUT8 KO TMD8 cells

The fucosylated glycopeptide levels in 2FF-treated and FUT8 KO TMD8 cells were investigated using SugarQuant. Six samples including TMD8 cells treated with 120 µM (samples 1) and 480 µM 2FF (sample2) for 3 days, FUT8 KO treated TMD8 cells using sgFUT8-II (sample 3) and sgFUT8-III (sample 4), TMD8 cells treated with the same amount of DMSO compared to 2FF as control 1 (ctr1) and TMD8 cells treated with the non-sense sgRNA (sgNT) as control 2 (ctr2), were used and labelled with TMT6plex reagents. (**Figure 4.36**).

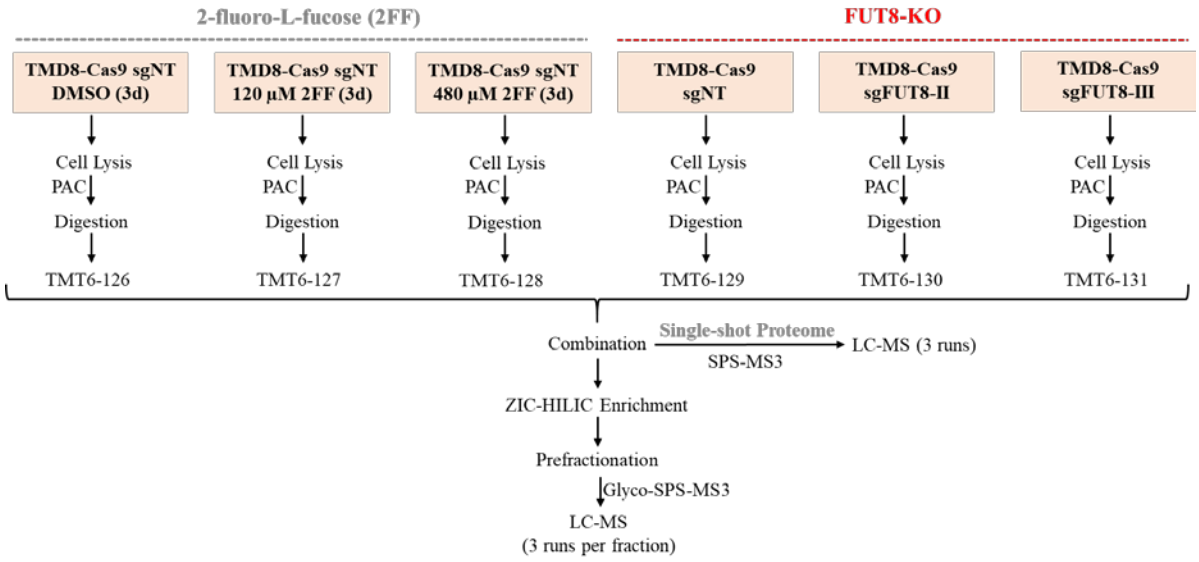


Figure 4.36 Workflow of site-specific quantitative N-glycoproteomics analysis in 2FF-treated and FUT8 KO TMD8 cells.

Briefly, 500 µg extracted proteins from each sample were used for PAC-based protein digestion. The resulting peptides were labelled with TMT6 reagents. Then we mixed TMT6 multiplexed labelled digested peptides and saved one aliquot (~10 µg) for single-shot proteome analysis using conventional SPS-MS3 method. Remaining samples was enriched for glycopeptides enrichment using ZIC-HILIC followed by bRP prefractionation. The collected 8 fractions were measured on Tribid Fusion Lumos with three injection replicates for LC-MS analysis using the Glyco-SPS-MS3 method.

4.3.3 Single-shot proteome analysis in 2FF-treated and FUT8 KO TMD8 cells

In single-shot proteomics analysis, log2-transformed ratios of all quantified proteins in four surveyed samples (refers to section 4.3.2) located between -0.5 and +0.5, indicating that the majority of expressed proteins in 2FF-treated and FUT8 KO samples were not regulated (**Figure 4.37**). Normalization factor (NF) was calculated using the median of ratios of all quantified proteins in each sample with different treatments and was used to calibrate the site-specific N-glycoproteome data in order to improve the accuracy. All the proteins and their quantitation information in each sample with individual treatment can be found in **Supplementary Data S9** in the appendices.

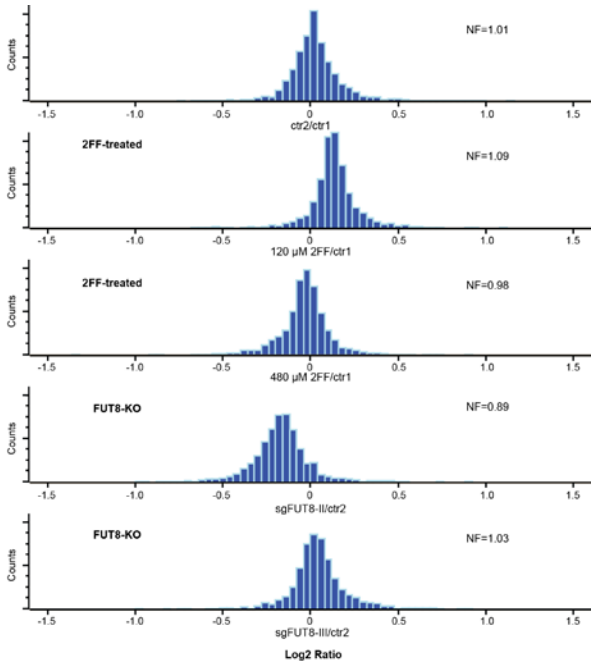


Figure 4.37 Quantitative proteome analysis in 2FF-treated and FUT8 KO TMD8 cells.

The normalization factor (NF) was marked on top right corner. Histogram showed the distribution of Log2 transformed ratios of total quantified proteins in each condition.

4.3.4 General characteristics of N-glycoproteome in 2FF-treated and FUT8 KO TMD8 cells

A total of 4051 unique glycoforms, which included 205 glycan compositions on 731 glycosites from 432 glycoproteins (Data can be found in **Supplementary Data S10** in the appendices). More than 99% (4026 of 4051) glycoforms were quantified. Substantial micro-heterogeneity of the site-specific N-glycosylation in the samples was revealed. More than five different glycan compositions were identified on about 27% (197/731) of N-glycosites, with an average of 5.5 glycoforms per site (**Figure 4.38a**). About 37% have two and more than two glycosites. Of that, 7 proteins were identified with more than 6 sites (**Figure 4.38b**). The majority of the glycoforms bore glycans containing 6–16 monosaccharides with the average of 10 (**Figure 4.38c**).

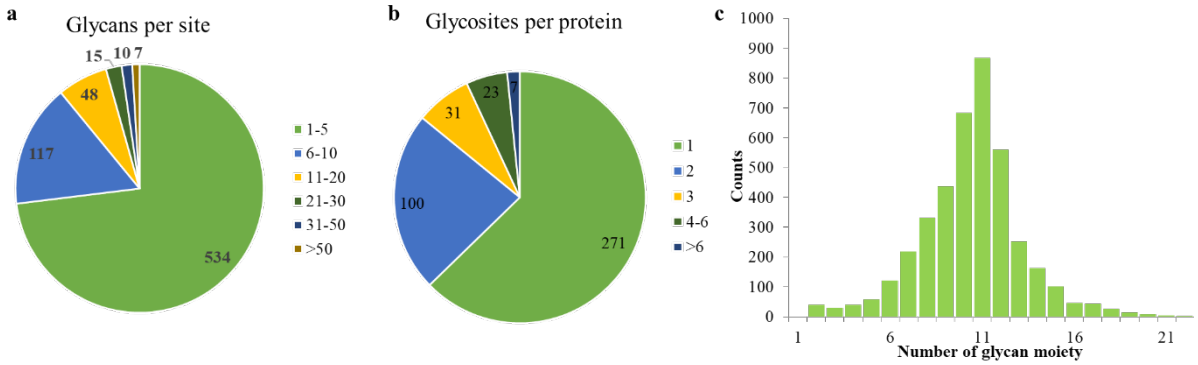


Figure 4.38 Microheterogenetiy of site-specific N-glycosylation in 2FF-treated and FUT8 KO TMD8 cells.

(a) The distribution of glycoforms identified on per glycosite. (b) The distribution of glycosites identified on per glycoprotein. (c) The distribution of the number of monosaccharides on glycan chain of all identified glycoforms.

The high correlations of quantified glycoforms in the conditions of 2FF treatment and FUT8 KO in TMD8 cells glycopeptides expression were revealed, indicating the majority of glycopeptide expressions in 2FF-treated and FUT8-KO TMD8 cells were not influenced (**Figure 4.39a**). It was consistent with that only 24.6% (992/4026) glycoforms were fucosylated in the dataset. In addition, the quantitation of glycopeptides in two controls was almost the same (**Figure 4.39b**), eliminating the controls' influence in 4 samples' comparison analysis.

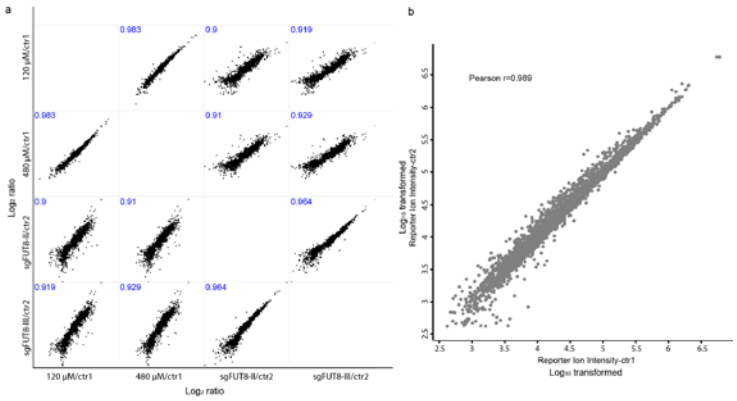


Figure 4.39 Correlation of quantified glycoforms in 2FF-treated and FUT8 KO TMD8 cells.

(a) Pearson correlation of quantified glycoforms in 4 samples with 2FF treatment or FUT8 KO. The correlation efficient r was marked in upper left corner in each comparison. (b) The comparison of reporter ion intensity in control 1 and control 2 samples. The addition of DMSO in control 1 during cell culture is the only difference with control 2.

4.3.5 Multi-dimensional quantitative N-glycoproteome in 2FF-treated and FUT8 KO TMD8 cells

From the analysis of the distribution of total quantified glycoforms in each 2FF treatment and FUT8 KO sample, some down-regulated (log2 ratio <-1) and up-regulated (log2 ratio >1) glycoforms were identified in each sample. The distribution of fucosylated glycoforms (24.6%, 992 in 4026) moved to the left side as a whole in each sample. The mean of 120 μM and 480 μM 2FF treated samples showed a 2-3 fold decrease and the mean of sgFUT8-III treated sample showed a near 8 fold decrease (**Figure 4.40a**). Interestingly, if we separated the fucosylated GPSMs (glycopeptide-spectrum-matches) into two types based on whether it was core-fucosylated or terminal-/subterminal-fucosylated (also called non-core-fucosylated), almost all the core-fucosylated GPSMs expression decreased in each sample. A large percentage of non-core-fucosylated GPSMs (<2% FDR) was up-regulated (**Figure 4.40b**). In non-core-

fucosylated GPSMs, about 82.3% (404/491) were sialyated and the majority of the sialyated non-core-fucosylated GPSMs were up-regulated with a ~4 fold and ~7-fold increase as a mean in 2FF-treated and FUT8 KO TMD8 cells, respectively (**Figure 4.40c**). The up-regulated sialyated peptides could be interpreted due to the inhibition of terminal fucosylation. The same terminal structure of glycan was competed by fucosyltransferases and sialytransferases.

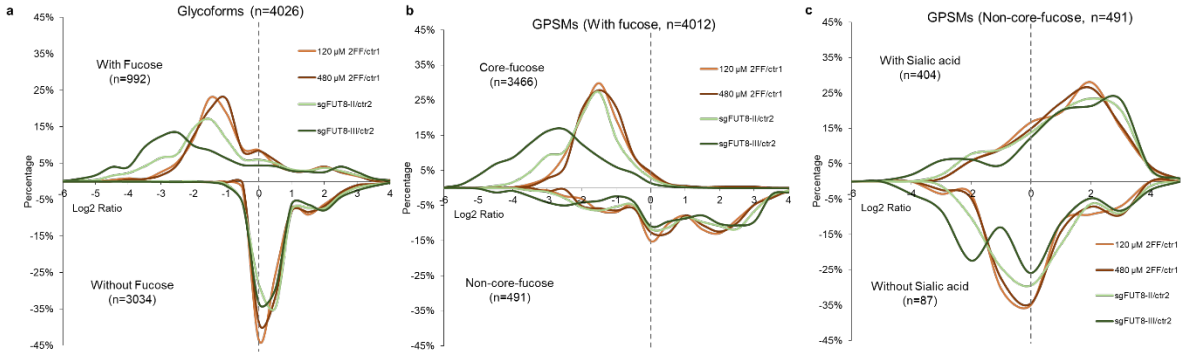


Figure 4.40 Quantitative N-glycoproteome (glycoform and GPSMs) in 2FF-treated and FUT8 KO TMD8 cells.

(a) Ratio distributions of fucosylated glycoforms (upper panel) and non-fucosylated ones (bottom panel) were shown separately. (b) Ratio distributions of core-fucosylated GPSMs (upper panel) and non-core-fucosylated ones (bottom panel) were shown separately. (c) Ratio distributions of sialyated (upper panel) and non-sialyated (bottom panel) non-core-fucosylated GPSMs were shown separately. 2FF treatment with different concentrations and FUT8 KO with 2 distinct sgRNAs were color coded as shown.

Subsequently, we separated all GPSMs into two groups depending on whether their glycan chains bore sialic acid or not. Expectedly, we observed the peaks in the positive and negative sides in the cure of the distribution of 9906 sialyated GPSMs (**Figure 4.41a, upper**). The abundance of 3.8% (405/10729) fucosylated and non-sialyated GPSMs in non-sialyated ones (**Figure 4.41b, upper**) and 36.4% (3607/9906) fucosylated and sialyated GPSMs in sialyated ones (**Figure 4.41c, upper**) decreased more than 2 folds, confirming the efficient fucosylation inhibition by 2FF treatment and FUT8 KO. In non-fucosylated sialyated GPSMs, a peak at the log2 ratio of 2 was found, showing that a large part of non-fucosylated but sialyated GPSMs were up-regulated (**Figure 4.41c, bottom**). The expressions of glycopeptides whose glycan chains bore core-fucose were impaired (**Figure 4.41d, upper**). In addition, the majority of non-core-fucosylated but sialyated GPSMs were up-regulated (**Figure 4.41d, bottom**). More interestingly, the increased extent of these GPSMs in FUT8 KO TMD8 cells (6 - 8 fold as a mean) was higher than those in 2FF treated TMD8 cells (~4 fold as a mean) (**Figure 4.41d, bottom**). It also showed that FUT8 involved the regulation of subterminal-/terminal-fucose synthesis except the core-fucose synthesis.

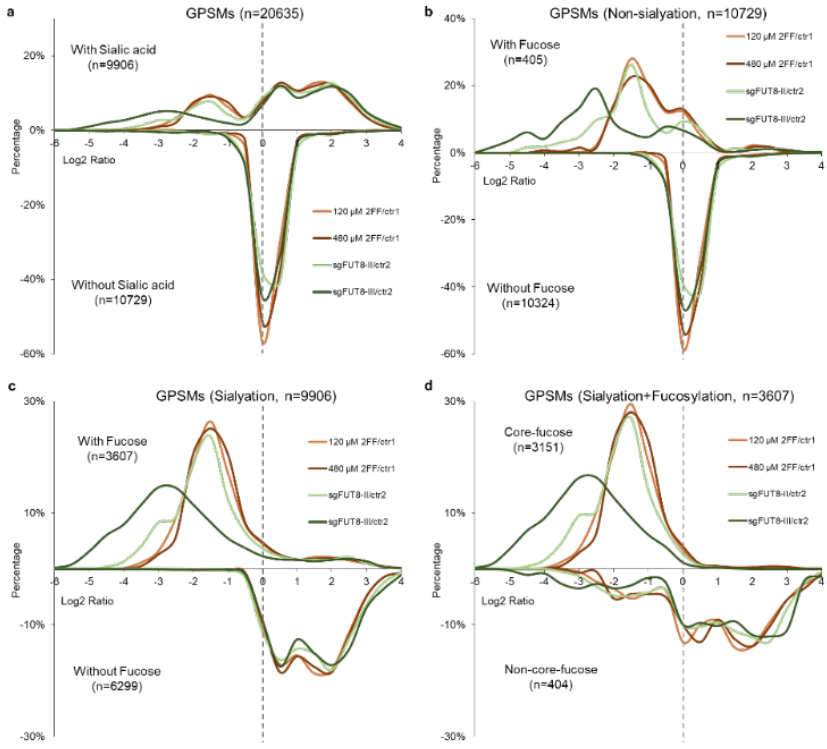


Figure 4.41 Quantitative N-glycoproteome (GPSMs) in 2FF-treated and FUT8 KO TMD8 cells.

(a) Ratio distributions of sialyated GPSMs (upper panel) and non-sialyated ones (bottom panel) were shown separately. (b) Ratio distributions of fucosylated (upper panel) and non-fucosylated (bottom panel) non-sialyated GPSMs were shown separately. (c) Ratio distributions of fucosylated (upper panel) and non-fucosylated (bottom panel) sialyated GPSMs were shown separately. (d) Ratio distributions of core-fucosylated (upper panel) and non-core-fucosylated (bottom panel) ones in sialyated and fucosylated GPSMs were shown separately. 2FF treatment with different concentrations and FUT8 KO with 2 distinct sgRNAs were color coded as shown.

Not surprisingly, the quantitation from glycan level showed the same result with the quantitation from glycoform level. Many down- and up-regulated glycans were identified. The increased and decreased extent was more significant in FUT8 KO than in 2FF treated samples (**Figure 4.42a**). The decreased glycans were mainly from the fucosylated ones (**Figure 4.42b**), whereas, the increased glycans were mainly from sialyated ones (**Figure 4.42c**).

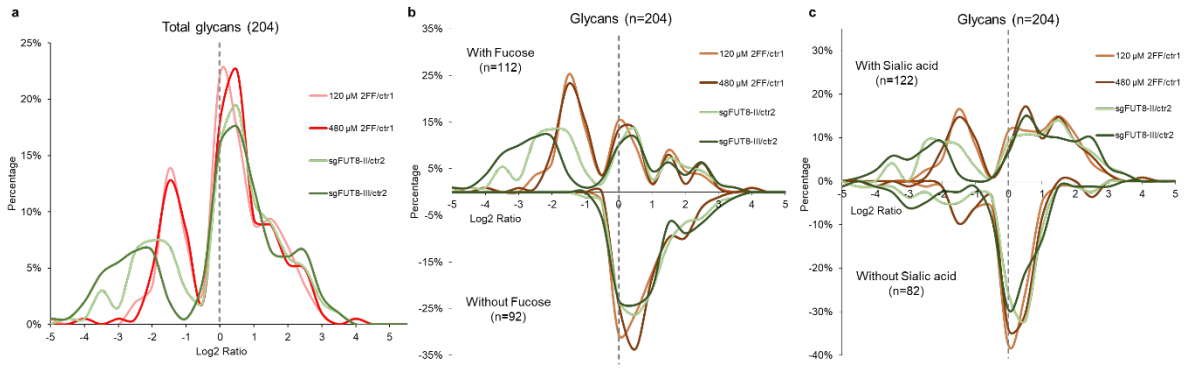


Figure 4.42 Quantitative N-glycoproteome (glycans) in 2FF-treated and FUT8 KO TMD8 cells.

(a) Ratio distribution of total glycans upon 2FF treatment and FUT8 KO. (b) Ratio distributions of fucosylated (upper panel) and non-fucosylated (bottom panel) glycans were shown separately. (c) Ratio distributions of sialyated (upper panel) and non-sialyated (bottom panel) glycans were shown separately. 2FF treatment with different concentrations and FUT8 KO with 2 distinct sgRNAs were color coded as shown.

4.4 Phosphoproteome analysis in DLBCL with BTK and SYK inhibitions

B cell receptor (BCR) is vital for B cells to perform its function. Many oncogenic mutations and chromosomal translocations in BCR and the downstream effectors sustain malignant B-cell growth and survival. Untill now, we know that BCR signaling is mainly involved in antigen-dependent NF-κB pathway in ABC DLBCL, PI3K and mTOR mediated pathway is active in GCB DLBCL. Two tyrosine kinases BTK and SYK are the important upstream effectors in BCR signaling in ABC DLBCL, whereas only SYK is vital for GCB DLBCL. The phosphorylation and dephosphorylation of effectors in BCR signaling pathway are vital for the survival of lymphomas. In this study, we investigated the change of phosphorylation of effectors in a systematic level after inhibition of upstream BTK and SYK effectors in ABC DLBCL and only SYK inhibition in GCB DLBCL using SILAC-based quantitative phosphoproteomics strategy.

4.4.1 Strategy for the phosphoproteome analysis in DLBCL

In total, 3 ABC cell lines of HBL1, TMD8 and OCI-Ly10 and 4 GCB cell lines of U2932, DOHH2, SUDHL4 and WSU-FSCCL were used to investigate the phosphoproteome expression under the BTK or SYK inhibition treatment. We used 10 nM Ibrutinib (IBR) and 250 nM PRT-060318 (PRT) were used for BTK and SYK inhibition, respectively (refers to section 3.2.1.3). A two-plex SILAC labeling strategy was used for the quantitation (**Figure 4.43**).

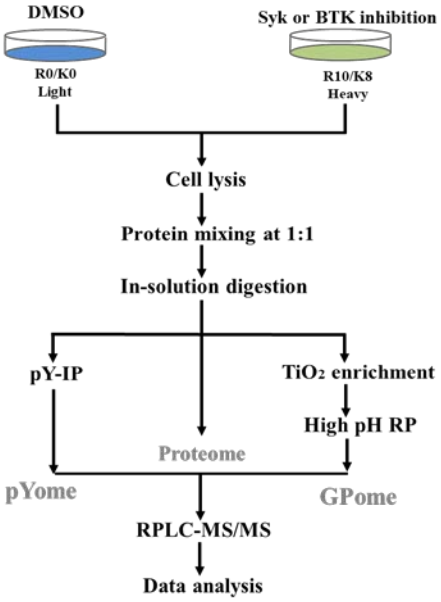


Figure 4.43 Workflow of Phosphoproteome analysis in DLBCL with BTK and SYK inhibition.

Cell lines with DMSO treatment were light labelled (K+0, R+0) as control. Cell lines with BTK or SYK inhibition were heavy labelled (K+8, R+10) for survey. Cell lysate of light and heavy labelled samples were mixed at equal protein amounts followed by in solution digestion with trypsin. Single-shot proteome analysis was performed by measuring each cell line's digested peptides in two or three injection replicates. Phosphoproteome analysis was done by two approaches: 1. unspecific phosphopeptide enrichment by titanium dioxide beads (TiO_2) followed by basic reverse phase prefractionation, which was defined as global phosphoproteome (GPome). 2. A specific antibody-based immunoaffinity purification (IP) of phosphorylated tyrosine peptides (phosphotyrosine, pY), which was noted as pYome-IP. All obtained phosphopeptides and digested peptides were then analyzed on Q Exactive HF.

4.4.2 Single-shot proteome in the phosphoproteome analysis of DLBCL

Single-shot proteome analysis showed that the majority of protein expressions in each cell line with individual inhibition were unchanged (less than 2-fold change) (**Figure 4.44**). The median of log2 transformed SILAC ratios of all quantified proteins in each sample was around zero indicating the accurate mixing of light and heavy labelled sample at equal protein amounts. All the proteins and their quantitation information in each sample can be found in **Supplementary Data S11** in the appendices.

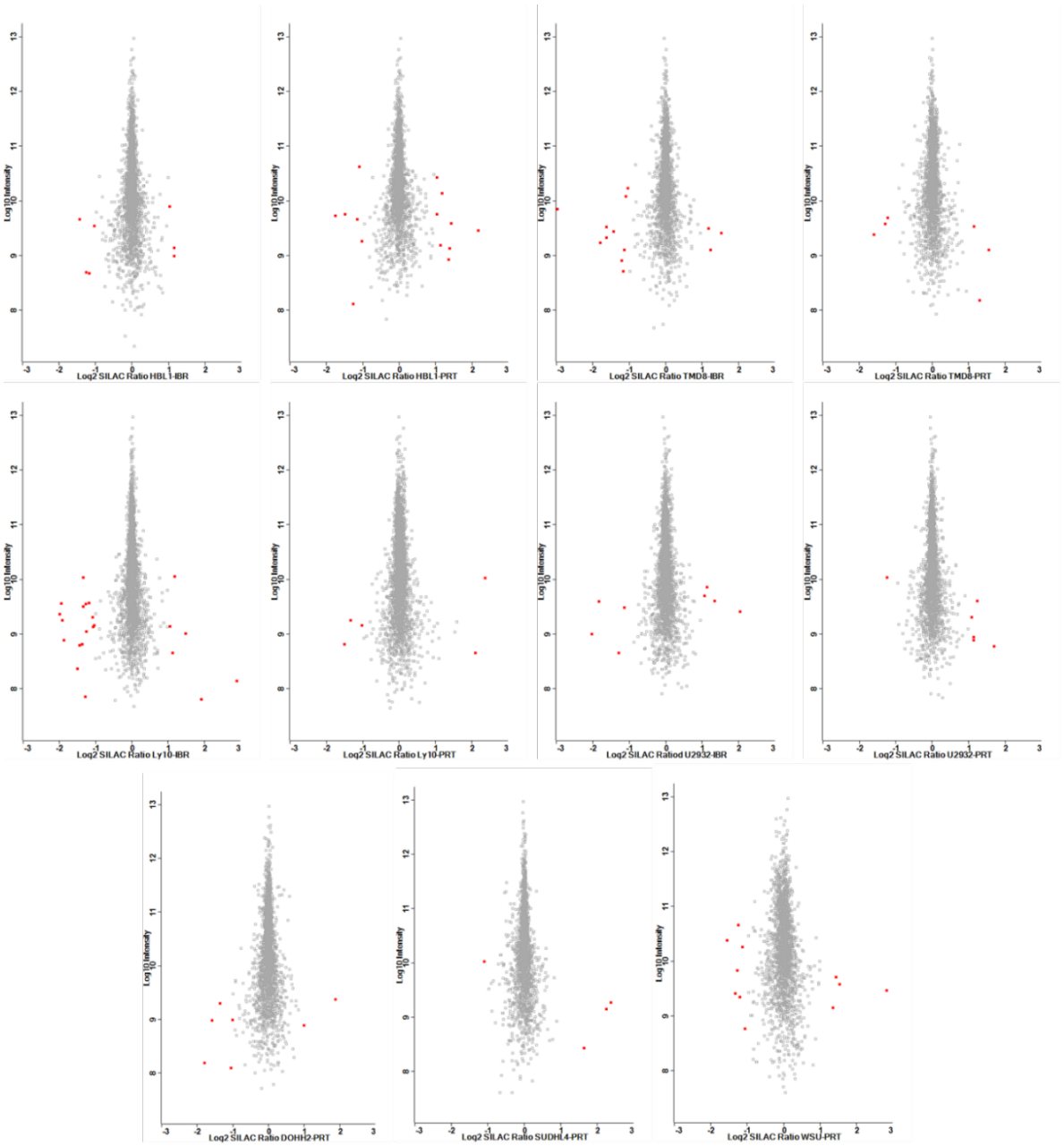


Figure 4.44 Protein expression in the phosphoproteome analysis of DLBCL with BTK or SYK inhibition.

Scatter plots show the distribution of Log2 transformed SILAC ratios of the proteins versus their intensities (Log10) in each cell line with BTK or SYK inhibition. Proteins with Log2 transformed SILAC ratios ≤ -1 or ≥ 1 were highlighted in red.

4.4.3 Statistical results of quantitative phosphoproteome in DLBCL with BTK or SYK inhibition

A total of 11 samples were surveyed, including three ABC cell lines of HBL1, TMD8 and OCI-Ly10, one GCB cell line of U2932 treated with individual BTK and SYK inhibitors, three GCB cell lines of DOHH2, SUDHL4 and WSU-FSCCL treated with SYK inhibitor. In the GPome of 11 samples, a total of 29508 unique class I phosphosites (phosphosites with localization probability of more than 0.75) were quantified in at least one sample, about 88% of them were serine phosphorylation, 11% happened on threonine residue and 1.1% were tyrosine phosphorylated peptides (**Figure 4.45a**). Because of the important role of tyrosine phosphorylation in signaling pathway and lower abundance than the other two types of phosphorylation, the antibody-based tyrosine phosphorylated peptides enrichment (pYome) was necessary to be performed. In the pYome of 11 samples, a total of 1207 unique class I phosphosites were quantified, of that more than 91% were tyrosine phosphorylation, 6% were serine phosphorylation and only 2.9 % were threonine phosphorylated peptides (**Figure 4.45b**). More importantly, 75% (907/1207) of phosphosites (p-sites) in pYome analysis were not included in the GPome, indicating the successfully complementary identification of p-sites in pYome. A total of 30405 class I p-sites were quantified in at least one of 11 surveyed samples from GPome and pYome analysis (**Table 4.2**). As far as we know, this is the largest dataset of phosphoproteome in diffuse large B-cell lymphoma.

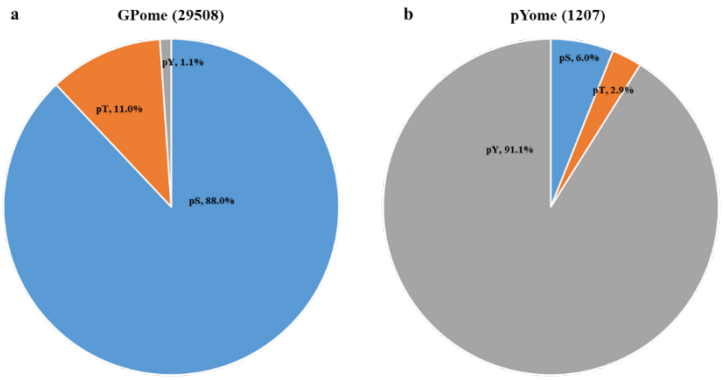


Figure 4.45 Numbers of quantified p-sites in GPome and pYome analysis of DLBCL cell lines with BTK and SYK inhibition.

The distributions of phospho-serine (pS), phospho-threonine (pT) and phospho-tyrosine (pY) in GPome (a) and pYome (b) were showed using the same color in the separate pie chart. P-sites with at least two valid ratios out of four replicates (two biological and two injection replicates) in each sample were regarded to be quantified.

The phenomena that one tryptic phosphopeptide bearing several phosphorylation modified sites is very common in organisms. The quantitation of phosphosites is based on the phosphopeptide expression. The same phosphopeptide bearing one p-site, two p-sites or more than two p-sites may be quantified together in the sample. In order to make full use of the quantitative phosphopeptides information, we analyzed the data from the p-events level. Almost 80% of p-events were the phosphopeptides with single p-site, 18% were the double p-sites modified peptides, only 2% of the quantified phosphopeptides were modified with more than two p-sites (refers to **Supplementary Data S12**). The number of quantified p-sites and p-events in each sample were listed in **Table 4.2**. All the quantified p-sites and p-events in each cell line with treatments can be found in **Supplementary Data S12** in the appendices.

Table 4.2 Summary of quantitative phosphoproteome in each DLBCL cell line

	Subtypes	Cell Line	Treatment (Inhibitor)	Unique p-sites	Unique p-events	Biological Replicates	Injection Replicates
GPome	ABC	HBL1	Ibrutinib	13680	15812	2	2
			PRT060318	13415	15455	2	2
		TMD8	Ibrutinib	13855	15995	2	2
			PRT060318	12737	14592	2	2
		Ly10	Ibrutinib	14527	16396	2	2
			PRT062607	12717	14236	2	2
	GCB	U2932	Ibrutinib	15186	17423	2	2
			PRT060318	14435	16476	2	2
		DOHH2	PRT060318	14215	16485	2	2
		SUDHL4	PRT060318	13267	14980	2	2
		WSU-FSCCL	PRT060318	11402	12597	2	2
	Total-GPome			29508	34976		2
pYome	ABC	HBL1	Ibrutinib	368	395	3	2
			PRT060318	362	376	2	2
		TMD8	Ibrutinib	521	560	2	2
			PRT060318	505	535	2	2
		Ly10	Ibrutinib	432	446	2	2
			PRT060318	509	523	2	2
	GCB	U2932	Ibrutinib	758	786	2	2
			PRT060318	259	279	2	2
		DOHH2	PRT060318	207	217	2	2
		SUDHL4	PRT060318	250	259	2	2
		WSU-FSCCL	PRT060318	189	203	2	2
	Total-pYome			1207	1278		
Total				30405	35945		

Note: 10 nM Ibrutinib was used for BTK inhibition; 250 nM PRT060318 was used for SYK inhibition.

4.4.4 The correlations of p-events in GPome and pYome analysis of DLBCL

Pearson correlation (r) and the coefficient of variation (CV) analysis were performed to investigate the reproducibility of quantified p-events among quadruplicates in each sample in GPome and pYome. In GPome, the majority of Pearson r located in the range of 0.6-0.8 and 0.8-1 (**Figure 4.46a**). The percentages of quantified p-events with CV of 4 ratios less than 20% in all 11 samples were above 78% (**Figure 4.46b**). Higher Pearson r was obtained in pYome analysis (**Figure 4.46c**). The percentage of quantified p-events with CV of 4 ratios less than 20% in all 11 samples located between 64% and 87% (**Figure 4.46d**). All the results indicated the high reproducibility among quadruplicates in the 11 samples in both GPome and pYome analysis.

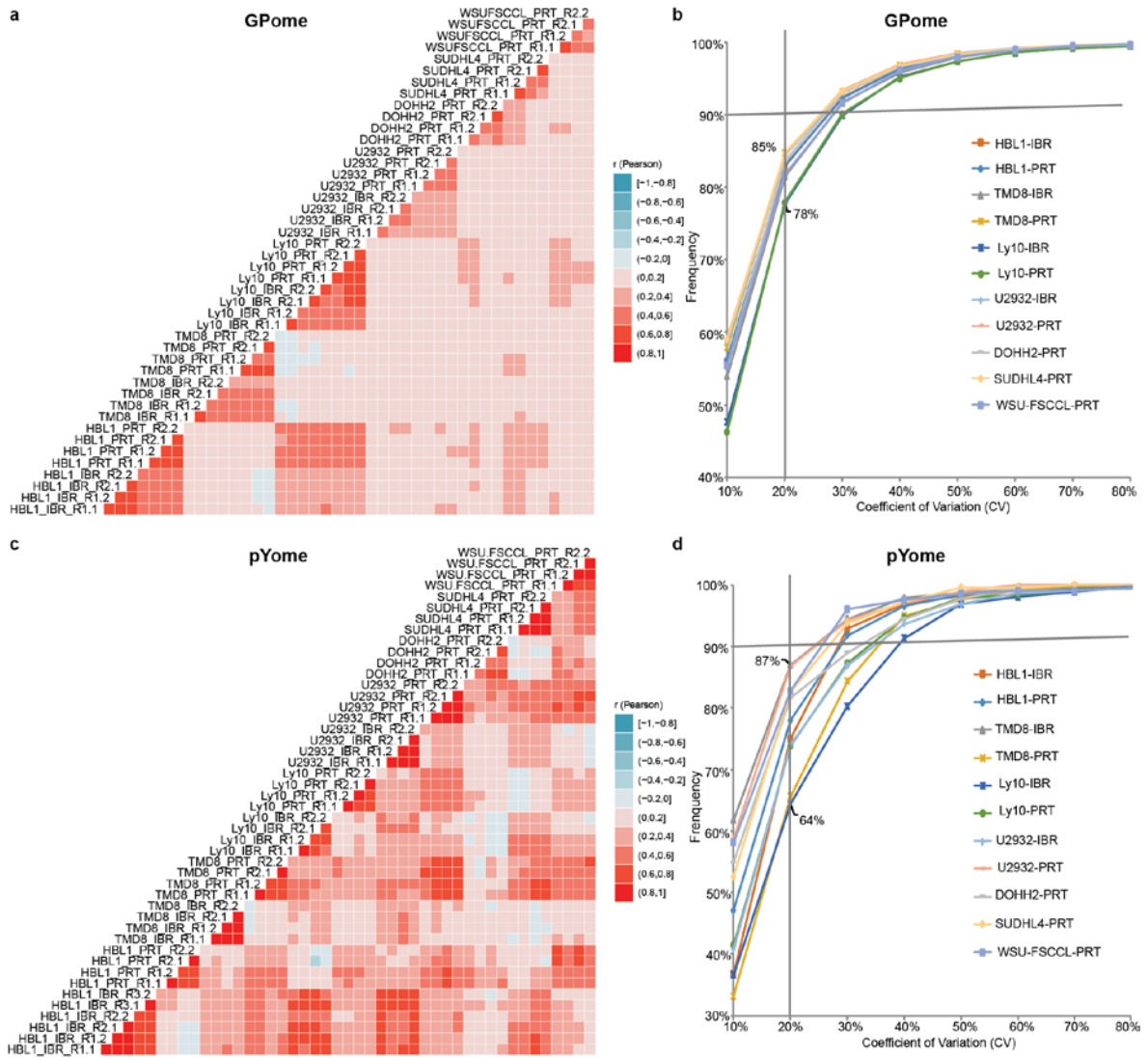


Figure 4.46 Correlations of quantified p-events in GPome and pYome analysis of DLBCL cell lines with BTK and SYK inhibition.

Pearson r of the replicates in each sample in GPome (a) and pYome (c) analysis. Two biological replicates were encoded as R1, R2, with the exception of 3 replicates for HBL1 with Ibrutinib treatment in pYome. Two injection replicates were encoded as -1, -2 suffix. Coefficient of variation (CV) profiling of ratios among 4 replicates in each sample in GPome (b) and pYome (d) analysis. The lowest and highest percentages of p-events with the CV less than 20% in the samples were marked.

4.4.5 Principal component analysis of quantified p-events in GPome analysis of DLBCL

In GPome analysis, a total of 34976 p-events were quantified within at least one sample. 2016 p-events were filtered out which were quantified in all 44 samples (11 treatments × 4 replicates) firstly. We performed the multiple-samples t-test using the Permutation-based FDR with 5% in ANOVA to keep the distinguishably expressed phosphopeptides across the 11 treatments in 3 ABC and 4 GCB cell lines with BTK or SYK inhibition, resulting in 1768 p-events for the confident comparison across all samples. From the PCA analysis, the quadruplicate measurements of each differently treated cell were clustered to each other, indicating the high reproducibility of our quantitative analysis (**Figure 4.47a**). No matter GCB or ABC cell lines, the specific cell line was the main driver for their segregation compared to IBR and PRT treatments. The phosphorylation profiling in 3 ABC cell lines of HBL1, TMD8 and Ly10 were distinguishably expressed based on the principal component 1 or component 2. However, the 4 GCB

cell lines located in the narrow zone based on principal component 1 and were segregated among each other based on principal component 2 (**Figure 4.47a**). Hierarchical clustering analysis showed the similar result with PCA. The quadruplicates in each cell line with IBR or PRT inhibition were co-clustered closely with each other locating in the primary branch of the dendrogram, proving that there was no batch effect and good sample coherency in the GPome analysis of 3 ABC and 4 GCB cell lines with IBR and PRT inhibition. The same cell line with IBR and PRT treatments were clustered into the secondary branch of the dendrogram. The ABC and GCB subtypes did not show the evident clustering among each other (**Figure 4.47b**).

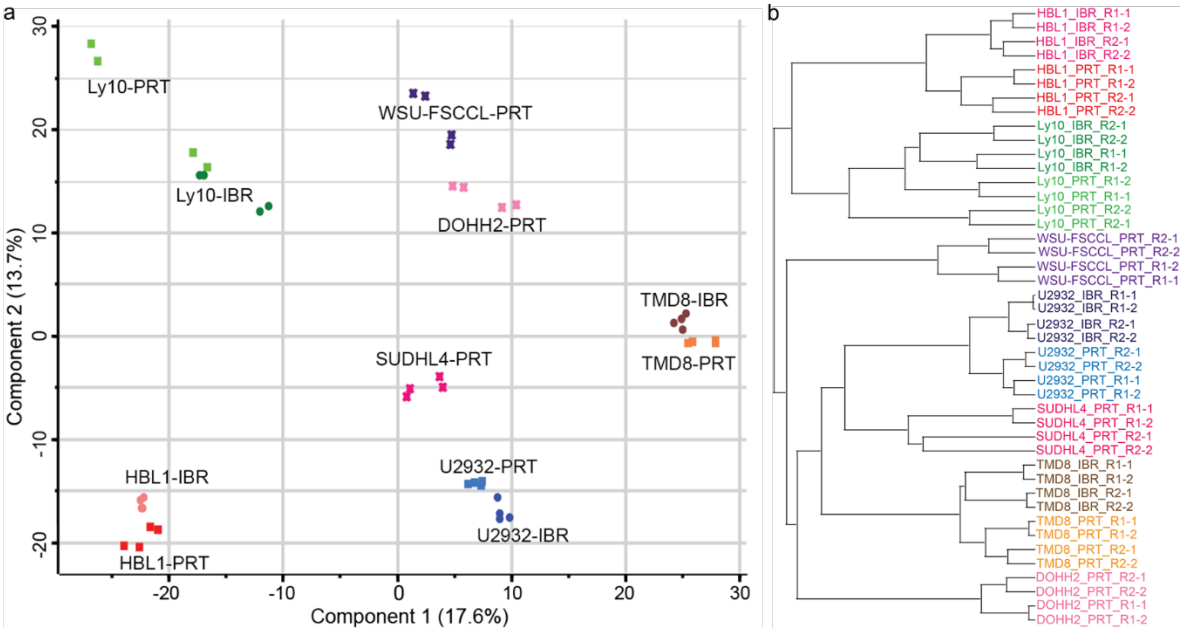


Figure 4.47 Principal component analysis and Hierarchical clustering of quantified p-events in GPome of DLBCL.

(a) PCA analysis of SILC ratios of quantified p-events across all samples. Samples were color-coded in cell line and treatment. Symbol types of ABC cell lines with Ibrutinib treatment, ABC cell lines with PRT treatment and GCB cell lines with PRT treatment were marked by filled circle, filled square and star separately. Note: filled circle for U2932 with IBR, filled square for U2932 with PRT. (b) Hierarchical clustering (Euclidian distance) of 44 samples using SILAC ratios of p-events quantified across all samples. Z-score conversion was performed for the log2 transformed SILAC ratios of each phosphopeptide quantified across all samples prior to PCA and Hierarchical clustering analysis.

4.4.6 Differential phosphopeptides analysis (DPA) in the GPome of DLBCL

Subsequently, the empirical Bayes approach, moderated t-test statistics, and a revised calling strategy were used to investigate the differential phosphopeptides (DP) expression in the used DLBCL cell lines with BTK and SYK inhibition. The differential phospho-events (DP-events) were filtered with the FDR threshold of 0.05. The numbers of down-regulated and up-regulated p-events in each cell with BTK or SYK inhibition were shown in the stacked bar plot (**Figure 4.48a)**. Much more DP-events in ABC DLBCL with either BTK or SYK inhibition than in GCB DLBCL cell lines with SYK inhibition were identified. A total of 9783 regulated p-events which were quantified in at least one sample. All the significantly regulated p-events in each sample can be found in **Supplementary Data S13** in the appendices. Then the analysis of overlapped DP-events across the 11 samples revealed that about 51% (4985/9783) were exclusively identified in any one of the 11 samples (**Figure 4.48b)**. Near 23.2%

(2266/9783) were overlapped identifications in random combinations of 2 of the 11 samples. Only 2 and 1 DP-events were in 10 of 11 samples and all the 11 samples separately.

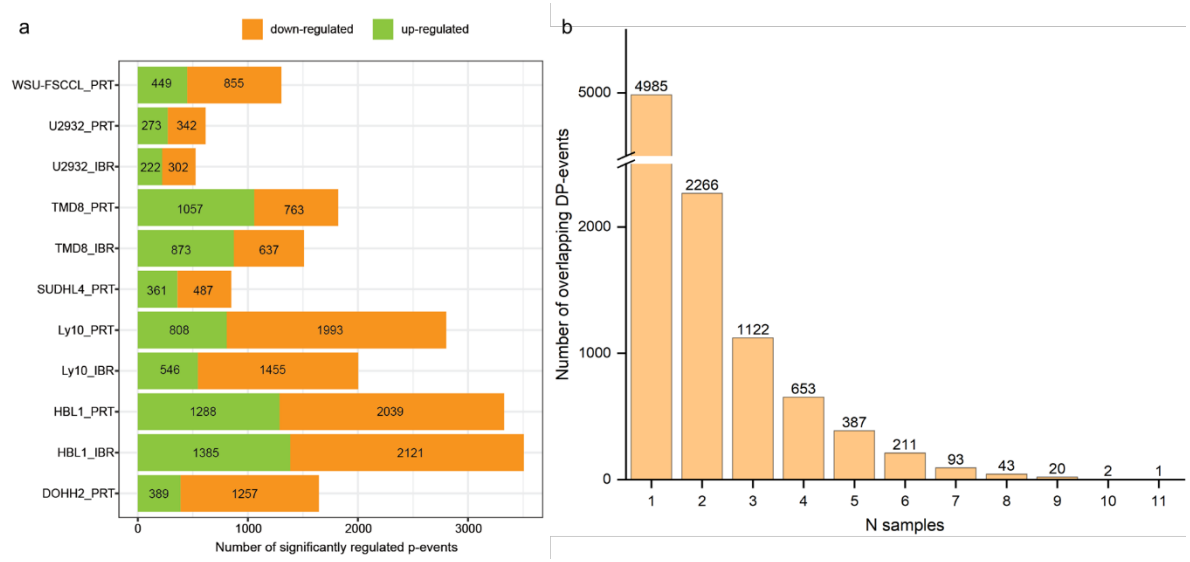


Figure 4.48 Differentially regulated p-events in the GPome analysis of ABC and GCB DLBCL cell lines.

(a) The number of down- (in deep blond) and up-regulated (in green) p-events in 11 samples. (b) Overlapping DP-events across 11 samples.

In addition, we performed the overlap analysis of DP-events for various selected conditions (**Figure 4.49**). DP-events exclusively identified in one sample took up the most percentages (**Figure 4.49a**). The most DP-events in common between two selected samples were in HBL1 with IBR and PRT treatments. The overlapped DP-events in more than two selected samples were less than 100 in most of these cases. The overlap analysis of DP-events among 3 ABC cell lines of HBL1, TMD8 and Ly10 with BTK or SYK inhibition revealed that DP-events in only one sample were the most, the overlapped ones were much fewer (**Figure 4.49b, c**). However, overlap analysis between PRT and IBR treatment in each of the 3 ABC cell lines, the overlapped DP-events in two conditions were similar numbers with those identified in only one sample (**Figure 4.49d, e, f**). 318 overlapped DP-events in TMD8 with IBR and PRT treatments and 301 overlapped DP-events in Ly10 with IBR and PRT treatments were obtained (**Figure 4.49 g**). The overlap analysis of DP-events showed that the regulation of phosphorylation in DLBCL was cell-type-specific. The BTK and SYK inhibition in the same cell displayed the similar numbers of changed phosphopeptides.

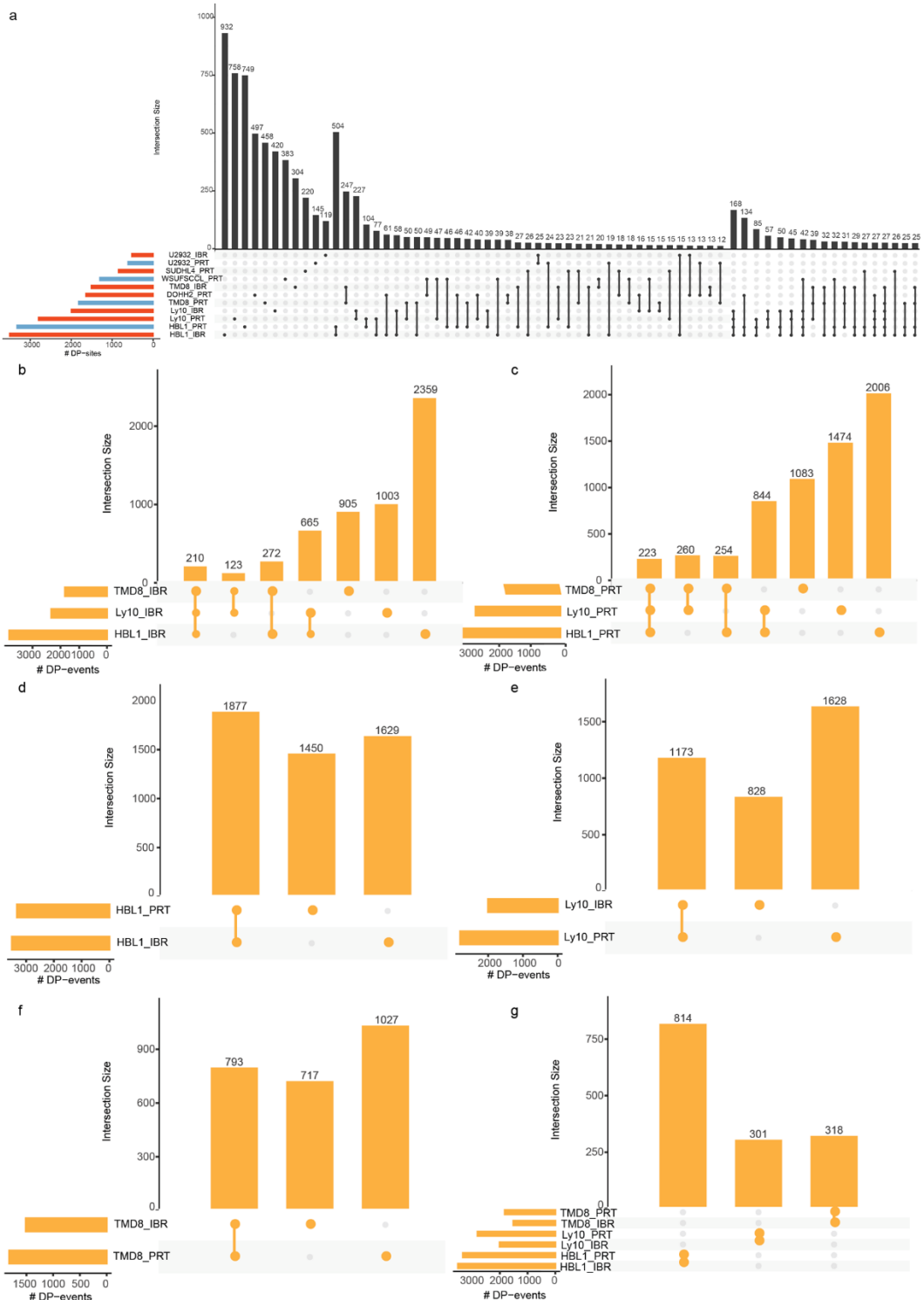


Figure 4.49 DP-events overlap analysis for selected conditions among 11 used samples in GPome.

(a) An overview of the overlap of DP-events in different samples. DP-events in only one condition, in common two or more than two conditions (not all the situations included) were shown. (b) Overlap analysis of DP-events in HBL1, TMD8 and Ly10 cell lines with Ibrutinib treatment. (c) Overlap analysis of DP-events in HBL1, TMD8 and Ly10 cell lines with PRT treatment. (d) Overlap analysis of DP-events in HBL1 with IBR and PRT treatments. (e) Overlap analysis of DP-events in Ly10 with Ibrutinib and PRT treatments. (f) Overlap analysis of DP-events in TMD8 with Ibrutinib and PRT treatments. (g) Overlap analysis of DP-events in HBL1, TMD8 and Ly10 with Ibrutinib and PRT treatments. The bar plot on the left showed the number of significant (FDR ≤ 5%) DP-events for each sample. The top bar plot showed the number of DP-events that are exclusively in common between the selected conditions (black circles in a, yellow circles in b-g). Selected combinations were highlighted.

Subsequently, the same signaling responses (Pearson r located in the range of 0.875 to 0.932) were obtained in the same cell line with the treatments of BTK and SYK inhibition (**Figure 4.50**). The responses in TMD8 with the IBR and PRT treatments were totally different with those in other cell lines with treatments, except in SUDHL4. The more similar responses were obtained in DOHH2 with PRT treatment with the other used GCB cell lines than with the used ABC cell lines with IBR or PRT treatment. The overall worse signaling responses between WSU-FSCCL with PRT and all the other samples, especially compared to the U2932 cell lines.

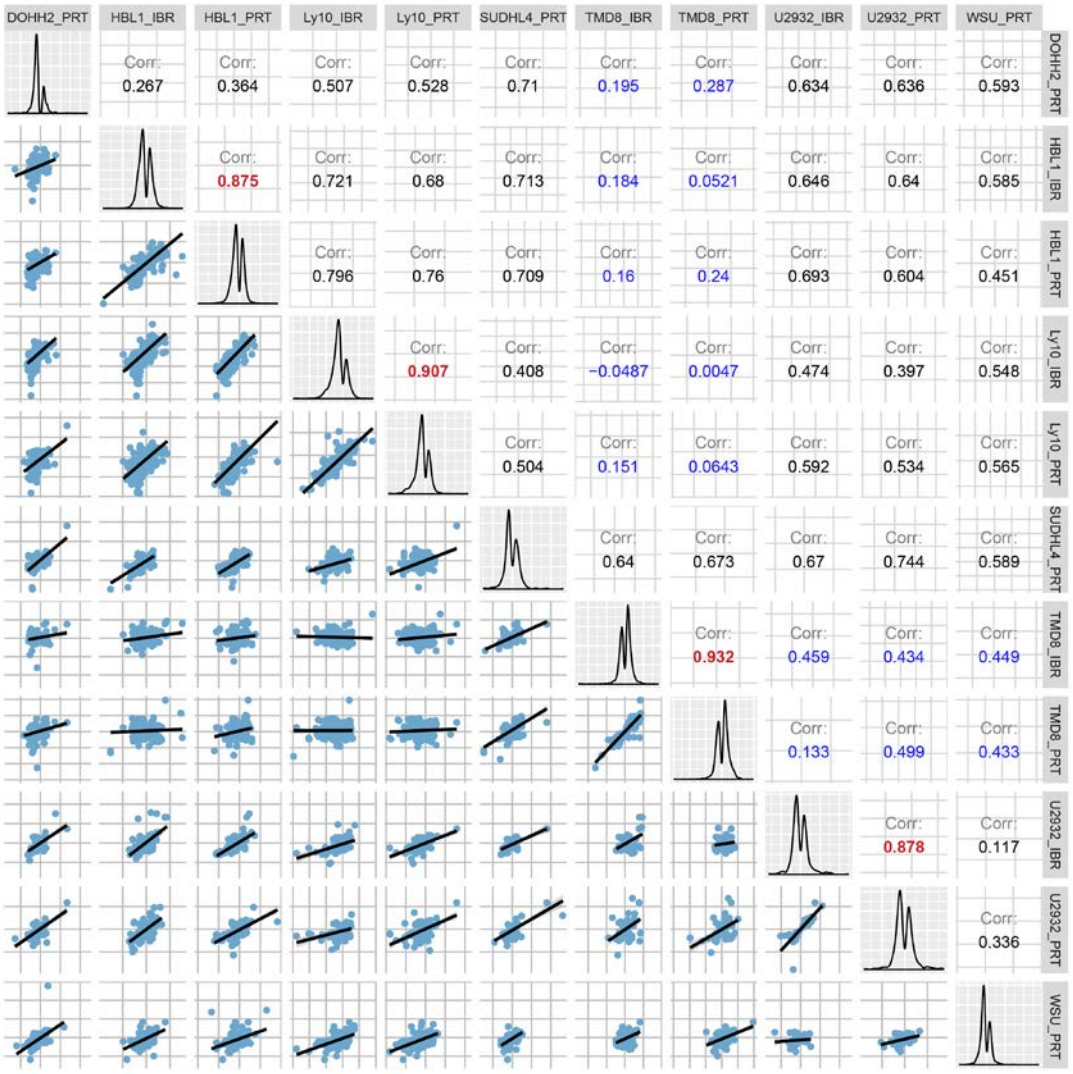


Figure 4.50 Overview of signaling perturbations in GPome analysis of DLBCL.

Common DP-events across two samples were only used for the plot. Scatter plots and Pearson correlations were shown separately for each pair plot of DP-events. The correlations of overlapped DP-events in HBL1, TMD8, Ly10 and U2932 with IBR and PRT treatments were marked in red. The correlations of overlapped DP-events between TMD8 with IBR treatment and other conditions, between TMD8 with PRT treatment and other conditions were marked in blue. “WSU” refers to WSU-FSCCL. “Corr” refers to Pearson correlation.

4.4.7 Pathway enrichment analysis of DP-events in the GPome of DLBCL

In order to understand the role of proteins which bore the altered phosphorylated peptides in used DLBCL cell lines with BTK or SYK inhibition, DP-events among each of 11 samples were used for Reactome enrichment analysis. These pathways with the Benjamini-Hochberg adjusted p-value less than 1% were shown (**Figure 4.51**).

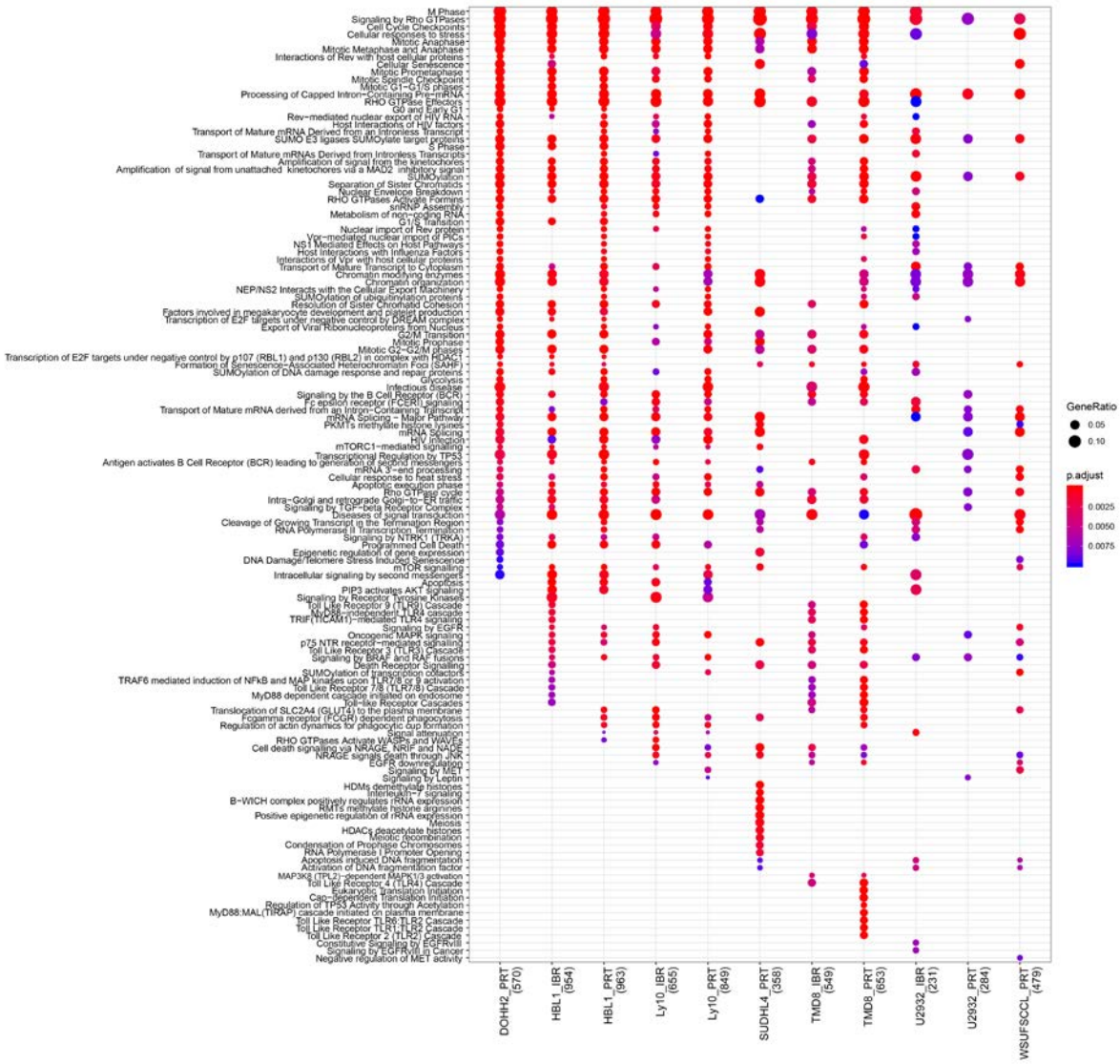


Figure 4.51 Reactome Pathway enrichment analysis of DP-events in the GPome of DLBCL with BTK and SYK inhibition.

The used numbers of DP-events in each sample were marked. The adjusted p-value for each pathway was noted with color-bar. The ratios of DP-events in each pathway compared to the total in each sample were shown with the size of filled circles.

Multiple pathways were involved by these phosphoproteins across all the 11 samples. Several pathways including M phase, signaling by Rho GTPases, diseases of signal transduction, were enriched in almost all samples. In addition, B cell receptor signaling was enriched in 8 samples, except in SUDHL4 with PRT, U2932 with IBR and WSU-FSCCL with PRT. The similar pathways were enriched between BTK and SYK inhibition in each of 3 ABC cell lines. Many of the pathways were highly enriched in only one or less than 4 samples. Toll-like receptors (TLR7/8/9) cascade and Myd88 dependent cascade initiated on endosome were enriched in TMD8 cell with either BTK or SYK inhibition.

Figure 4.52 Gene Ontology (GO) enrichment analysis of DP-events in the GPome of DLBCL with BTK and SYK inhibition.

The used numbers of DP-events in each sample were marked. The adjusted p-value for each pathway was noted with color-bar. The ratios of DP-events in each pathway compared to the total in each sample were shown with the size of filled circles.

In addition, Gene Ontology (GO) enrichment analysis revealed that these phosphoproteins with changed phosphorylated peptides after BTK or SYK inhibition in each sample were involved in the majority of the same biological processes (**Figure 4.52**). They were mainly related with the gene regulation process, such as chromosome organization and medication, DNA metabolic process, nuclear export. A series of RNA regulation processes were also highly enriched in all the samples, like RNA transport, localization, mRNA catabolic process. GTPases mediated signal transduction and the regulation of GTPases activity were influenced. Importantly, B cell differentiation was enriched in the IBR and PRT treated samples.

4.4.8 Differential phosphorylation analysis (DPA) in the pYome of DLBCL

In pYome analysis, a total of 1207 p-events were quantified within at least one sample. Log_2 SILAC ratio with ≥0.5 and ≤-0.5 were used as the threshold for the significantly regulated p-events in pYome analysis. The numbers of down-regulated and up-regulated p-events in each cell with BTK or SYK inhibition in pYome were shown in the stacked bar plot (**Figure 4.53a)**. All the significantly regulated p-events in each sample in pYome analysis can be found in **Supplementary Data S14** in the appendices.

Then we surveyed the overlapped DP-events across the 11 samples, which revealed that about 56.4% (253/448) were exclusively identified in only one of the 11 samples (**Figure 4.53b)**. Near 20.1% (90/448) were overlapped identifications in random combinations of 2 of the 11 samples. Only 6 DP-events were commonly identified in more than 6 of 11 samples. From the distribution of all quantified p-events in each sample, the mean of all ratios of p-events was close to zero (log2 transformed) in the majority of 23 samples (**Figure 4.53c, d**). Moreover, we performed the overlap analysis of DP-events for various selected conditions (**Figure 4.54**). The majority of DP-events were exclusively identified in one sample, eg, 118 p-events in U2932 with IBR treatment.

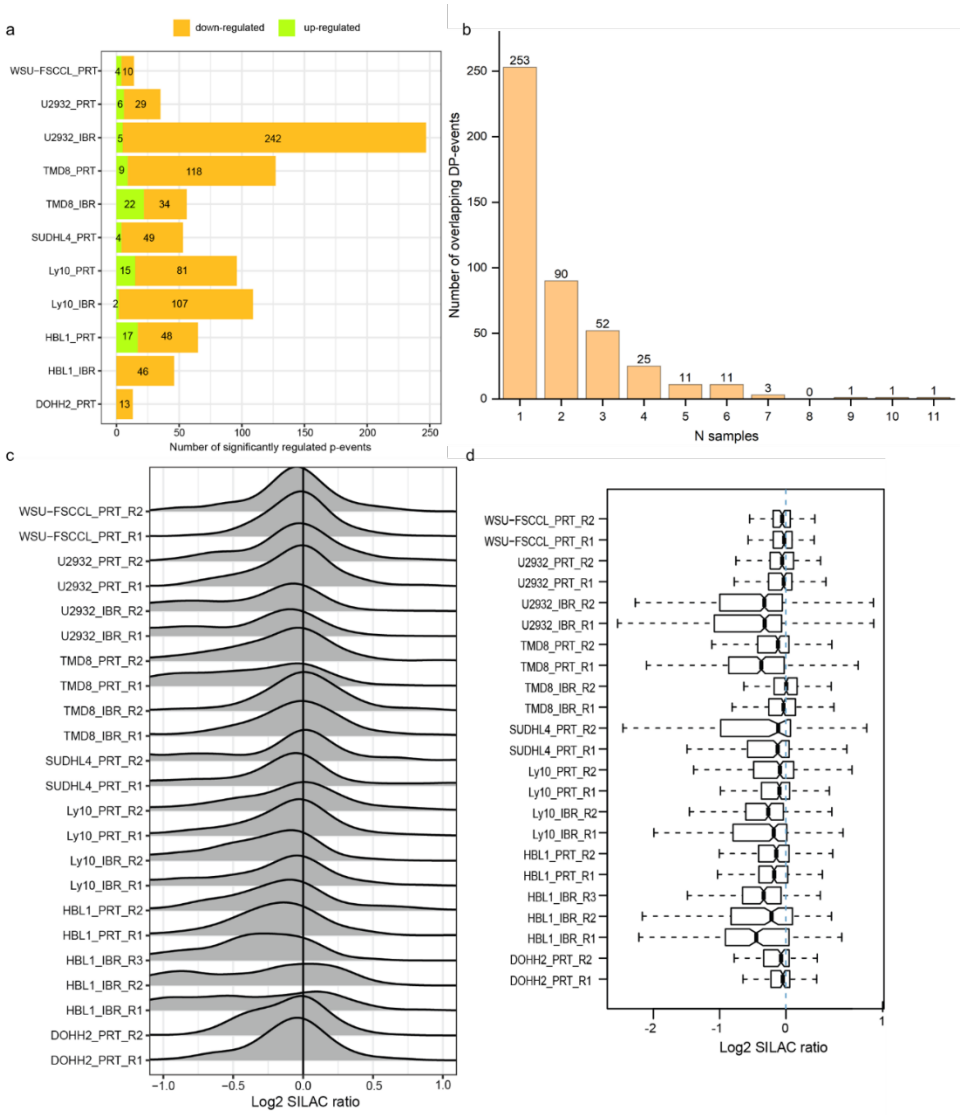


Figure 4.53 Differentially regulated p-events in the pYome analysis of used ABC and GCB cell lines with BTK or SYK inhibition.

(a) The number of down- (in deep blond) and up-regulated (in green) p-events in 11 samples. (b) Overlapping DP-events across 11 samples. (c) Distributions of quantified p-events in pYome analysis of DLBCL. (d) Box plots show the distribution of quantified p-events in pYome analysis of DLBCL. The average SILAC ratio from two injection replicates was calculated for the biological replicate for each p-event in each sample.

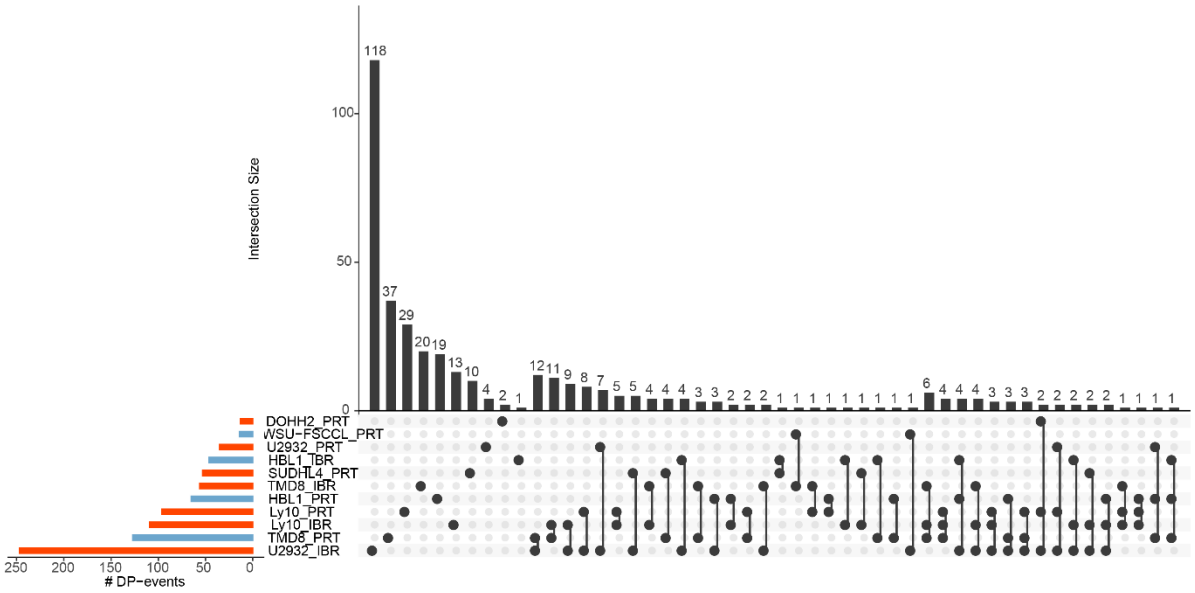


Figure 4.54 DP-events overlap analysis for selected conditions among 11 used samples in pYome.

An overview of the overlap of DP-events in different samples. DP-events in only one condition, in common two or more than two conditions (not all the situations included) were shown. The bar plot on the left showed the number of significant (FDR ≤ 5%) DP-events for each sample. The top bar plot showed the number of DP-events that are exclusively in common between the selected conditions (black circles). Selected combinations were highlighted.

4.4.9 Pathway enrichment analysis of DP-events in the pYome of DLBCL

Similarly, Reactome pathway and GO enrichment analysis of phosphoproteins which bore the altered phosphopeptides in the pYome of DLBCL with BTK and SYK inhibition were performed. The Benjamini-Hochberg adjusted p-values with the threshold of 1% were set. The majority of these pathways were involved in receptor-mediated signaling **(Figure 4.55a**). For instance, signaling by Interleukins and Fc-gamma receptor (FCGR) dependent phagocytosis were enriched in 10 of 11 samples. Many pathways were exclusively enriched in TMD8 with PRT treatment, such as, PKMTs methylate histone lysine, Nonhomologous End-joining (NHEJ), SUMOylation of chromatin organization proteins. More importantly, the reactome pathways “signaling by the B cell receptor”, “antigen activates B cell receptor leading to generation of second messengers” and “signaling by receptor tyrosine kinases” were enriched in all samples, except in TMD8 with IBR.

Regarding the involved biological processes (**Figure 4.55b**), the peptidyl-tyrosine phosphorylation were involved due to the enrichment of tyrosine-phosphopeptides. The important processes, such as B cell receptor signaling pathway, B cell activation and B cell differentiation, were high enriched in almost all samples. In addition, Few DNA replication-related processes such as telomere capping, chromatin silencing at rDNA, DNA replication-dependent nucleosome organization/assembly were highly enriched in TMD8 with PRT treatment.

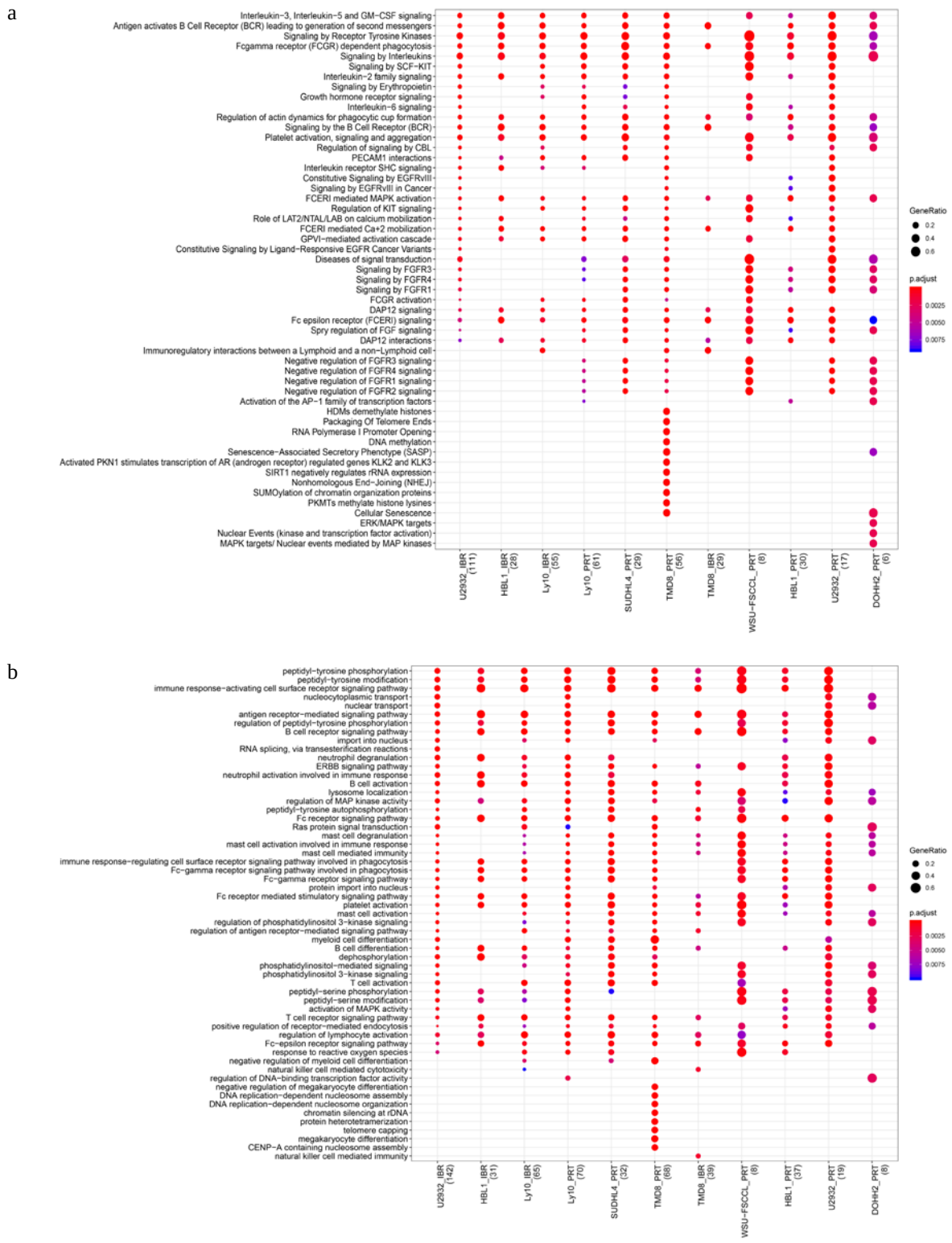


Figure 4.55 Reactome Pathway and GO enrichment analysis of DP-events in the pYome of DLBCL with BTK and SYK inhibition.

The used numbers of DP-events in each sample were marked. The adjusted p-value for each pathway was noted with color-bar. The ratios of DP-events in each pathway compared to the total in each sample were shown with the size of filled circles.

4.5 Ubiquitinome analysis in DLBCL with BTK and SYK inhibitions

We used the di-Glycine-lysine-specific antibody (K-ε-GG) to enrich the ubiquitinated peptides in combination with SILAC technology to discover the altered ubiquitylated peptides in DLBCL cell lines after BTK and SYK inhibitions. It enabled us to generate quantitative ubiquitin-modified proteome (Ubiquitinome) dataset in the subtypes of DLBCL cell lines with the treatments of two tyrosine kinase inhibitors (Ibrutinib for BTK inhibition, PRT-062607 for SYK inhibition).

4.5.1 Strategy for the ubiquitinome analysis in DLBCL

In ubiquitinome analysis, we investigated the ubiquitinome expression in two ABC cell lines of HBL1, TMD8 and one GCB cell line of U2932 with BTK and SYK inhibition using a two-plex SILAC labeling strategy. 10 nM Ibrutinib (IBR) and 250 nM PRT-062607 (PRT) were used for BTK and SYK inhibition separately (**Figure 4.56**).

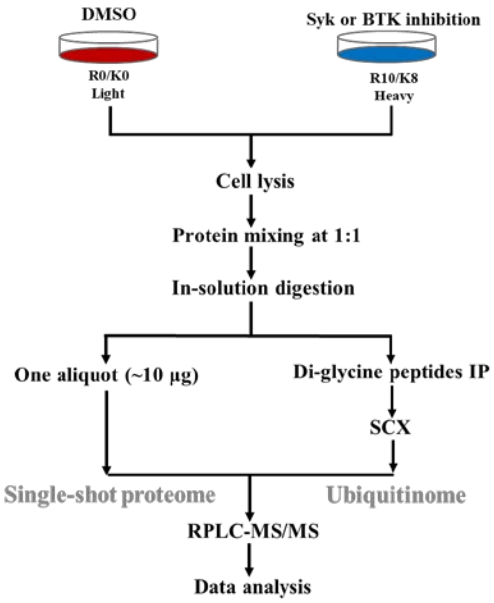


Figure 4.56 Workflow of Ubiquitinome analysis in DLBCL with BTK and SYK inhibition.

Heavy labelled (K+8, R+10) cell lines were treated with IBR or PRT. Light labelled (K+0, R+0) cells were treated with same amount of DMSO as control. We combined the equal protein amounts of light and heavy labelled samples followed by in solution digestion with trypsin. One aliquot of digested peptides from the mixed labelled samples was used for Single-shot proteome analysis in two injection replicates for each biological replicate after desalting. The leftovers were used for di-glycine peptides enrichment by Ubiquitin Remnant Motif (K-ε-GG) Kit (CST). Then strong cation exchange (SCX) was performed to fractionate the enriched peptides. The resulting digesting peptides and 3 fractions of enriched ubiquitinated peptides were analyzed on Q Exactive HF.

4.5.2 Single-shot proteome in the ubiquitinome analysis of DLBCL

The single-shot proteome result showed that the majority of proteins in HBL1, TMD8 and U2932 with IBR and PRT treatments were unchanged (**Figure 4.57**). The accurate mixing of light and heavy labelled sample at equal protein amounts was validated based on that the median of log2 transformed SILAC ratios of all quantified proteins in each sample was around zero. All the proteins and their quantitation information in each sample can be found in **Supplementary Data S15** in the appendices.

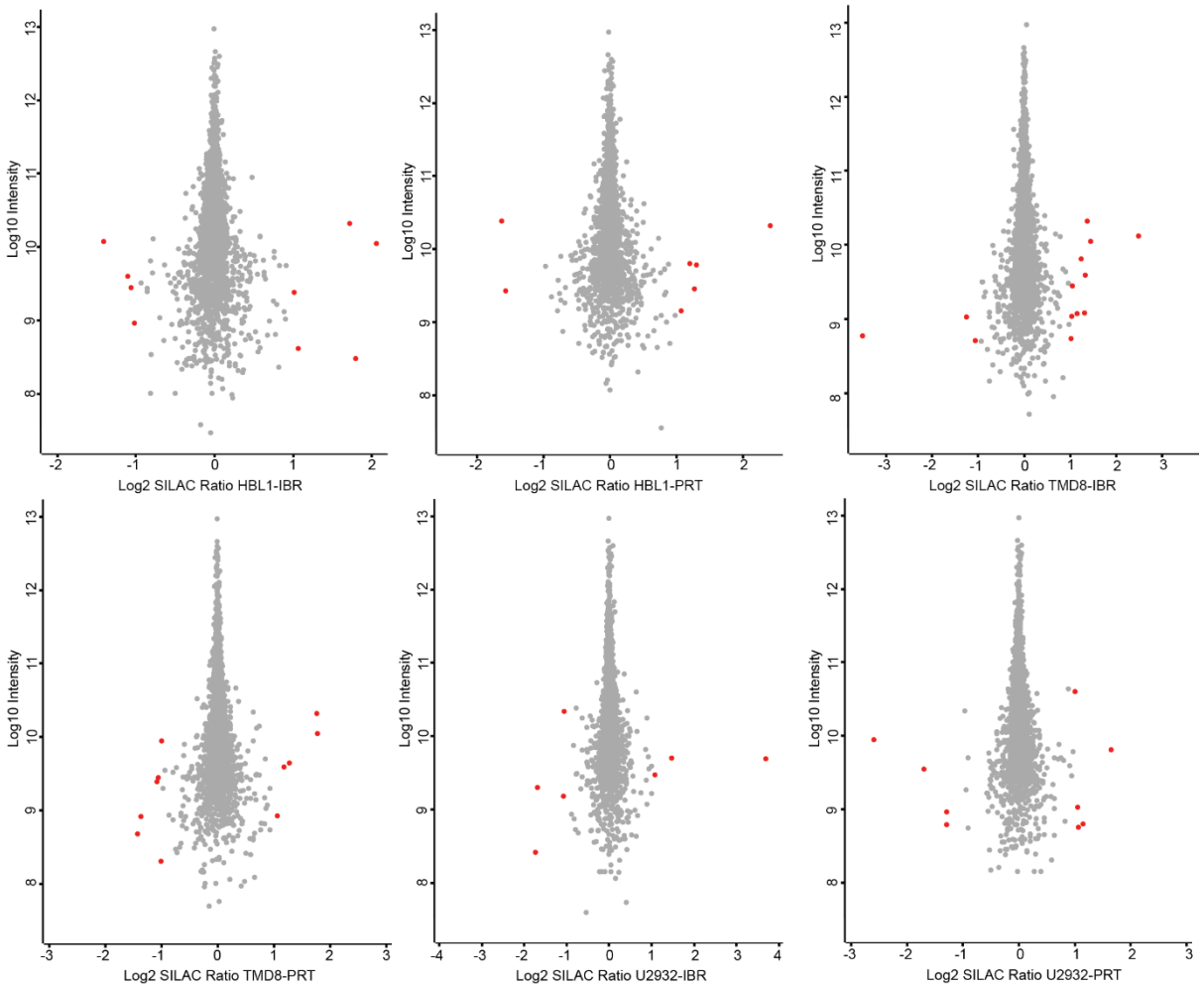


Figure 4.57 Protein expression in the ubiquitinome analysis of DLBCL with BTK or SYK inhibition.

Scatter plots showed the distribution of Log2 transformed SILAC ratios of the proteins versus their intensities (Log_{10}) in each cell line with BTK or SYK inhibition. Proteins with Log_2 transformed SILAC ratios ≤ -1 or ≥ 1 were highlighted in red.

4.5.3 Statistical results of quantitative ubiquitinome in DLBCL with BTK or SYK inhibition

A total of 14992 unique class I ubiquitylation sites (ubi-sites) with the localization probability of more than 0.75 on 3961 proteins were quantified in at least one of six samples, which included three cell lines of HBL1, TMD8, U2932 with separate IBR and PRT treatment (**Table 4.3**). To my knowledge, this is the largest dataset of ubiquitinome in diffuse large B-cell lymphoma. Multi-ubiquitination also exist on the same tryptic ubiquitylated peptides, despite the relatively lower frequency compared to multi-phosphorylation. Almost 97% of ubi-events were the ubiquitylated peptides with single ubi-site, 2.5% were the double ubi-sites modified peptides, less than 0.5% of the quantified ubiquitylated peptides were modified with more than two ubi-sites (refers to **Supplementary Data S16**). The number of quantified unique ubi-sites and ubi-events in each sample were listed in **Table 4.3**. The ubi-events with at least two valid ratios among replicates were regarded to be quantified in each sample. All the quantified ubi-sites and ubi-events in each cell line with IBR and PRT treatments can be found in **Supplementary Data S16** in the appendices.

Table 4.3 Summary of quantitative ubiquitinome results in each DLBCL cell line

Cell Line	Treatment (Inhibitor)	Unique Ubi-sites	Unique Ubi-events	Biological Replicates	Injection Replicates
HBL1	Ibrutinib	7777	7935	2	2
	PRT062607	5723	5804	2	2
TMD8	Ibrutinib	10740	10911	3	2
	PRT062607	8099	8259	3	2
U2932	Ibrutinib	5764	5839	2	2
	PRT062607	3899	3953	2	2
	Total	14992	15311		

Note: 10 nM Ibrutinib was used for BTK inhibition; 250 nM PRT062607 was used for SYK inhibition.

4.5.4 The correlations of ubi-events in ubiquitinome analysis of DLBCL

Pearson correlation r of quantified ubi-events between four or six replicates in each sample were in the range of 0.4-0.6 and 0.6-0.8 (**Figure 4.58a**). The percentage of quantified ubi-events with CV less than 20% in all 6 samples located between 80% and 89%. The Pearson correlation and CV analysis indicated that the high reproducibility of quantified ubi-events among biological and injection replicates in 6 samples.

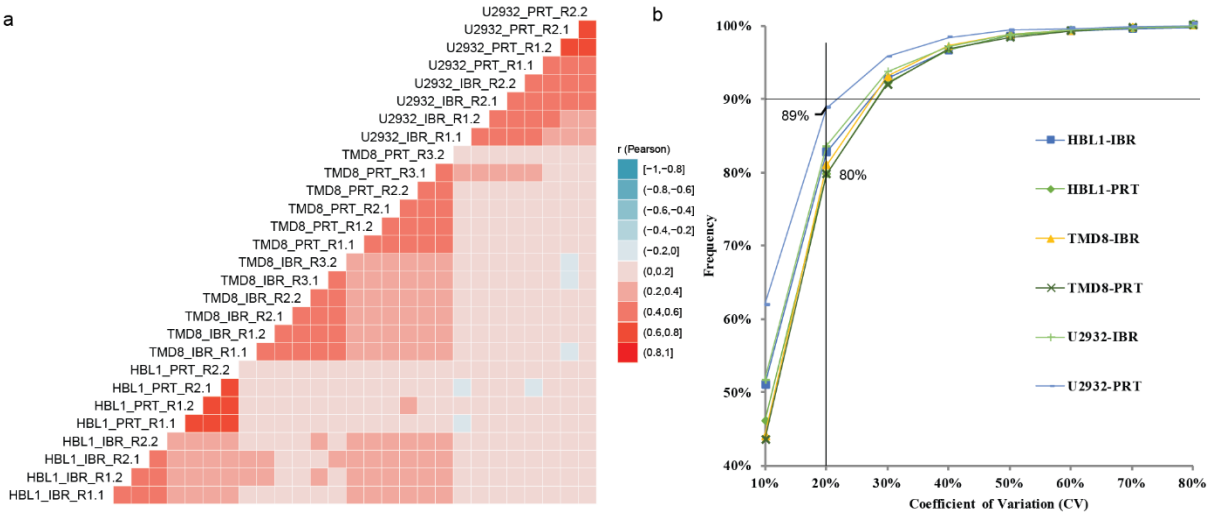


Figure 4.58 Correlations of quantified ubi-events in ubiquitinome analysis of DLBCL cell lines with BTK and SYK inhibition.

(a) Pearson r of pairwise quantified ubi-events in replicates was computed using the $\log_2$ SILAC ratios. Two biological replicates were encoded as R1, R2, with the exception of 3 replicates for TMD8 with IBR and PRT treatments. Two injection replicates were encoded as -1, -2 suffix. (b) Coefficient of variation (CV) profiling of ratios of quantified ubi-events among 4 or 6 replicates in each sample. The lowest and highest percentages of ubi-events among total quantified ubi-events with CV less than 20% in the samples were marked.

4.5.5 Principal component analysis of quantified ubi-events in ubiquitinome of DLBCL

In ubiquitinome analysis, a total of 15311 ubi-events were quantified within at least one sample. 887 ubi-events quantified in all 28 samples (four replicates in HBL1 and U2932, six replicates in TMD8) were filtered out firstly. We performed the multiple-samples t-test using the Permutation-based FDR with 5% in ANOVA to choose the differentially expressed ubiquitylated peptides across the 28 samples, leading to 766 ubi-events for the confident comparison. PCA analysis revealed that 2 ABC cell lines of HBL1 and TMD8 were separated into the different zones with 1 GCB cell line of U2932 based on the component 1 (25.9%) (**Figure 4.59a**). The IBR and PRT treatments in U2932 didn't induce the distinct ubiquitylation profilings based on either component 1 or component 2 (18.7%). Whereas, the ubiquitylation level in HBL1 with PRT treatment was different with that in HBL1 with IBR treatment from their location in PCA. The variance of ubiquitylation also existed between TMD8 with PRT and IBR treatments depending on component 2. Compared with the cell-type specific concordance in phosphorylation regulation by IBR and PRT, different ubiquitylation regulations were found in the same ABC cell line under the IBR and PRT treatment. Hierarchical clustering analysis supported the result from PCA analysis. The replicate measurements were co-clustered closely with each other, which located in the 1st branch of the dendrogram (**Figure 4.59b**).

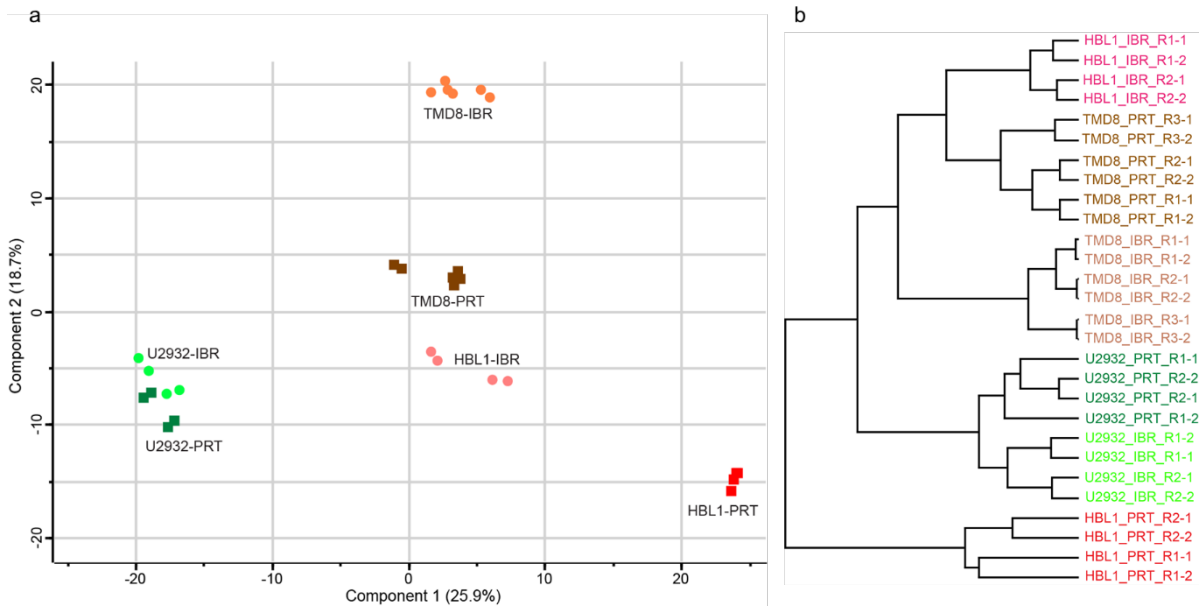


Figure 4.59 Principal component analysis and hierarchical clustering of quantified ubi-events in DLBCL.

(a) PCA analysis of SILC ratios of quantified ubi-events across all samples. Samples were color-coded in cell line and treatment. Symbol types of ABC and GCB cell lines with Ibrutinib and PRT treatment were marked by filled circle, filled square separately. (b) Hierarchical clustering (Euclidian distance) of 28 samples using SILAC ratios of ubi-events quantified across all samples. Z-score conversion was performed for the log2 transformed SILAC ratios of each ubi-event across all samples prior to PCA and hierarchical clustering analysis.

4.5.6 Differential ubiquitylated peptide analysis (DUB) in the ubiquitinome of DLBCL

The differential ubiquitylated peptide expression was surveyed using the empirical Bayes approach, moderated t-test statistics and a revised calling strategy with the FDR less than 0.05. The numbers of down- and up-regulated DUB-events in each cell with IBR or PRT treatment were listed in the stacked bar plot (**Figure 4.60a**). A total of 4208 DUB-events were quantified in at least one of six used samples. All the significantly regulated ubi-events in each sample can be found in **Supplementary Data S17** in

the appendices. The most DUB-events were found in HBL1 with PRT treatment. The down-regulated ubi-events were about two-fold of the up-regulated ones in HBL1 with PRT. However, in the other 5 samples, more up-regulated ubi-events were identified. Then we investigated the overlapped DUB-events among the 6 samples (**Figure 4.60b**). About 70% (2955/4208) were identified in only one sample. Almost of 21% (867/4208) were found commonly in two of six samples. Less than 3.1% DUB-events were common in four or more than four samples. Subsequently, we performed the overlap analysis of DUB-events for various selected conditions (**Figure 4.60c**). The majority of DUB-events exclusively identified in one sample. The common DUB-events in two conditions were the most in TMD8 between IBR and PRT. The overlapped DUB-events in more than two selected samples were less than 50.

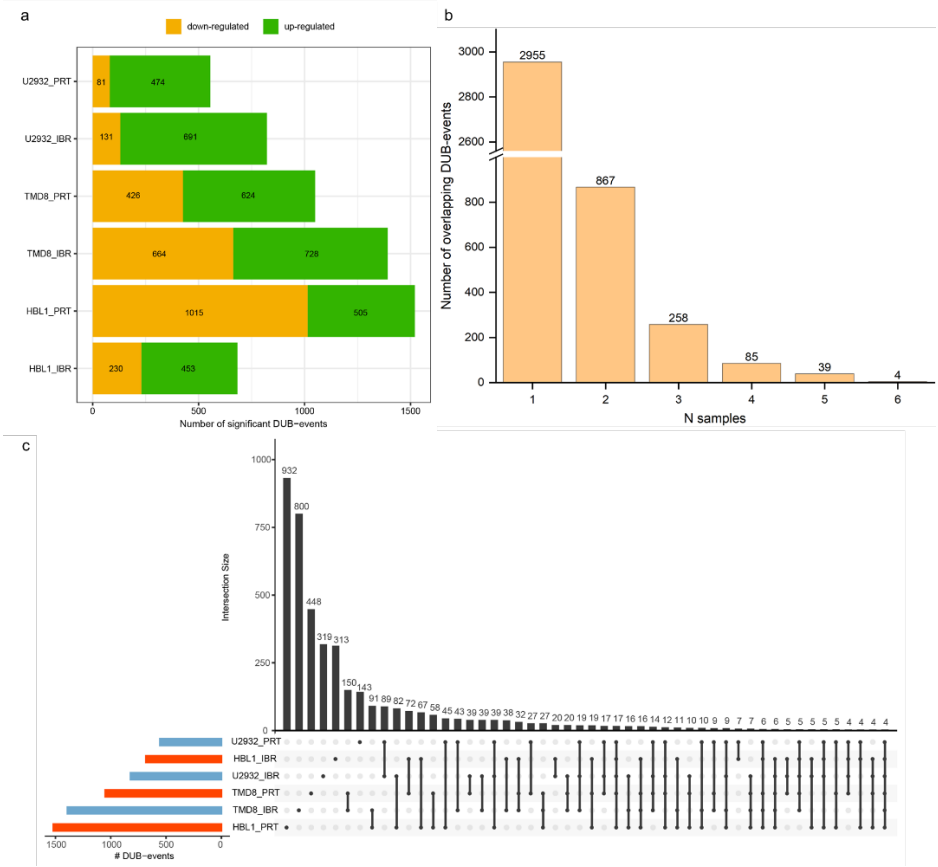


Figure 4.60 Differentially regulated ubi-events in the ubiquitinome of DLBCL.

(a) The number of down- (in deep yellow) and up-regulated (in green) ubi-events in 6 samples separately. (b) Overlapping DUB-events across 6 samples. (c) An overview of the overlap of DUB-events in different samples. DUB-events in only one condition, in common two or more than two conditions were shown. The bar plot on the left showed the number of significant (FDR ≤ 5%) DUB-events for each sample. The top bar plot showed the number of DUB-events that are exclusively in common between the selected conditions. Selected combinations were highlighted using the lines.

The correlations of common DUB-events across two random conditions were surveyed. The same signaling responses were obtained in the same cell line with two distinct treatments and their Pearson r located in the range of 0.8 to 0.9 (**Figure 4.61**). In addition, the similar signaling responses were found between TMD8 with PRT and HBL1 with IBR treatment. The responses in GCB cell line of U2932 with PRT treatment were totally different with those in 2 ABC cell lines with IBR or PRT treatment.

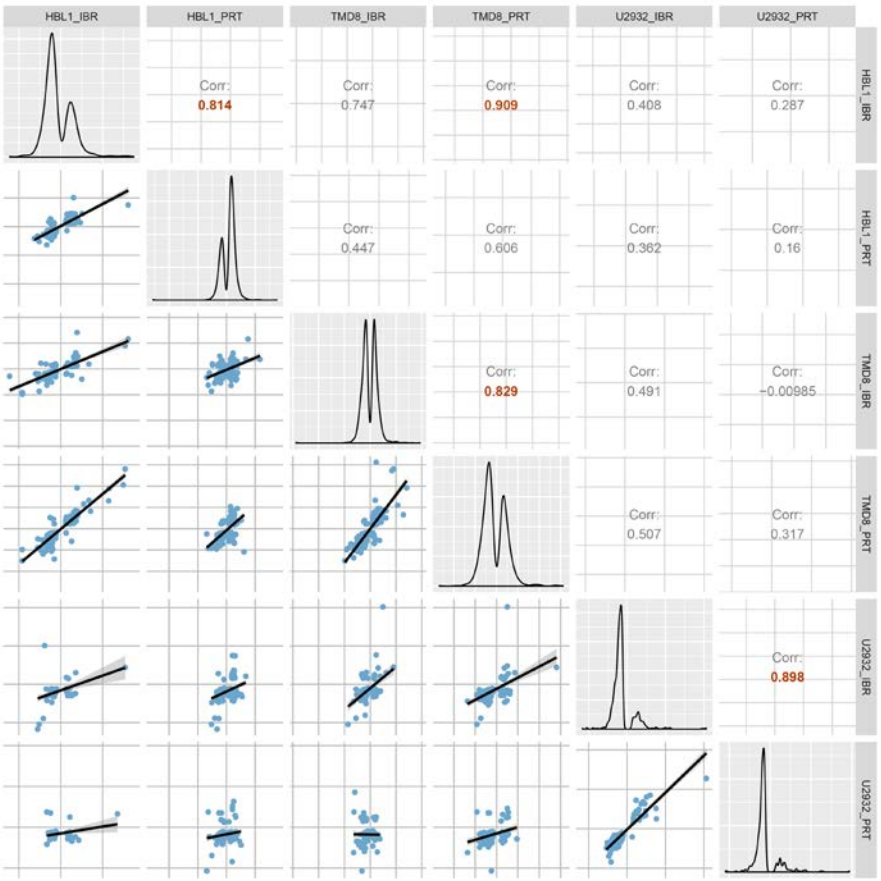


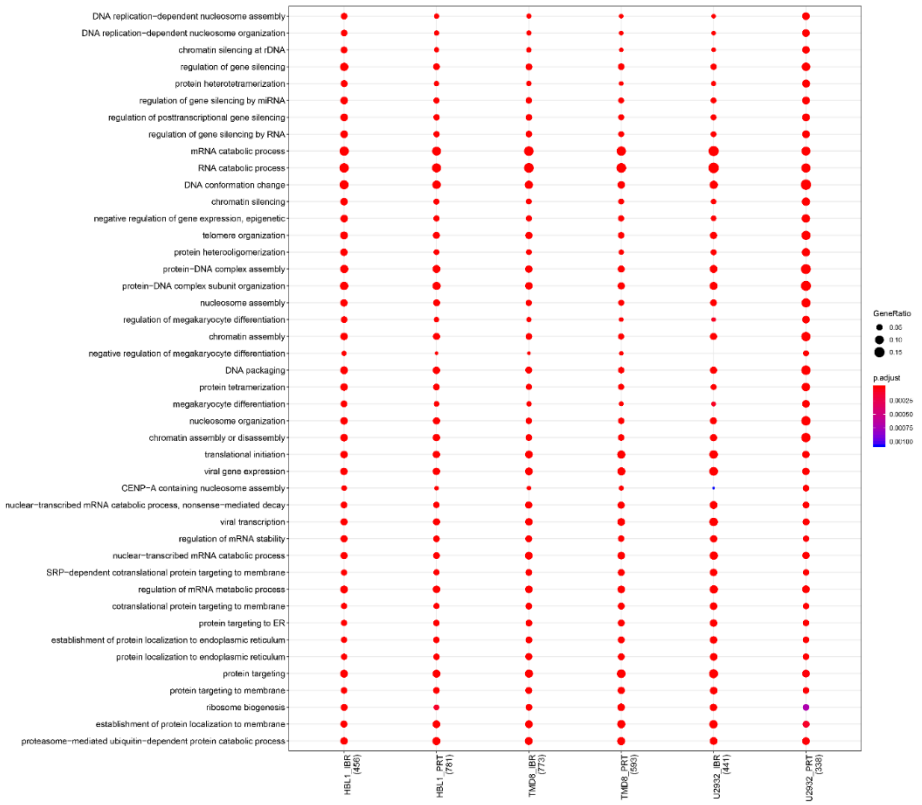
Figure 4.61 Overview of signaling perturbations in the ubiquitinome analysis of DLBCL with IBR and PRT treatments.

Common DUB-events among two samples were used for the plot. Scatter plots and Pearson correlations were shown separately for each pair plot of DUB-events. The high correlations of overlapped DUB-events were marked in red. "Corr" refers to Pearson correlation r.

4.5.7 Pathway enrichment analysis of DUB-events in the ubiquitinome of DLBCL

Reactome pathway and GO enrichment analysis of ubiquitylated proteins which bore the altered DUB-events were performed. The Benjamini-Hochberg adjusted p-values were set less than 1%. Most of the enriched biological processes were related with gene regulation (**Figure 4.62a**), for example, regulation of gene silencing, DNA packaging, regulation of mRNA stability, translational initiation, protein-DNA complex assembly or subunit organization. These ubiquitylated proteins involved in the same multiple pathways across each sample. Cellular responses to stress, infectious disease, M phase were highly enriched (**Figure 4.62b**).

a



b

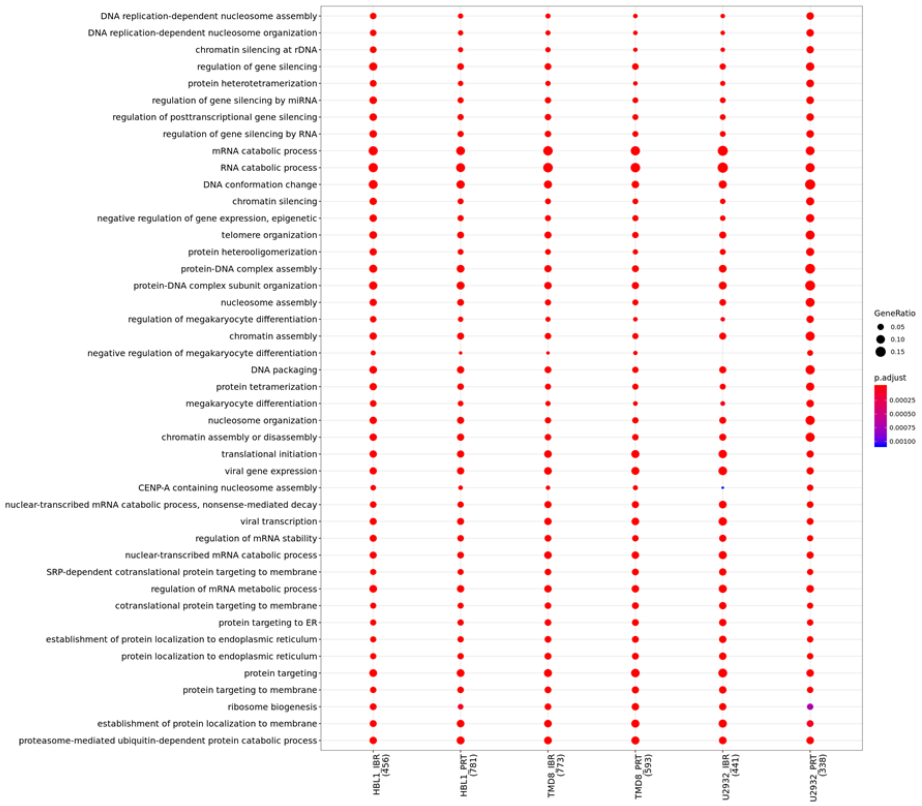


Figure 4.62 GO enrichment analysis and Reactome Pathway analysis of DUB-events in DLBCL.

The used numbers of DUB-events in each sample were marked. The adjusted p-value for each pathway was noted with color-bar. The ratios of DUB-events in each pathway compared to the total ones in each sample were shown with the size of filled circles.

5. Discussion

5.1 Benefits of the SugarQuant platform

Site-specific quantitative N-glycoproteomics analysis remains much more challenging than the conventional quantitative proteomics or phosphoproteomics analysis due to the inherent characteristics of glycosylation. SugarQuant integrates fast PAC-based sample preparation, optimized Glyco-SPS-MS3 and semi-automatic GlycoBinder script and enables to achieve confident identification and quantitation of intact N-glycopeptides in complex biological samples on a large scale. The whole workflow including sample preparation, MS acquisition, database search and data analysis was evaluated and optimized in SugarQuant. So the benefits are in many aspects. Until now, only two publications covered the large scale site-specific N-glycoproteome quantitative analysis [151, 153]. In order to represent the benefits clearly, I will make a comparison between SugarQuant and the two existing platforms as a follows. Firstly, regarding sample preparation, Sun et al [153] introduced a chemoenzymatic method called solid phase extractions of N-linked glycans and glycosite-containing peptides (NGAG), in which we have to identify potential glycosylation sites firstly and then determine the glycan heterogeneity on the sites. It is time-consuming and labor-intensive. Moreover, Stadlmann et al [151] used urea and HILIC respectively for protein extraction and glycopeptides enrichment. This workflow is relatively simple but the residual urea can negatively impact TMT labelling efficiency. In contrast, the PAC-based methods in SugarQuant enables to achieve sufficient solubilization of membrane-associated glycoproteins, complete and fast removal of detergents, efficient TMT-labeling, and selective glycopeptide enrichment with reduced handling time and sample loss. It is easy to handle even for non-specialist researchers. Secondly, a systematic evaluation and optimization of experimental parameters for chemically labelled glycopeptides was not performed in these two studies. We provided optimized Glyco-SPS-MS3 method and demonstrated its performance in both glycopeptide identification and quantification. Thirdly, their own database search algorithms termed GPQuest and SugarQB were developed respectively. In Sun's publication, 4,562 oxonium ion-containing spectra to 1,562 unique glycopeptides (containing 518 glycosites and 81 glycans) in OVCAR-3 cells were identified using GPQuest software with filtering based on the presence of peptide+HexNAc and/or peptide ions, as well as $\geqslant$7 observed b and y ions (1% FDR). In Stadlmann's study, it reported 1,100 glycopeptides mapping to 576 proteins in human embryonic stem cells with a Mascot-based peptide FDR cut-off using SugarQB. In contrast, we integrated well-tested and freely available software tools such as RawTools, MSConvert, pGlyco and pParse in GlycoBinder for data processing and database search. We identified and quantified over 5000 glycoforms containing 855 glycosites from 528 glycoproteins in DG75 cells using GlycoBinder with FDR control at three levels of matched glycan, peptide, and glycopeptide. In a word, we believe that SugarQuant outperforms the existing platforms.

5.1.1 Development and optimization of sample preparation for SugarQuant

The majority of mature glycoproteins locate on the cell surface or in the extracellular matrix. The first step of sample preparation in glycoproteomics should enable to achieve as much complete solubilization of membrane-associated glycoproteins as possible using detergents or chaotropic reagents. 8M urea is one of the commonly used lysis buffers to denature proteins in proteomics and glycoproteomics. A multi-folds dilution is necessary before enzymatic digestion due to its inhibitory effects on proteases at these high concentrations. For example, at least an 8-fold dilution is necessary prior to tryptic digestion of sample lysates in 8M urea. Then urea needs to be removed in digested peptide samples by desalting. Rapigest, an acid-labile detergent from Waters, has shown to facilitate complete digestion and increase

the solubility of hydrophobic proteins and peptides [227]. It undergoes hydrolysis under acidic condition to form pellets, so the clean-up is much more convenient by centrifugation instead of desalting. In addition, sodium dodecyl sulfate (SDS) is commonly used for the solubilization of various biological samples, especially for cell and tissue samples, owing to its outstanding capacity for extracting membrane proteins [228]. However, detergents diminish enzymes activities for protein digestion and cause serious ion suppression in MS analysis [229]. So the removal of SDS is prerequisite before MS analysis. There are already many approaches used for detergent removal, such as transmembrane electrophoresis [229], filter-aided sample preparation (FASP) [213], dialysis [230] and acetone precipitation [154]. Unluckily, almost all these methods make the whole sample preparation procedure more laborious and time-consuming, resulting in the higher risk of sample loss.

In order to obtain the optimal proteins solubilization and glycoprotein extraction, we used SDS, urea and Rapigest for protein extraction from DG75 cells separately. Our results showed that more proteins and glycoforms were identified using SDS and urea than using Rapigest as detergents (**Figure 4.2**), although the Rapigest workflow was faster. We chose SDS-containing lysis buffer for multiplex quantitative glycoproteomics analyses, due to the fact that incomplete removal of urea during desalting impairs TMT labeling efficiency seriously. However, removal of SDS via protein precipitation is too time-consuming, especially re-dissolving the protein pellets needs more efforts and causes sample loss to different extents. Hence, we sought to simplify the whole sample preparation workflow by reducing handling time. Recently, single-pot solid-phase enhanced sample preparation (SP3) has been developed which enables to achieve fast and efficient sample preparation for MS analysis [215]. Olsen group further investigated the underlying mechanism and found that the protein clean-up occurred irrespective of microparticle surface chemistry but instead via protein aggregation capture (PAC). During PAC protocol, proteins aggregate on the surface of magnetic beads followed by the removal of SDS by washing with high percentage of organic buffer. Thus, the whole handling time can be decreased to a large extent without any need to re-dissolve the protein pellets. Several different magnetic beads bearing various surface functional groups were chosen for optimization of the PAC workflow. Our results show that similar protein identifications were obtained (**Figure 4.2**). In our experience, acidifying the mixed TMT labelled peptides to ensure that the final concentration of TFA is around 0.5%-1% before drying is very important. Then glycopeptide enrichment is performed using ZIC-HILIC.

5.1.2 Development of Glyco-SPS-MS3 for high throughput intact glycopeptide identification and quantitation

Abundant glycan and peptide fragment ions on the spectrum are necessary for confident characterization of intact glycopeptides using MS. Till now, many different MS acquisition methods have been used for intact glycopeptide identification, mainly including utilization of one type of fragmentation method, or a combination of several different fragmentation methods. The single fragmentation method simplifies the MS workflow but often provides limited fragment ions information [148]. For example, CID mainly generated Y ions for glycan composition with little b, y ions for amino acid sequences, HCD produced diagnostic oxonium ions, Y ions and b, y ions which needs the energy optimization, and ETD yielded c/z-ion series for glycosite and peptide identity with little information on glycan. The combination of different fragmentation methods for simultaneous glycan-peptide characterization benefits the quality of acquired spectrum with complementary information on glycan or peptide fragments in some extents. However, these pipelines often involve complex MS workflows, long duty cycle, complex data processing and special bioinformatics tools. Recently, the fragmentation methods of stepped collisional energy HCD (sNCE-HCD) and AI-ETD were used for intact glycopeptides analysis and proved to achieve large-scale identification [154, 156]. However, none of

those methods dealt with chemically labelled glycopeptides for multiplexing quantification. We found that different types of fragment ions from TMT labelled intact glycopeptides under HCD including oxonium ions, glycan-product Y ions, peptide b-/y-ions, and TMT reporter ions need distinct NCEs to achieve the most intense responses during MS analysis (**Figure 4.6**). A multi-stage fragmentation method would be a good choice to cover the different optimal NCEs. In addition, it has been reported that MS3 methods brought benefits to the glycopeptide identification [152, 231]. By targeting the Y1 ion (whole peptide sequence carrying a single HexNAc) for further MS3 fragmentation, more peptide b-/y-ions will be recorded in the MS3 spectrum in order to improve peptide sequence identification [232]. However, prior knowledge of the targeted peptides selected for MS3 fragmentation is required and enough parent ions intensity to generate available information for MS3 analysis should be guaranteed. So its throughput and sensitivity are limited. Furthermore, currently there is no software available to automatically interpret and combine information from MS2 and MS3 fragmentation stages, which further limits the applicability of the MS3 approach, particularly for complex biological samples in a large scale analysis. Later, a synchronous precursor selection (SPS) to select multiple numbers of fragment ions from MS2 fragmentation was developed as SPS-MS3 method, which enables to reduce the co-isolation interference and increase the reporter ions intensity resulting in more accurate quantitation in proteome and phosphoproteome analysis [183, 184]. In order to achieve multi-stage fragmentation and multi-precursors selection, we introduced the SPS-MS3 method into the analysis of TMT labelled intact glycopeptides and named this method as Glyco-SPS-MS3 which we then optimized to obtain more confident identification and more accurate quantitation.

In Glyco-SPS-MS3, different HCD NCEs were used for MS2 and MS3 fragmentation to produce the complementary fragments of oxonium ions, glycan-related product Y ions, peptide b-/y-ions, and TMT reporter ions. Orbitrap detection for both MS2 and MS3 fragments enabled high-resolution and high-mass accuracy data acquisition. Synchronous precursor selection (multi-notches) of 10 MS2 fragments for MS3 improved the detection sensitivity of reporter ions and peptide ions, and even more importantly, decreased the co-isolation interference. The key fine-tuned MS parameters were described in detail in **Table 4.1**. The optimized MS settings for Glyco-SPS-MS3 which are applicable on Fusion and Lumos are listed **in Supplementary Table S1** in the appendices.

5.1.2.1 Detector and fragmentation type selection

In the original SPS-MS3 method, selected precursors are fragmented with CID followed by detection in an ion trap for MS2. Then multiple MS2 fragments were selected for HCD in the ion-routing multipole (IRM) followed by transferring all the fragment ions to Orbitrap for MS3 detection. In the Glyco-SPS-MS3 method, precursors are dissociated under HCD to generate more glycan-product Y ions and the resulting fragments are detected in high resolution and accuracy in Orbitrap in order to obtain the correct identification. The workflow of Glyco-SPS-MS3 is described as follows. Peptide ions injecting from an ion source firstly store in an IRM followed by Orbitrap detection for the MS1 scan. Intense precursor ions are selected in a quadrupole, transferred to an IRM with HCD fragmentation using lower NCE. The product ions are detected in Orbitrap for the MS2 scan. In the MS3 scan, the product ions from the MS2 HCD fragmentation in the IRM are transferred to an ion tap and the top ten most intense ions are selected using SPS technology followed by transferring them back to the IRM for HCD fragmentation with higher NCE. Then the fragment ions are sent to an Orbitrap for detection. Since Orbitrap is used for both MS2 and MS3 detection, the MS2 and MS3 scans are not performed in parallel resulting in longer cycle time. Our results showed that despite of 30% fewer triggered precursors in Glyco-SPS-MS3 compared to MS2, higher GPSM identification rates were obtained due to the improved qualities of MS2 and MS3 spectra in Glyco-SPS-MS3 (**Figure 4.11**).

5.1.2.2 Collisional energy optimization

12 TMT-labelled IgM glycopeptides were used to evaluate the fragmentation effects under various NCEs settings including single NCE in MS2, single NCE in MS3, stepped NCE (sNCE) in MS2, sNCE in MS2 plus single NCE in MS3 and single NCE in MS2 plus sNCE in MS3. Our results showed that optimal NCEs for different fragment ions from TMT-labelled intact glycopeptides located in a wide range to obtain the most intense responses (**Figure 4.6**). sNCE in MS2 couldn't generate enough reporter ions for quantitation although sufficient amounts of Y ions and peptide ions were obtained. One reason for that might be the more labile glycosidic bonds firstly absorbed most of the energy to break off and did not leave enough energy for fragmentation of reporter ions [181]. We found that the use of lower NCE (25%) in MS2 in combination with a higher energy (35-40%) in MS3 obtained the best identification results and more intense reporter ions for quantitation. In the MS3 scan, the Y ions with m/z in the range of 700-2000 which carry fewer oligosaccharides on glycan chain were selected for MS3 fragmentation resulting in more peptide ions and higher reporter ions due to the decreased inhibition by the generation of oxonium ions.

5.1.2.3 Trade-off between quality and speed (Ion target and injection time)

High quality of spectra with enough various types of fragment ions are critical for confident glycopeptide identification. Our results showed that higher automatic gain control (AGC) and longer injection time (IT) improve the quality of spectra despite of the prolonged cycle time. When the AGC was set to target values of $5e^4$, e^5, and $5e^5$, about 81%, 70%, and 5% of the spectra reached the pre-defined AGC within an IT of 100 ms, respectively. With prolonged ITs of 250 and 500 ms, about 29% and 68% of the spectra reached the AGC of $5e^5$. However, the duty cycle would be too long to allow coupling with the LC separation if using 500 ms IT to the MS2 and MS3. Then we allocated 500 ms IT in total to the MS2 and MS3 scans in different proportions. Interestingly, a longer IT in MS3 led to increased peptide score along with slightly decreased glycan score. This result was in accordance with our observations that low-NCE MS2 provided more glycan Y ions, while high-NCE MS3 generated more peptide b/y ions. Higher AGC always produces excessive space charge resulting in the impairment of mass accuracy [233]. However, our results showed that the precursor mass accuracy only located within ±4 ppm with the higher AGC $5e^5$.

5.1.3 Cross-talks between phosphorylation and N-glycosylation in DG75 upon 2FF treatment

Some major players in BCR signaling were found to be affected upon 2FF treatment. For example, we quantified 19 glycoforms mapping to two glycosites (Asn101, Asn127) on CD79B, which is the B-cell antigen receptor complex-associated protein beta chain for initiation of BCR signal transduction in cooperation with CD79A. Five fucosylated glycoforms on Asn 101 and two fucosylated glycoforms on Asn 127 showed a decreased expression upon 2FF-treatment, but there was no significantly changed glycosylation occupancy on these two sites. Also, 170 glycoforms on 3 sites (Asn46, Asn209 and Asn440) on IgM, which is the membrane-bound antibody portion of the BCR, were quantified. The expression of two fucosylated glycoforms on Asn 46 and Asn 209 decreased, but these two sites did not show changed glycosylation occupancy. In addition, a total 192 glycoforms on 6 glycosites Asn234 (22 glycoforms), Asn337 (29 glycoforms), Asn380 (104 glycoforms), Asn421 (2 glycoforms), Asn470 (11 glycoforms) and Asn531 (24 glycoforms) were quantified on human receptor-type tyrosine-protein phosphatase C (PTPRC or CD45), which is a tyrosine phosphatase involved in the regulation of B lymphocyte activation. 14, 11 and 7 significantly reduced fucosylated glycoforms were detected on Asn234, Asn337 and Asn531 respectively. However, only the glycosite Asn 531 exhibited decreased glycosylation occupancy. Our results suggest that 2FF site-specifically regulates glycosylation of

human Burkitt's lymphoma cells on important BCR players, which may in turn influence downstream BCR signaling. Besides the changed glycosylation in PTPRC, we detected significantly reduced phosphorylation on Ser-975, which lies within the tyrosine-protein phosphatase 2 domain and potentially affect the phosphatase activity of PTPRC according to PROSITE database [234]. Except PTPRC, down-regulated phosphorylation hits were also identified on 6 other proteins including SYK, BLNK, NFATC1, NFATC2, NFKB1 and CD19. Each of them involves B-cell activation pathway. For instance, SYK is phosphorylated by binding to phosphorylated ITAMs of CD79A and CD79B, resulting in the activation of SYK. BLNK as an adaptor protein is recruited by the active BCR signalosome. CD19, a glycoprotein on the surface of membrane, is activated by SYK and other SFKs and involved in PI3K pathway in BCR signaling. NFATC1 and NFATC2, as the nuclear factor of activated T cells family of transcription factors, regulate genes vital for cell growth and survival in hematopoietic cells. They mainly induces the expression of cytokine genes in T cells, such as, IL-2 or IL-4. Recently, it was reported that NFATC1 was involved in the induction of programmed death ligand (PD-L1) signaling in DLBCL [235]. NFKB1 is one effector in NF-κB pathway, which is a main downstream pathway used by B-cells to maintain the viability. In addition, the functional enrichment analysis revealed that these proteins, whose phosphorylation decreased upon 2FF treatment, were highly involved in the pathways of cell cycle, M phase, gene expression and DNA repair. It suggested a possible route of 2FF-modulated phosphorylation signaling. Therefore, we assumed that changes in glycosylation on a receptor could affect the phosphorylation levels of its phosphosites.

The combination of N-glycoproteomics and phosphoproteomics analyses revealed cross-talks between phosphorylation and N-glycosylation in Burkitt's lymphoma cells. 2FF-induced de-fucosylation on glycan structures induced the changes in site-specific glycoforms and overall glycosite occupancy. In addition, 2FF treatment also affects phosphorylation of the key players involving in B cell receptor signaling. Consequently, our results broadened the understanding of the roles of PTMs in BCR signaling in lymphoma and benefited the lymphoma therapies.

5.1.4 The applicability of SugarQuant platform

Protein glycosylation is one of the most structurally complicated and diverse type of PTMs. Each glycoprotein can own multiple glycosites, and each glycosite can be modified by multiple different glycan structures. Our developed SugarQuant platform enables to achieve glycoprotein characterization nicely from multiple aspects, such as, which protein is glycosylated, the location of glycosite, the occupancy of each glycosite, and which glycan compositions are attached with each glycosite. What's more, it attains the relative quantitation of the intact N-glycopeptides in multiple samples. Until now, more and more researches reveal that site-specific N-glycosylation patterns regulate the physiological and pathological process of various diseases [76, 149, 185]. We believe that SugarQuant is a robust and powerful platform to discover the aberrant glycosylation and quantify the dynamic glycopeptide changes during the progression of diseases. It can be applied to biological and biomedical research extensively.

In particular, since the beginning of this year, a novel coronavirus (SARS-COV-2) emerged and rapidly spread, developing into an epidemic of severe acute pneumonia syndrome (COVID-19), which poses a major threat to global human health. The highly glycosylated spike protein (S) on the virus surface mediates cell entry by binding to the angiotensin-converting enzyme II (ACE2) receptor on host cells [236]. The S protein comprises two functional subunits of S1 subunit responsible for receptor binding and S2 subunit responsible for membrane fusion. 22 potential N-glycosylation sites exist in each monomer of S protein. Some glycosylation sites were observed in S protein using cryogenic electron

microscopy (cryo-EM) [236]. Viral glycosylation plays important roles in viral pathobiology, including mediating protein folding, stability and protein priming by host proteases. N-glycans may shield specific amino acid residues from cell and antibody recognition, leading to the immune evasion of virus. Significantly, SugarQuant enables to rapid identification and comparative quantitation of intact N-linked glycopeptides from S protein in virus and the proteins in host cells transfected with viruses experimentally. It will provide more evidences for us to decipher the mechanisms of virus invasion and to guide vaccine design and antiviral therapeutics development.

5.2 Site-specific quantitative glycoproteome analysis of DLBCL subtypes

More than 7000 glycoforms were quantified in five ABC DLBCL and four GCB DLBCL cell lines using SugarQuant. The segregation of DLBCL subtypes could achieve by their intact glycopeptide profilings. In addition, we also found the difference on glycosylation levels in the same subtype of DLCBCL cell lines (refers to section 5.2.1 and 5.2.2).

5.2.1 Differential glycosylation level in WSU-FSCCL

No matter in IG and DG analysis, we found that the site-occupancy (**Figure 4.28**) and intact glycopeptide expressions (**Figure 4.30, Figure 4.33**) in WSU-FSCCL were significantly different with those in other DLBCL cell lines, even in GCB DLBCL cell lines. We guess that it might be related to the genetic subtypes in DLBCL. Till now, six genetic subtypes were revealed including MCD, N1, A53, BN2, ST2 and EZB, which have a complex mapping relationship to the DLBCL subgroups [237]. HBL1, TMD8 and OCI-Ly10 mapped to MCD. RIVA mappd to BN2. DOHH2, SUDHL4 and SUDHL5 were catogerized into EZB. Whereas, a large portion of DLBCL cell lines was not found their corresponding genetic subtypes yet, such as WSU-FSCCL. The further research is needed to reveal the comprehensive relationships between the DLBCL subgroups and genetic subtypes and their roles in the pathogenesis.

5.2.2 Differential expression of fucosylated and sialyated glycopeptides in RIVA and OCI-Ly3

In IG analysis, we observed that the fucosylated and sialyted glycopeptide expressions in RIVA and OCI-Ly3 were different from those in the other 3 ABC cell lines (**Figure 4.31**). However, all the five ABC cell lines were categorized into one group based on their site-occupancy level in DG analysis (**Figure 4.28**). SugarQuant helps us to obtain the site-specific glycosylation levels in different DLBCL cell lines. Then we deduced that the fucosylated and sialyated glycopeptide levels in cell lines from ABC DLBCL subtype were probably different. It might revealed that the sialyation and fucosylation played roles in DLBCL subtypes. Although, till now, no publication reported the roles of fucosylation and sialylation in DLBCL. In B16 melanoma cell line, inhibition of sialylation by fluorinated sialic acid analogue (3Fax - peracetyl Neu5Ac) was found to reduce cell adhesion, migration, and growth [238]. In addition, in liver cancer cell, it was reported that inhibition of fucosylation suppress cell proliferation, migration and tumor formation [192].

5.2.3 Glycoforms as the driver for the segregation of ABC and GCB DLBCL subtypes

From IG analysis using SugarQuant, we revealed that the specific glycoforms could be used for the segregation of ABC and GCB DLBCL. The majority of these glycoforms were from the key cell-surface BCR effectors. For instance, on CD44 which was repressed by BCL-6 in GCB DLBCL, 15 glycoforms were identified on Asn57. On CD53, which was required for cell fusion, 10 glycoforms were identified on Asn129. On CD166, which helps to identify subsets of proinflammatory B lymphocytes and drive their transmigration across different CNS barriers in mouse and human [239], 1 glycoform on Asn265,

6 on Asn306 and 3 on Asn361 were found. On IL4I1, which was reported as a negative immune checkpoint controlling B cell differentiation and activation [240], 2 glycoforms on Asn54, 2 glycoforms on Asn134 and 1 glycoform on Asn559 were identified.

We identified 371 highly expressed glycoforms in ABC DLBCL. For example, on Ly75 (CD205), which captured foreign antigens and transported to the antigen-processing compartment, resulting in the decreased B-lymphocytes proliferation, 4 glycoforms on Asn865, 9 glycoforms on Asn1103 and 1 glycoform on Asn1593 were highly expressed in ABC DLBCL. On SIGLEC10, which is a negative regulator of B-cell receptor-mediated calcium signaling, 1 glycoforms on Asn100 and 2 glycoforms on Asn355 were highly expressed in ABC subtypes [241]. 2 glycoforms on Asn195 and Asn431 on SEL1L were identified as signature for ABC subtype, while SEL1L binding to Hrd1 mediates the endoplasmic reticulum associated degradation to regulated the B cell development and enables to recognize and target the pre-B cell receptor [242]. 4 glycoforms on Asn170 and 1 glycoform on Asn188 on CD37 antigen, which were demonstrated to be one therapeutic target in B cell NHL, [243] were highly expressed in ABC DLBCL. 16 glycoforms on Asn64 (4), Asn210 (2), Asn242 (2), Asn694 (5) and Asn731 (3) on TLR9 were up-regulated in ABC. Recent study revealed that TLR9 binding to MYD8 and BCR as My-T-BCR supercomplex co-localized with mTOR on endolysosomes. This supercomplex drived pro-survival NF-κB and mTOR signaling in ABC subtype [40]. The regulated processes might be involved by the glycosylation of TLR9. 7 glycoforms on Asn46 (3) and Asn209 (4) on IgM and 3 glycoforms on Asn101 on CD79B were highly expressed in ABC DLBCL. Interestingly, 2 glycoforms on Asn46 on IgM and 2 glycoforms on Asn101 on CD79B were up-regulated in GCB subtype (**Figure 5.1 a,b**).

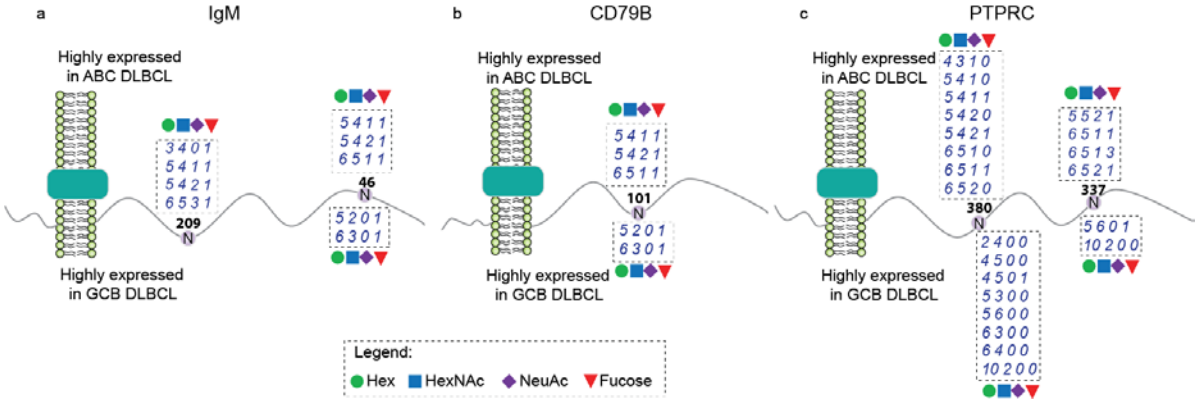


Figure 5.1 Visualization of highly expressed glycoforms in ABC and GCB DLBCL on IgM, CD79B and PTPRC.

Glycan compositions were presented as the numbers of Hex, HexNAc, NeuAc and Fucose in dashed box.

202 glycoforms were up-regulated in GCB DLBCL subtypes. 20 glycoforms on Asn177 were identified on IgG, which was the pre-dominant IgH isotype in GCB subtypes [244]. In addition, 7 glycoforms on Asn71 on IgJ were up-regulated in GCB DLBCL, giving rise to that IgJ might be one IgH isotype in BCR in some cases of GCB DLBCL. 2 glycoforms on Asn94, 3 ones on Asn205, 2 ones on Asn212 and 2 on Asn406 on ITGB1 which might involve B cell differentiation [245] were found. 4 glycoforms on Asn732 on SUN1, 2 glycoforms on Asn160 on CREG1, 2 glycoforms on Asn259 on STS, 7 glycoforms on Asn212 on PPT1, 5 glycoforms on Asn467 on P3H1 and 6 glycoforms on Asn527 on SE1L3 were also up-regulated in GCB subtypes. However, all these genes' roles in lymphoma are still unclear.

We also found that some other glycoforms which happened on one specific site on the same protein had distinct expression in ABC and GCB DLBCL subtypes. 73 glycoforms with high expression in GCB subtype and 77 glycoforms with high expression in ABC subtype occurred on 24 sites from 19 proteins. For example, 4 glycoforms on Asn337 and 8 glycoforms on Asn380 on PTPRC which is a tyrosine phosphatase involved in the regulation of B lymphocyte activation were up-regulated in ABC, 2 and 8 other glycoforms on Asn337 and Asn380 on PTPRC were up-regulated in GCB DLBCL (**Figure 5.1c**). 14 glycoforms on Asn84 (1), Asn103 (2), Asn249 (6) and Asn261 (5) on LAMP1 as a lysosomal and plasma membrane protein which contributes to tumor metastatic potential and differentiation were up-regulated in ABC DLBCL. 15 glycoforms on Asn249 (7), Asn261 (6) and Asn322 (2) on LAMP1 were up-regulated in GCB DLBCL. Moreover, LAMP1 was proved that its expression was related with the poor prognosis DLBCL patients [246]. On LAMP2, which was regarded as a marker of EBV-mediated lymphocyte transformation [247], 8 glycoforms on Asn49 (1), Asn58 (1) and Asn356 (6) separately were highly expressed in ABC DLBCL, 20 glycoforms which mapped on Asn58 (2), Asn275 (4) and Asn356 (14) separately were up-regulated in GCB DLBCL. Our results provide new insights from the intact glycopeptide level to investigate the relationship of N-glycosylation with the distinct mechanisms of tumorigenesis in ABC and GCB DLBCL.

5.3 Reduced fucosylation and increased sialyation in 2FF-treated and FUT8 KO TMD8 cell line

In order to investigate the specific change of fucosylated glycopeptide expression, we performed a site-specific quantitative N-glycoproteomics analysis in 2FF-treated and FUT8 KO TMD8 cells using SugarQuant. We found that the distinguished glycopeptide expression existed in the conditions of 2FF-treatement and FUT8-KO TMD8 cells (**Figure 4.39**).

Due to that about 86% fucosylated glycopeptide-spectrum-matches (GPSMs) (3466/4012) were core-fucosylation, whose synthesis was regulated by FUT8, we found that the expressions of fucosylated glycoforms were inhibited more serious by FUT8 KO than 2FF treatment (**Figure 4.40)**. In addition, the result of better FUT8 knockout effect induced by sgFUT8-III compared to sgFUT8-II was consistent with the western and lectin blot analysis.

5.3.1 The role of FUT8 in TMD8

We didn’t observe the down-regulated non-fucosylated glycoforms, whereas, the up-regulated ones in non-fucosylated glycoforms were identified (**Figure 4.40a, bottom**). In addition, there were also the up-regulated fucosylated glycoforms (**Figure 4.40a, upper**). The unique glycan composition on specific position of modified Asn on specific protein was used for the definition of unique glycoform, inducing that one unique glycoform may include the core-fucosylated and non-core-fucosylated glycan chains with the same glycan composition. Then this unique glycoform was quantified using the merged quantitation information of the used core- and non-core fucosylated ones. In order to obtain the correct core-fucosylated glycopeptides expression, we analyzed the core-fucosylation on GPSMs level to eliminate the interference of non-core-fucosylated boring the same glycan composition with core-fucosylated. Interestingly, if we separated the fucosylated GPSMs into two types based on whether it was core-fucosylated or terminal-/subterminal-fucosylated (also called non-core-fucosylated), almost all the core-fucosylated GPSMs expression decreased compared to control in each sample. Previous study revealed that sialyation was up-regulated in 2FF treated HepG2 cells due to the lack of terminal fucosylation impaired by 2FF and the competition of the same terminal structure by fucosyltransferases and sialytransferases [191, 192]. Our result showed that the core-fucosylation inhibition by specific

knockout of FUT8 was also able to facilitate the sialyation expression. One possible explanation is that FUT8 also regulated the synthesis of terminal fucosylation in some way with the cooperation of FUT1 and FUT2, which are regarded to be responsible for terminal fucosylation synthesis. Most of the 87 non-sialyated non-core-fucosylated GPSMs were down-regulated even in FUT8 KO samples (**Figure 4.40c**), further indicating that FUT8 was mostly probably responsible for the synthesis of non-core-fucose on N-glycans in lymphoma cells. Moreover, if we can identify terminally fucosylated N-glycopeptides in FUT1 and FUT2 KO cells, it will help to support our deduction.

5.3.2 Increased sialyation in fucosylation inhibited TMD8

We found that down- and up-regulated GPSMs whose glycan chains bore sialic acid were quantified in each sample (**Figure 4.43a**). And many up-regulated non-fucosylated GPSMs with sialyation were revealed (**Figure 4.43c**). The sialyated glycopeptides whose glycan chain did not bear fucose or bore non-core-fucose were up-regulated, which was probably derived from the competitive binding to the same oligosaccharides during subterminal-/terminal-fucose and sialic acid synthesis.

5.4 Differential phospho-events in DLBCL with BTK and SYK inhibition

More than 30,000 phosphorylation sites in ABC and GCB DLBCL cell lines with BTK and SYK inhibition using two-plex SILAC labeling strategy were quantified. More importantly, 9783 differentially regulated phosphorylated peptides, mapping to 3561 phosphoproteins, under pharmaceutical inhibition of effectors in BCR signaling, were identified.

5.4.1 Criterion for filtering differential phospho-events in pYome of DLBCL

In pYome analysis, a huge variances of the numbers of identified p-events were obtained among the 11 samples, like only 189 p-events in WSU-FSCCL with PRT treatment, 758 p-events in U2932 with IBR treatment. Only 64 p-events were quantified across all 11 samples. What's worse, only 33 p-events were commonly quantified in all 46 samples (each replicate as one sample). Due to the very high missing values and no feasible imputation way, the principal component analysis and hierarchical clustering of pYome data were not meaningful. In addition, t-test statistics analysis was not suitable for the pYome dataset, which had much fewer p-events and huge missing-values. So we used the Log_2 SILAC ratio with ≥ 0.5 and ≤ -0.5 as the threshold for the significantly regulated p-events.

5.4.2 Differential phospho-events in DLBCL

Regarding to the DP-events in pYome analysis of DLBCL, very few down-regulated p-events were identified in all 11 samples, even no up-regulated p-events in HBL1 with IBR and DOHH2 with PRT samples. The most DP-events were in U2932 with IBR treatment. The DP-events in DOHH2 and WSU-FSCCL with PRT treatments were less than 15 due to the least identification of total p-events among them. Our quantitation showed that the overall tyrosine-phosphopeptides expression decreased after BTK or SYK inhibition in DLBCL (**Figure 4.53d**).

Regarding the DP-events in GPome analysis of DLBCL, the numbers of down-regulated and up-regulated p-events in each cell with BTK or SYK inhibition were different. For example, fewer DP-events in GCB DLBCL cell lines were identified than in ABC DLBCL. The most DP-events were in HBL1 with BTK inhibition. The most DP-events after SYK inhibition were found in HBL1. In U2932 cell line with BTK and SYK inhibitions there were the minimums of DP-events. Interestingly, except in TMD8 cell lines with Ibrutinib and PRT treatments, the down-regulated p-events were much more than up-regulated ones in all the other cell lines with BTK or SYK inhibition. Combined with the

analysis of overlapped DP-events across the 11 samples (**Figure 4.49**), we assumed that the altered phosphorylation expression after BTK or SYK inhibition was regulated differentially in different lymphoma cell lines, which might be weaker relationship with the subtypes of DLBCL (**Figure 4.47**).

Subsequently, we focused on some BCR signaling proteins, in which the change of phosphorylation was identified. MYC, a transcription factor, is involved in disease pathogenesis. The chromosomal aberrations in MYC are found in BL and the oncogenic synergy between MYC and PI3K pathway might occur in BL [25]. The phosphorylation of MYC at Thr58 leaded to MYC degradation in BL [248]. In HBL1 and SUDHL4 with SYK inhibition sample, we detected the increased phosphorylation at Thr58 on MYC. The decreased phosphorylation at Thr58 on MYC in Ly10 with SYK inhibition was found. Three other phosphorylation sites on MYC in Ly10 with SYK inhibition were reduced. The phosphorylation on MYC might affect BCR signaling in different ways in distinct DLBCL cell lines. BLNK, as a central linker protein, bridges the SYK kinase to a downstream signaling pathways. We found decreased phosphorylation of BLNK at six distinct sites in BTK or SYK inhibited DLBCL cell lines. In NF-κB pathway, BLNK is regarded to bridge SYK and BTK and mediates the signaling transducing. Therefore, it suggested that the phosphorylation of BLNK directly regulated BCR signaling. My-T-BCR supercomplex (MYD88, TLR9 and BCR) was uncovered to drive pro-survival NF-κB and mTOR signaling in DLBCL. Previous study showed that inhibitors of BCR and mTOR signaling cooperatively decreased the formation and function of the My-T-BCR supercomplex [40]. However, our result showed that the increased phosphorylation of MYD88 at Ser257 in HBL1 and TMD8, the increased phosphorylation of TLR9 at Ser532 in TMD8. It revealed that the inhibition of BTK or SYK might not decrease the expression of MYD88 and TLR9 directly and their phosphorylation could be affected via the regulation process of their protein synthesis. Moreover, we found the decreased phosphorylation of mTOR at Ser567 in HBL1 and Ly10 with SYK inhibition and reduced phosphorylation of mTOR at Ser2448 in HBL1 with BTK inhibition. It might be related with the decreased BCR signaling upon BTK and SYK inhibition.

Together, the changed phosphorylation of about one-third of all quantified sites howed that the disturbance of BCR signaling by the inhibition of BTK or SYK caused a reorganization of phosphorylation signaling networks. Many previously reported BCR signaling effectors, such as tyrosine kinases, were included in our dataset. However, the regulated phosphorylation from the majority of quantified proteins was uncovered for the first time.

5.5 Differential ubi-events in DLBCL with BTK and SYK inhibition

We investigated the regulatory ubiquitylation in DLBCL cell lines with BTK and SYK inhibition systematically by the analysis of di-Gly captured ubiquitylated peptides. Notably, the significantly differentially ubiquitylation was identified on many of the members in BCR signaling, including CD79A and CD79B, tyrosine kinases (BTK, SYK, LYN). No matter in ABC DLBCL and GCB DLBCL, the activation of PI3K pathway remains in BCR signaling. In our result, six decreased ubiquitylated peptides at six distinct sites on PIK3 adapter protein 1 (PIK3AP1) and the reduced ubiquitylation on PIK3 subunits (PIK3C2A, PIK3CD) were identified in three cell lines upon BTK or SYK inhibition. It showed the successful disturbance of BCR signaling upon the impairment of upstream tyrosine kinases. Tumor necrosis factor alpha-induced protein 3 (TNFAIP3, also known as A20), contains deubiquitinase activities, is a key negative regulator of NF-κB pathway. Three decreased ubiquitylated peptides of TNFAIP3 at Lys81, Lys301 and Lys643 in HBL1 and TMD8 and one increased ubiquitylated peptides of TNFAIP3 at Lys722 were found. In addition, we also identified the decreased ubiquitylation on MALT1 and BCL10 in ABC and GCB DLBCL. It reported that A20 deubiquitylated MALT1 and

inactivated the CBM complex resulting in the inhibition of NF-κB signaling. It revealed that the ubiquitylation of TNFAIP3, MALT1 and BCL10 might mediate their regulation in BCR signaling. Together, our results showed that ubiquitylation is regulated at thousands of sites upon the inhibition of upstream tyrosine kinases within 3 hours. A huge amount of ubiquitylation-regulating proteins including E3 ubiquitin-protein ligases and deubiquitinases were identified in our dataset. Moreover, we found the changed ubiquitylation from BCR related effectors, DNA replication factors, transcriptional effectors and translational activators. It indicated that ubiquitylation-dependent signaling plays a more ubiquitous role in BCR signaling.

5.6 Utilization of GPome and Ubiquitinome data in DLBCL with BTK and SYK inhibition

5.6.1 Comparison of Phosphoproteome and ubiquitinome analysis in DLBCL

In this study, we obtained two datasets including essential gene hits in DLBCL. One is from significantly differentially phosphorylated sites in phosphoproteome analysis. The other one derives from significantly differentially ubiquitinated sites in ubiquitinome analysis. In order to know whether significant gene hits from phosphoproteome and ubiquitinome analysis preferentially involve different biological processes, or whether there are primarily redundant information, we firstly investigated the overlap between significant protein hits using UNIPROT identifiers in phosphoproteome and ubiquitinome analysis. Because ubiquitinome analysis was only performed in HBL1, TMD8 and U2932 cell lines, the datasets from these three cell lines were used for further overlap analysis. 592 proteins that are significantly modified by both phosphorylation and ubiquitylation were found. In contrast, the majority of regulated proteins were solely identified by one type of PTM, for example, 80% (2321/2913) proteins modified with only phosphorylation, 70% (1364/1956) proteins modified with only ubiquitylation.

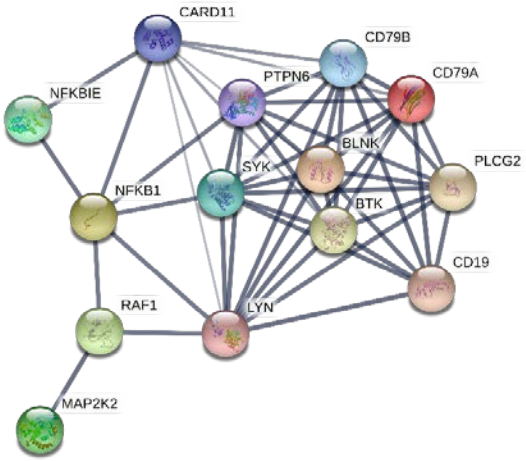


Figure 5.2 STRING interaction networks of the essential proteins with the co-occurrence of phosphorylation and ubiquitylation in BCR signaling.

Regarding these proteins co-regulated by phosphorylation and ubiquitylation, some of them were known effectors implicated in BCR signaling pathway (**Figure 5.2**). The components of BCR, CD79A and CD79B, both of them were ubiquitylated and phosphorylated. In CD79B, phosphorylation and ubiquitylation happened on the adjacent position-Ser221 and Lys219. Interestingly, phosphorylation at Ser221 was decreased in HBL1 with SYK inhibition. The increased ubiquitylation at Lys219 was found in SYK inhibited HBL1. Co-occurrence of phosphorylation and ubiquitylation on BTK and SYK was also identified, such as, 4 and 14 regulated ubiquitylation sites on BTK and SYK, 4 and 8 regulated

phosphosites on BTK and SYK, respectively. LYN and PTPN6, as the known BCR negative regulators, also co-occurred both PTMs. Reduced phosphorylation level on PTPN6 was found in HBL1 and TMD8 cells. Whereas, both increased and decreased ubiquitylation modified peptides on PTPN6 were identified. We found 5 increased phosphorylation modified peptides on CARD11, one component of the CBM complex, in all three cell lines with BTK or SYK inhibition. The increased ubiquitylation of CARD11 at Lys994 was only identified in TMD8 with BTK inhibition. Although these proteins were known to be essential in BCR signaling, most of the regulated sites identified here, especially the ubiquitylation sites, and how to be regulated in BCR were unknown.

The activation of NF-κB pathway is a hallmark for the survival of ABC DLBCL. Upon the inhibition of BTK and SYK in ABC DLBCL, we found decreased phosphorylation of NFKB1 at Ser893 and Ser907. While in BTK and SYK inhibited U2932 cells, the increased phosphorylation of NFKB1 at Ser903 and Ser907 was found. The decreased ubiquitylation of NFKB1 at Lys243 was only identified in TMD8 cell with SYK inhibition. In addition, we also identified the decreased phosphorylation of NFKBIE at Ser44 and Ser28 in HBL1. One ubiquitylated peptide of NFKB1 at Lys177 decreased in HBL1 with BTK inhibition. Together, these results provided a foremost overview of co-regulated phosphorylation and ubiquitylation in BCR signaling in DLBCL.

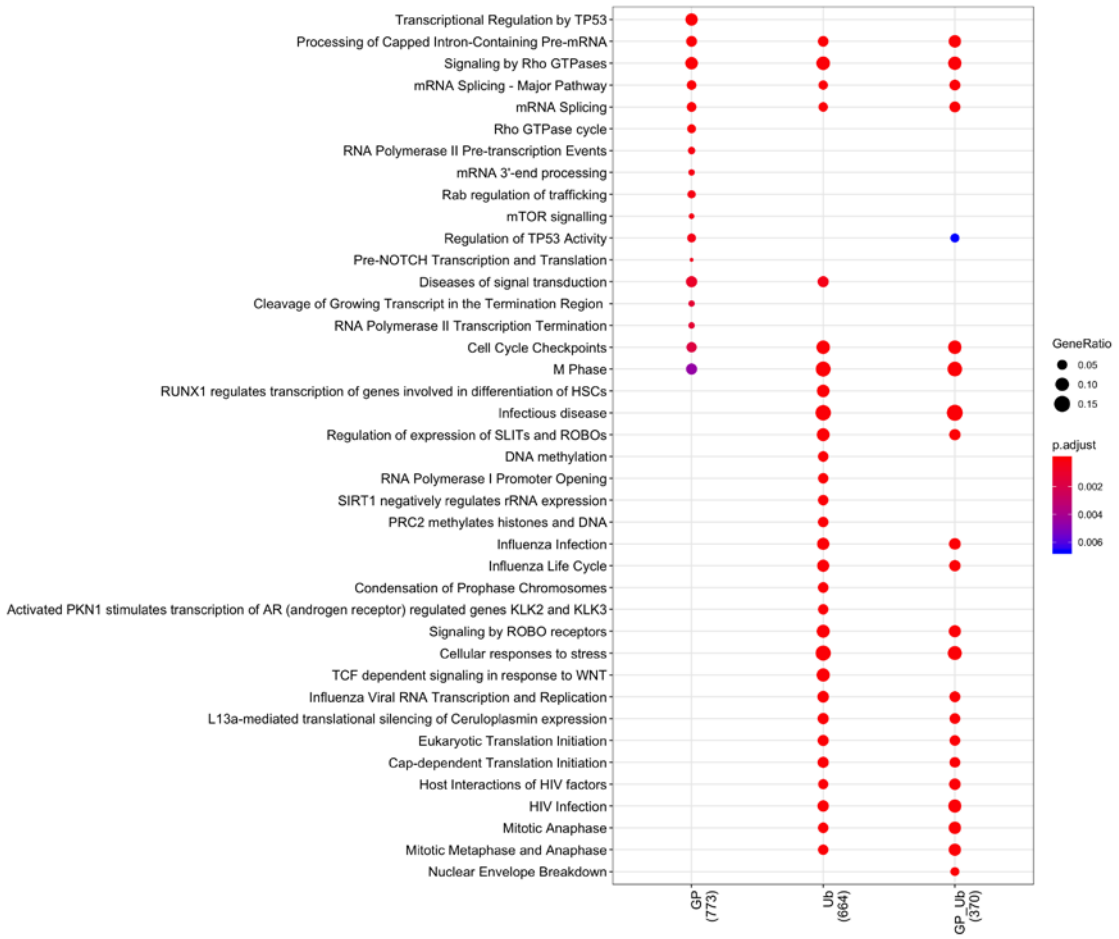


Figure 5.3 Reactome Pathway analysis of essential hits identified from phosphoproteome and ubiquitinome analysis in DLBCL.

The adjusted p-value for each pathway was noted with color-bar. The ratios of p-events and ubi-events in each pathway compared to the total ones in each sample were shown with the size of filled circles. GP is phosphoproteome, Ub is ubiquitinome.

Next, in order to know which biological pathways were involved by these essential proteins, we performed a Reactome pathway analysis using hits identified by both phosphoproteome and ubiquitinome, as well as those identified by a single modification (**Figure 5.3**). These significant hits from different cell lines were pooled together. This analysis revealed that these gene hits from phosphoproteome and ubiquitinome analysis cover primarily distinct pathways. Moreover, these pathways, such as, mRNA processing, cell cycle checkpoints, M phase, and signaling by Rho GTPases were highly involved by essential hits from either phosphoproteome or ubiquitinome analysis. All in all, we assumed that these gene hits from both PTM analysis are largely complementary.

5.6.2 Integration of phosphoproteome, ubiquitinome and CRISPR dropout screen analysis in DLBCL

In collaboration with Prof. Dr. Louis M.Staudt lab (National Institutes of Health), we used a library of small guide RNAs (sgRNAs) to conduct loss-of-function CRISPR dropout screens for essential genes in DLBCL cell lines engineered with inducible Cas9. Untill now, we already screened four ABC DLBCL cell lines, five GCB DLBCL cell lines and three multiple myeloma cell line as controls. CRISPR dropout screens help us to find the viability regulators in each DLBCL cell line. For each gene, we perform basic diagnostics on the derived CRISPR screen score (CSS) from a gene-level statistic. CSS is the number of standard deviations away from the average effect of inactivating a gene [40]. In this study, we performed the global phosphoproteome and ubiquitinome analysis of DLBCL cell line treated with two tyrosine kinase inhibitors using SILAC-based quantitative proteomics technology. We successfully identified multiple differentially phosporylated and ubiquitinated sites for each condition using a sensitive differential detection algorithm. It enabled us to look for potential BCR signaling effectors from these phosphorylation and ubiquitination modified proteins whose phosphorylation and ubiquitination were changed. The integration of global phosphoproteome and ubiquitinome results with the CRISPR screen hits will allow us to look for essential gene hits exhibiting significant differential phosphorylation and ubiquitination in further. The filtered gene hits most probably work as viability regulators in the BCR network. The integrative analysis enables a more comprehensive understanding of the molecular landscape of lymphoma and offers an opportunity for more precise diagnosis and treatment.

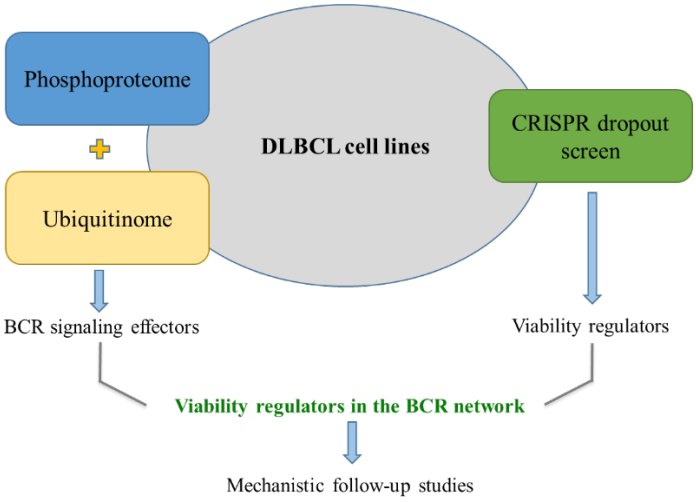


Figure 5.4 Integrative analysis of phosphoproteome, ubiquitinome and CRISPR screens in DLBCL.

6. Conclusions and Perspectives

In the first part, we developed a platform for multiplexed quantitation of site-specific N-glycoproteomics termed SugarQuant. We successfully applied SugarQuant to detect the site-specific glycosylation changes in DG75 cell line upon pharmacological interference with fucosylation. Furthermore, expression of intact glycopeptides in cell line of DLBCL subtypes was quantified on a global level, whose accuracy of quantitation was validated by deglycoproteomics analysis using the same sample. In addition, we successfully determined the altered expression of fucosylated glycopeptides in the cell line TMD8 upon knockout of the fucosyltransferase (FUT8) and by the inhibition of fucose synthesis using 2FF. Our results provide the largest N-glycoproteome data in Burkitt's lymphoma (DG75) and in few ABC and GCB DLBCL cell lines. This platform which achieves efficient large-scale profiling and accurate quantitation of glycoforms, glycosites and glycans, will help us to understand the roles of site-specific N-glycosylation regulation in health and diseases. Moreover, an in-depth quantitation of significant differential phosphorylation events in ABC and GCB DLBCL cell lines upon inhibition of the tyrosine kinases BTK and SYK was achieved in our SILAC-based quantitative phosphoproteome analysis. The results revealed a cell-type-specific signaling responses in distinct cell lines upon BTK and SYK inhibitions. The identified phosphoproteins which bear significantly regulated phosphorylation sites will serve as a valuable resource for us to look for essential effectors in BCR signaling in lymphomas. The ubiquitinome data provides complementary information for ubiquitylated proteins in BCR signaling. In collaboration with Louis Staudt lab, we identified a pool of essential effectors in few ABC and GCB DLBCL cell lines using the CRISPR dropout screen technology. The integration of phosphoproteome, ubiquitinome and functional genomics data will help us to find the therapeutically most meaningful effectors in the BCR signaling network.

7. Zusammenfassung

Posttranslationale Modifikationen (PTM) sind ein wichtiger biologischer Regelmechanismus für verschiedenste Signalwege sowie die Kontrolle des Zellschicksals. In dieser Arbeit wurde massenspektrometrische, quantitative Proteomik zur Untersuchung unterschiedlicher PTM wie Glykosylierung, Phosphorylierung und Ubiquitinierung in diffus-großzelligen B-Zell-Lymphomen und Burkitt Lymphomen unter verschiedenen Behandlungsansätzen im großen Maßstab genutzt.

Die Etablierung der quantitativen N-Glykoproteomik-Plattform SugarQuant

Einhergehend mit dem zunehmenden Verständnis der N-Glykosylierung von Proteinen und seiner ortsspezifischen Regulation in Physiologie und Pathophysiologie, erwächst ein dringendes Bedürfnis nach einer geeigneten LC-MS Methode zur Quantifizierung intakter N-Glykopeptide (z.B. Peptide mit intaktem Glykanrest, welche durch endoproteolytische Spaltung von glykosylierten Proteinen entstanden sind) in verschiedensten biologischen Proben auf Omics-Niveau. Konventionelle, analytische Ansätze sehen häufig eine chemische oder enzymatische Freisetzung der Glykane vor, welche zu getrennten Oligosacchariden und deglykosylierten Peptiden führt. Allerdings ist die Analyse intakter Glykopeptide unerlässlich für die Verknüpfung zwischen Protein- und Glykanprofilen, sowie der Untersuchung der Glykan-Mikroheterogenität einer bestimmten Bindungsstelle. Trotz der technischen Schwierigkeiten, welche sich aus der heterogenen und komplexen Struktur von Glykanen und deren wenig informativem Fragmentierungsverhalten in der Tandem-Massenspektrometrie (MS/MS) ergeben, gab es erhebliche Verbesserungen in MS-basierter, intakter Glykopeptid-Charakterisierung. Hierzu zählen unter anderem Verbesserungen im Bereich der Glykopeptidanreicherung, neuen Fragmentierungsmethoden, optimierten MS Messmethoden und verbesserten Datenbank-Suchalgorithmen. Im Fokus dieser Verbesserungen stand aber meist die Qualität der intakten Glykopeptididentifikation und weniger die Quantifizierung. Bis heute bleibt die akkurate Quantifizierung intakter Glykopeptide im großen Maßstab eine technische Herausforderung und es bestehen weiterhin Defizite bei Programmen zur Verarbeitung quantitativer Daten.

Als Methoden der Wahl für die MS-basierte, quantitative Proteomik und Analyse von PTM haben sich, neben anderen Lösungen, Tandem Massenlabel (TMT) und isobarische Label für relative und absolute Quantifizierung (iTRAQ) als Methoden der Wahl etabliert. Diese Ansätze ermöglichen eine Nutzung für ein breites Spektrum unterschiedlicher Probenarten sowie die Möglichkeit zur Parallelisierung, welche die Messzeit und -varianz der Läufe reduziert. Ein „Poolen“ unterschiedlicher Proben verbessert zudem die allgemeine Sensitivität, besonders für Spezies geringeren Vorkommens. Die in jüngster Zeit entwickelte synchrone Präkursorauswahl Technologie (SPS-MS3) verbessert die quantitative Präzision durch die Verminderung der Interferenzen, welche durch Präkursor-Ko-Selektion entstehen können, auf Kosten der Scan-Geschwindigkeit. Mit ihren einzigartigen Vorteilen für die intakte Glykopeptidanalytik werden sowohl iTRAQ und TMT Reagenzien für quantitative Glykoproteomik Studien im großen Maßstab eingesetzt. Aktuelle Studien weißen darauf hin, dass sich physikalische und chemische Eigenschaften der Glykopeptide durch das Labelling verändern, was zu einer Verminderung der Zuverlässigkeit von Identifikation und Quantifizierung führen kann (siehe unten). Zudem macht das chemische Labelling die konventionelle Glykopeptid-Aufreinigung aufwändiger und komplizierter. Die aufgeführten technischen sowie softwareseitigen Limitierungen der Datenverarbeitung und -interpretation haben die Entwicklung eines robusten Workflows für die quantitative Glykoproteomik eingeschränkt.

Im Folgenden stellen wir SugarQuant vor, ein integraler Workflow für globale Glykopeptid Identifikation und Quantifizierung im großen Maßstab (Abb. 1). SugarQuant setzt sich zusammen aus (i) der Zelllyse unter Verwendung von SDS, (ii) der Proteinextraktion, (iii) Proteinanreicherung und endoproteolytischer Verdau mittels *Protein Aggregation Capture* (PAC) 17,18, (iv) parallelisiertes TMT Labelling, (v) N-Glykopeptidanreicherung mittels zwitterionischem HILIC (ZIC-HILIC) gefolgt von (vi) rudimentärer Umkehrphasen-Präfraktionierung und (vii) einer LC-MS3-Messung mittels Glyko-SPS-MS3, welche hochaufgelöste MS2 und MS3 Produktionenscans hervorbringt. Die Prozessierung der Daten mittels des neuen GlycoBinder Tools (iix) kombiniert MS2 und MS3 Fragmentionen für die N-Glykopeptid Identification und extrahiert die TMT Reporter-Ionenintensitäten aus den MS3 Scans für jeden identifizierten N-Glykopeptid-Spektrum-Match (GPSM). GlycoBinder konsolidiert redundante GPSMs mit ihren quantitativen Werten und meldet mehrdimensionale Quantifizierungsergebnisse für eindeutige Glykoformen, Glykosylierungsstellen und Glykanzusammensetzungen in einem zugänglichen tabellenbasierten Format.

Im Zuge einer Machbarkeitsstudie nutzten wir SugarQuant zum quantitativen Mapping der Proteinglykosylierung von Burkitt Lymphomzellen, welche mit unterschiedlichen Konzentrationen von 2-Desoxy-2-fluoro-L-fucose (2FF), einer Guanosindiphsophatfucose welche die zelluläre Fucosylierung unterbindet. Wir haben demonstriert, dass SugarQuant die hierdurch resultierende, niedrigere Expression fucosylierter N-Glykane, welche bei üblicherweise genutzten MS2 Methoden nicht detektiert wird, ortsspezifisch nachweisen kann. Unsere Ergebnisse deuten auf 2FF-sensitive N-Glykosylierungsstellen hin und zeigen 2FF-mediierte Änderungen der N-Glykosylierung essentieller Bestandteile der B-Zellenrezeptorsignaltransduktion auf. Zeitgleich zeigten diese 2FF-behandelten DG75 Proben bei der Analyse des Phosphoproteoms eine Beeinflussung der Phosphorylierung der BCR Effektoren durch 2FF. Die Kombination von N-Glyko- und Phosphoproteomik-Analysen zeigte eine gegenseitige Beeinflussung zwischen Phosphorylierung und N-Glykosylierung in Burkitt Lymphomzellen. 2FF-induzierte Defucosylierung an Glykanstrukturen führte zu Veränderungen der ortsspezifischen Glykoformen und der Gesamtbelegung der Glykosylierungsdömane. Darüber hinaus beeinflusst die 2FF-Behandlung auch die Phosphorylierung essenzieller Bestandteile, die an der Signalübertragung von B-Zell-Rezeptoren beteiligt sind. Folglich erweiterten unsere Ergebnisse das Verständnis der Rolle von PTMs bei der BCR-Signalübertragung in Lymphomen.

Quantitative N-Glykoproteomik Analyse von diffus-großzelligen B-Zell-Lymphomsubtypen mittels SugarQuant

Die Klassifizierung heterogener Krebssubtypen ist hilfreich und wichtig für die klinische Diagnostik und Behandlung: eine molekulare Klassifizierung von histologisch und morphologisch nicht unterscheidbaren diffusen großzelligen B-Zell-Lymphomen (DLBCL) in drei Subtypen, einschließlich aktiviertem B-Zell-ähnlichen (ABC) und Keimzentrum-B-Zell-ähnlichen (GCB) und nicht klassifizierten DLBCLs wurde durch Genexpressionsanalysen erreicht. Es wurde auch gezeigt, dass die konsistente Klassifizierung von DLBCL in ABC- und GCB-Subtypen, basierend auf ihrer Proteinexpression, unter Verwendung quantitativer Proteomanalysen auf SILAC-Basis möglich ist. Obwohl immer mehr Beweise den Zusammenhang zwischen aberranter Glykosylierung und Krebsentstehung zeigen, sind die genauen Mechanismen der Rezeptorregulation aufgrund ihrer Glykosylierung immer noch sehr schwer zu interpretieren. Dies begründet sich durch die inhärente Komplexität der Glykosylierungssynthese, die durch verschiedene Glykosyltransferasen vermittelt wird und der daraus resultierenden Mikro- und Makroheterogenität der Glykane. Eine frühere Studie zeigte, dass quantitative Glykodomänen-Belegungsprofile aus der Deglykoproteomikanalyse die

Klassifizierung der beiden Subtypen ABC- und GCB-DLBCL ermöglichten. Aufgrund der damaligen, technologischen Grenzen wurden deglykosylierte Peptide für die MS-Analyse verwendet, nachdem N-Glykane enzymatisch freigesetzt wurden, was zum Verlust der wichtigen Informationen der Glykankomponenten und der Verknüpfung zwischen Glykan und Protein führte. In dieser Studie untersuchten wir das ortsspezifische, quantitative N-Glykoproteom in den Subtypen von DLBCL mittels SugarQuant: mehr als 7000 einzigartige, intakte Glykopeptide (Glykoformen) wurden in fünf ABC- und vier GCB-DLBCL-Zelllinien quantifiziert. Ebenso war es möglich die DLBCL-Subtypen aufgrund ihres Glykopeptidexpressionsprofils zu unterscheiden. Die Mehrzahl dieser Glykoformen stammte von den wichtigsten BCR-Effektoren auf der Zelloberfläche wie IgM, CD79 und PTPRC. Wir fanden auch Unterschiede bei der intakten Glykopeptidexpression in demselben Subtyp von DLCBCL-Zelllinien: zum Beispiel einen unterschiedlichen Glykosylierungsgrad in WSU-FSCCL im Vergleich zu den anderen GCB-DLBCL-Zelllinien, sowie differenzielle Expression von fucosylierten und sialyierten Glykopeptiden in RIVA und OCI-Ly3 im Vergleich zu HBL1, TMD8 und OCI-Ly10. Darüber hinaus hat die Deglykoproteomikanalyse, unter Verwendung der Glykopeptide nach Entfernung der Glykanreste, die Genauigkeit der Quantifizierung intakter N-Glykopeptide mittels SugarQuant weiter bewiesen.

N-Glykoproteomanalyse von FUT8 Knock-Out und 2FF- behandelter TMD8 Zellen

Um die spezifische Veränderung der Expression von fucosylierten Glykopeptiden zu untersuchen, führten wir eine ortsspezifische, quantitative N-Glykoproteomikanalyse in 2FF-behandelten und FUT8-Knockout (KO) TMD8-Zellen unter Verwendung von SugarQuant durch. Die Fucosylierung (Bindung von Fucose an den Glykanrest) ist eine der wichtigsten Formen der Glykosylierung im Zusammenhang mit dem Fortschreiten einer Krebserkrankung. Fucose ist eine Desoxyhexose, eine Hexose welcher die Hydroxylgruppe am Kohlenstoff der sechsten Position fehlt. In unserem Labor haben die CRISPR-Dropoutscreendaten in ABC und GCB-DLBCL-Zelllinien gezeigt, dass FUT8 kein essenzielles Gen für das Überleben der TMD8-Zellen ist, sodass wir die Veränderung der Fucosylierung in FUT8 KO TMD8-Zellen untersuchen konnten. Bei der Behandlung der Zellen mit 2FF zeigte sich eine Unterdrückung der gesamten Fucosylierung, welche Core- und terminale/subterminale Fucosylierung umfasst. In den FUT8-KO TMD8-Zellen trat ausschließlich eine verringerte Core-Fucosylierung auf. In dieser Studie war die Korrelation der intakten Glycopeptid-Expression in 2FF-Behandlungen mit zwei verschiedenen Konzentrationen und in FUT8 KO mit zwei sgRNAs höher als unter den Bedingungen der 2FF-Behandlung und FUT8-KO, was die unterschiedliche Glycopeptidexpression unter den beiden Bedingungen der Fucosylierungshemmung darstellt. Unsere Ergebnisse zeigten auch, dass FUT8 die Synthese von terminaler/subterminaler Fucose an den Glykanrest reguliert und die Hemmung der Fucosylierung die Expression von sialyierten Glykopeptiden induzierte, was wahrscheinlich auf die kompetitive Bindung an dieselben Oligosaccharide während der terminalen/subterminalen Fucosylierung bzw. Sialylierung zurückzuführen ist. Diese Entdeckung hilft beim Verständnis der physiologischen und pathophysiologischen Prozesse, welche durch fucosylierte und sialylierte Glykosylierung reguliert werden.

Phosphoproteomanalyse in DLBCL mit BTK und SYK Inhibitoren

Die Funktion des B-Zell-Rezeptors (BCR) ist für B-Zellen von entscheidender Bedeutung: viele onkogene Mutationen und chromosomale Translokationen im BCR und den nachgeschalteten Effektoren unterstützen das Wachstum und Überleben von malignen B-Zellen. Bekannt ist bisher, dass die BCR-Signalübertragung hauptsächlich am Antigen-abhängigen NF-KB-Signalweg in ABC-

DLBCL beteiligt ist, wohingegen PI3K- und mTOR-vermittelte Signalwege in GCB-DLBCL aktiv sind. Die Tyrosinkinasen BTK und SYK sind die wichtigen Upstream-Effektoren bei der BCR-Signalübertragung in ABC-DLBCL, während nur SYK für GCB-DLBCL von entscheidender Bedeutung ist. Die Aufrechterhaltung der BCR-Signalübertragung wird über die Phosphorylierung und Dephosphorylierung von Effektoren im Signalweg reguliert. In dieser Studie untersuchten wir die systematische Änderung der Phosphorylierung von Effektoren nach Hemmung der vorgeschalteten BTK- und SYK-Effektoren in ABC-DLBCL und der SYK-Effektoren in GCB-DLBCL unter Verwendung einer SILAC-Phosphoproteomik-Strategie. Schließlich wurde ein großer Datensatz der quantifizierten Phosphorylierungsstellen in den verwendeten drei ABC- und vier GCB-Zelllinien nach der Inhibition von BTK und SYK generiert (Ibrutinib für die BTK-Inhibition, PRT-060318 für die SYK-Inhibition). Unsere Ergebnisse zeigten, dass die Phosphorylierungsregulation in DLBCL mit BTK- und SYK-Inhibition zelltypspezifisch übereinstimmt. Diese signifikant regulierten Phosphorylierungsereignisse in jeder Zelllinie mit unterschiedlichen Behandlungen waren an mehreren Reaktomwegen beteiligt. Darüber hinaus waren die, mit der Genregulation verbundenen, biologischen Prozesse in den verwendeten DLBCL-Zelllinien stark angereichert.

Ubiquitinomanalyse in DLBCL Zelllinien BTK und SYK Inhibitoren

Wir verwendeten den Di-Glycin-Lysin-spezifischen Antikörper (K-ε-GG), um die ubiquitinierten Peptide in Kombination mit der SILAC-Technologie anzureichern und die veränderten, ubiquitinierten Peptide in den DLBCL-Zelllinien nach BTK- und SYK-Hemmungen zu entdecken. Dies ermöglichte uns, einen quantitativen, Ubiquitin-modifizierten Proteom-Datensatz (Ubiquitinom) der behandelten Subtypen von DLBCL-Zelllinien zu generieren. Unsere Ergebnisse umfassten mehr als 15.000 Ubiquitylierungsstellen aus der Ubiquitinomanalyse von zwei mit Ibrutinib und PRT-062607 behandelten ABC- und einer GCB-Zelllinie. Zusätzlich stellten wir einen Datensatz differentiell regulierter ubiquitinierter Peptide für jede Zelllinie mit BTK- und SYK-Inhibition bereit. Die Analyse der differentiell, ubiquitinierten Peptiden in jeder Zelllinie nach BTK- und SYK-Inhibition ergab, dass die BCR-abhängige Ubiquitinierungsregulation auch in DLBCL zelllinienspezifisch war. Die ubiquitinierten Proteine, die diese signifikant regulierten Ubi-Peptide in den untersuchten Proben trugen, waren ebenfalls stark an Genregulationsprozessen beteiligt.

In Zusammenarbeit mit dem Labor von Prof. Dr. Louis M.Staudt (National Institutes of Health), verwendeten wir eine Bibliothek kleiner Leit-RNAs (sgRNAs), um CRISPR-Cas9-Screenings mit Funktionsverlust auf essentielle Gene in DLBCL-Zelllinien durchzuführen, die mit induzierbarem Cas9 konstruiert wurden. Bis jetzt haben wir bereits vier ABC-DLBCL-Zelllinien, fünf GCB-DLBCL-Zelllinien und drei multiple Myelomzelllinien als Kontrollen gescreent. CRISPR-Dropout- Screenings helfen uns, die Überlebensregulatoren in jeder DLBCL-Zelllinie zu finden. Für jedes Gen führen wir eine grundlegende Diagnose des abgeleiteten CRISPR-Screen-Scores (CSS) aus einer Statistik auf Genebene durch. CSS ist die Anzahl der Standardabweichungen vom durchschnittlichen Effekt der Inaktivierung eines Gens. In dieser Studie führten wir globale Phosphoproteom- und Ubiquitinomanalyse in DLBCL-Zelllinien durch, die mit zwei Tyrosinkinase-Inhibitoren (dem BTK-Inhibitor Ibrutinib und dem SYK-Inhibitor PRT-060318) unter Verwendung der quantitativen Proteomics-Technologie auf SILAC-Basis behandelt wurden. Wir haben erfolgreich mehrere differentiell phosporylierte und ubiquitinierte Stellen für jeden Zustand unter Verwendung eines sensitiven Differentialdetektionsalgorithmus identifiziert. Es ermöglichte uns, nach potenziellen BCR-Signaleffektoren innerhalb dieser Phosphorylierungs- und Ubiquitinierungs-modifizierten Proteinen zu suchen. Durch die Integration der globalen Phosphoproteom- und Ubiquitinom-Ergebnisse mit den

CRISPR-Screen-Daten wurden essentielle Gen-Treffer gefunden, die eine signifikant unterschiedliche Phosphorylierung und Ubiquitinierung aufwiesen. Die gefilterten Gentreffer wirken höchstwahrscheinlich als Überlebensregulatoren im BCR-Netzwerk. Die integrative Analyse ermöglicht ein umfassenderes Verständnis der molekularen Landschaft des Lymphoms und bietet die Möglichkeit einer genaueren Diagnose und Behandlung.

Zusammenfassend haben wir eine robuste, genaue und ortsspezifische, quantitative N-Glykoproteomik-Methode entwickelt. Sie wurde verwendet, um die intakte N-Glykopeptidexpression in Zelllinien von DLBCL Subtypen und die Expression von fucosylierten Glykopeptiden in TMD8-Zelllinien nach Inhibition der Fucosylierung zu untersuchen. Darüber hinaus haben wir die Phosphorylierungs- und Ubiquitylierungsänderungen in DLBCL-Subtypen bei der Inhibition der Tyrosinkinasen BTK und SYK durch quantitative Phosphoproteom- und Ubiquitinomanalyse auf SILAC-Basis charakterisiert. Diese lieferten wertvolle Datensätze für die Entdeckung unbekannter Effektoren und führten zu einem besseren Verständnis der BCR-Signalnetzwerke.

8. References

1. Birbrair, A. and P.S. Frenette, *Niche heterogeneity in the bone marrow.* Ann N Y Acad Sci, 2016. **1370**(1): p. 82-96.
2. Morrison, S.J. and J. Kimble, *Asymmetric and symmetric stem-cell divisions in development and cancer.* Nature, 2006. **441**(7097): p. 1068-74.
3. Baldwin, C.L., *Immune response overview.* Vet Microbiol, 2002. **90**(1-4): p. 365-6.
4. Lenz, G. and L.M. Staudt, *Mechanisms of Disease: Aggressive Lymphomas.* New England Journal of Medicine, 2010. **362**(15): p. 1417-1429.
5. Arber, D.A., et al., *The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia.* Blood, 2016. **127**(20): p. 2391-405.
6. McGuire, S., *World Cancer Report 2014. Geneva, Switzerland: World Health Organization, International Agency for Research on Cancer, WHO Press, 2015.* Adv Nutr, 2016. **7**(2): p. 418-9.
7. Juliusson, G. and R. Hough, *Leukemia.* Prog Tumor Res, 2016. **43**: p. 87-100.
8. Barbui, T., et al., *The 2016 WHO classification and diagnostic criteria for myeloproliferative neoplasms: document summary and in-depth discussion.* Blood Cancer J, 2018. **8**(2): p. 15.
9. Hsi, E.D., *2016 WHO Classification update-What's new in lymphoid neoplasms.* Int J Lab Hematol, 2017. **39 Suppl 1**: p. 14-22.
10. Armitage, J.O., et al., *Non-Hodgkin lymphoma.* Lancet, 2017. **390**(10091): p. 298-310.
11. Gobbi, P.G., et al., *Hodgkin lymphoma.* Crit Rev Oncol Hematol, 2013. **85**(2): p. 216-37.
12. Lenz, G. and L.M. Staudt, *Aggressive lymphomas.* N Engl J Med, 2010. **362**(15): p. 1417-29.
13. Shaffer, A.L., 3rd, R.M. Young, and L.M. Staudt, *Pathogenesis of human B cell lymphomas.* Annu Rev Immunol, 2012. **30**: p. 565-610.
14. Raab, M.S., et al., *Multiple myeloma.* Lancet, 2009. **374**(9686): p. 324-39.
15. Molyneux, E.M., et al., *Burkitt's lymphoma.* Lancet, 2012. **379**(9822): p. 1234-44.
16. Burkitt, D., *A sarcoma involving the jaws in African children.* Br J Surg, 1958. **46**(197): p. 218-23.
17. O'Conor, G.T. and J.N. Davies, *Malignant tumors in African children. With special reference to malignant lymphoma.* J Pediatr, 1960. **56**: p. 526-35.
18. Epstein, M.A., B.G. Achong, and Y.M. Barr, *Virus Particles in Cultured Lymphoblasts from Burkitt's Lymphoma.* Lancet, 1964. **1**(7335): p. 702-3.
19. Morrow, R.H., et al., *Burkitt's lymphoma in the Mengo Districts of Uganda: epidemiologic features and their relationship to malaria.* J Natl Cancer Inst, 1976. **56**(3): p. 479-83.
20. Burkitt, D.P., *Observations on the geography of malignant lymphoma.* East Afr Med J, 1961. **38**: p. 511-4.
21. Panea, R.I., et al., *The whole-genome landscape of Burkitt lymphoma subtypes.* Blood, 2019. **134**(19): p. 1598-1607.
22. Lovisa, F., et al., *IGH and IGK gene rearrangements as PCR targets for pediatric Burkitt's lymphoma and mature B-ALL MRD analysis.* Laboratory Investigation, 2009. **89**(10): p. 1182-1186.
23. Schmitz, R., et al., *Oncogenic mechanisms in Burkitt lymphoma.* Cold Spring Harb Perspect Med, 2014. **4**(2).
24. Dalla-Favera, R., et al., *Human c-myc onc gene is located on the region of chromosome 8 that is translocated in Burkitt lymphoma cells.* Proc Natl Acad Sci U S A, 1982. **79**(24): p. 7824-7.
25. Schmitz, R., et al., *Burkitt lymphoma pathogenesis and therapeutic targets from structural and functional genomics.* Nature, 2012. **490**(7418): p. 116-20.

26. Meyer, N., S.S. Kim, and L.Z. Penn, *The Oscar-worthy role of Myc in apoptosis.* Semin Cancer Biol, 2006. **16**(4): p. 275-87.
27. Murre, C., P.S. McCaw, and D. Baltimore, *A new DNA binding and dimerization motif in immunoglobulin enhancer binding, daughterless, MyoD, and myc proteins.* Cell, 1989. **56**(5): p. 777-83.
28. Richter, J., et al., *Recurrent mutation of the ID3 gene in Burkitt lymphoma identified by integrated genome, exome and transcriptome sequencing.* Nat Genet, 2012. **44**(12): p. 1316-20.
29. Love, C., et al., *The genetic landscape of mutations in Burkitt lymphoma.* Nat Genet, 2012. **44**(12): p. 1321-5.
30. Schmitz, R., et al., *Genetics and Pathogenesis of Diffuse Large B-Cell Lymphoma.* N Engl J Med, 2018. **378**(15): p. 1396-1407.
31. Kaymaz, Y., et al., *Comprehensive Transcriptome and Mutational Profiling of Endemic Burkitt Lymphoma Reveals EBV Type-Specific Differences.* Mol Cancer Res, 2017. **15**(5): p. 563-576.
32. Grande, B.M., et al., *Genome-wide discovery of somatic coding and noncoding mutations in pediatric endemic and sporadic Burkitt lymphoma.* Blood, 2019. **133**(12): p. 1313-1324.
33. Siegel, R.L., K.D. Miller, and A. Jemal, *Cancer statistics, 2019.* CA Cancer J Clin, 2019. **69**(1): p. 7-34.
34. Alizadeh, A.A., et al., *Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling.* Nature, 2000. **403**(6769): p. 503-11.
35. Wright, G., et al., *A gene expression-based method to diagnose clinically distinct subgroups of diffuse large B cell lymphoma.* Proceedings of the National Academy of Sciences of the United States of America, 2003. **100**(17): p. 9991-9996.
36. Davis, R.E., et al., *Constitutive nuclear factor kappaB activity is required for survival of activated B cell-like diffuse large B cell lymphoma cells.* J Exp Med, 2001. **194**(12): p. 1861-74.
37. Read, J.A., et al., *Evaluating cell-of-origin subtype methods for predicting diffuse large B-cell lymphoma survival: a meta-analysis of gene expression profiling and immunohistochemistry algorithms.* Clin Lymphoma Myeloma Leuk, 2014. **14**(6): p. 460-467 e2.
38. Young, R.M., et al., *Taming the Heterogeneity of Aggressive Lymphomas for Precision Therapy.* Annual Review of Cancer Biology, 2019. **3**(1): p. 429-455.
39. Wilson, W.H., et al., *Targeting B cell receptor signaling with ibrutinib in diffuse large B cell lymphoma.* Nat Med, 2015. **21**(8): p. 922-6.
40. Phelan, J.D., et al., *A multiprotein supercomplex controlling oncogenic signalling in lymphoma.* Nature, 2018. **560**(7718): p. 387-391.
41. Murphy, K., et al., *Janeway's immunobiology*. 2012, New York: Garland Science.
42. Clark, M.R., et al., *Receptors, subcellular compartments and the regulation of peripheral B cell responses: the illuminating state of anergy.* Mol Immunol, 2011. **48**(11): p. 1281-6.
43. Dogan, I., et al., *Multiple layers of B cell memory with different effector functions.* Nat Immunol, 2009. **10**(12): p. 1292-9.
44. Davis, R.E., et al., *Chronic active B-cell-receptor signalling in diffuse large B-cell lymphoma.* Nature, 2010. **463**(7277): p. 88-92.
45. Davis, R.E., et al., *Constitutive nuclear factor kappa B activity is required for survival of activated B cell-like diffuse large B cell lymphoma cells.* Journal of Experimental Medicine, 2001. **194**(12): p. 1861-1874.
46. Lenz, G., et al., *Oncogenic CARD11 mutations in human diffuse large B cell lymphoma.* Science, 2008. **319**(5870): p. 1676-9.

47. Agathangelidis, A., et al., *Stereotyped B-cell receptors in one-third of chronic lymphocytic leukemia: a molecular classification with implications for targeted therapies.* Blood, 2012. **119**(19): p. 4467-4475.
48. Hadzidimitriou, A., et al., *Is there a role for antigen selection in mantle cell lymphoma? Immunogenetic support from a series of 807 cases.* Blood, 2011. **118**(11): p. 3088-3095.
49. Young, R.M., et al., *Survival of human lymphoma cells requires B-cell receptor engagement by self-antigens.* Proc Natl Acad Sci U S A, 2015. **112**(44): p. 13447-54.
50. Young, R.M., et al., *Pathogenic B-cell receptor signaling in lymphoid malignancies: New insights to improve treatment.* Immunol Rev, 2019. **291**(1): p. 190-213.
51. Ngo, V.N., et al., *Oncogenically active MYD88 mutations in human lymphoma.* Nature, 2011. **470**(7332): p. 115-9.
52. Soderberg, O., et al., *Characterizing proteins and their interactions in cells and tissues using the in situ proximity ligation assay.* Methods, 2008. **45**(3): p. 227-232.
53. Sancak, Y., et al., *Ragulator-Rag Complex Targets mTORC1 to the Lysosomal Surface and Is Necessary for Its Activation by Amino Acids.* Cell, 2010. **141**(2): p. 290-303.
54. Pongas, G.N., C.M. Annunziata, and L.M. Staudt, *PI3Kdelta inhibition causes feedback activation of PI3Kalpha in the ABC subtype of diffuse large B-cell lymphoma.* Oncotarget, 2017. **8**(47): p. 81794-81802.
55. Griner, L.A.M., et al., *High-throughput combinatorial screening identifies drugs that cooperate with ibrutinib to kill activated B-cell-like diffuse large B-cell lymphoma cells.* Proceedings of the National Academy of Sciences of the United States of America, 2014. **111**(6): p. 2349-2354.
56. Ezell, S.A., et al., *Synergistic induction of apoptosis by combination of BTK and dual mTORC1/2 inhibitors in diffuse large B cell lymphoma.* Oncotarget, 2014. **5**(13): p. 4990-5001.
57. Chen, L.F., et al., *SYK Inhibition Modulates Distinct PI3K/AKT-Dependent Survival Pathways and Cholesterol Biosynthesis in Diffuse Large B Cell Lymphomas.* Cancer Cell, 2013. **23**(6): p. 826-838.
58. Kraus, M., et al., *Survival of resting mature B lymphocytes depends on BCR signaling via the Igalpha/beta heterodimer.* Cell, 2004. **117**(6): p. 787-800.
59. Rickert, R.C., K. Rajewsky, and J. Roes, *Impairment of T-Cell-Dependent B-Cell Responses and B-1 Cell-Development in Cd19-Deficient Mice.* Nature, 1995. **376**(6538): p. 352-355.
60. Engel, P., et al., *Abnormal B-Lymphocyte Development, Activation, and Differentiation in Mice That Lack or Overexpress the Cd19 Signal-Transduction Molecule.* Immunity, 1995. **3**(1): p. 39-50.
61. Lionakis, M.S., et al., *Inhibition of B Cell Receptor Signaling by Ibrutinib in Primary CNS Lymphoma.* Cancer Cell, 2017. **31**(6): p. 833-843 e5.
62. Wu, J., M. Zhang, and D. Liu, *Acalabrutinib (ACP-196): a selective second-generation BTK inhibitor.* J Hematol Oncol, 2016. **9**: p. 21.
63. Mato, A.R., et al., *Toxicities and outcomes of 616 ibrutinib-treated patients in the United States: a real-world analysis.* Haematologica, 2018. **103**(5): p. 874-879.
64. Cheng, S., et al., *SYK inhibition and response prediction in diffuse large B-cell lymphoma.* Blood, 2011. **118**(24): p. 6342-52.
65. Young, R.M., et al., *Mouse models of non-Hodgkin lymphoma reveal Syk as an important therapeutic target.* Blood, 2009. **113**(11): p. 2508-2516.
66. Chen, L., et al., *SYK-dependent tonic B-cell receptor signaling is a rational treatment target in diffuse large B-cell lymphoma.* Blood, 2008. **111**(4): p. 2230-7.

67. Gobessi, S., et al., *Inhibition of constitutive and BCR-induced Syk activation downregulates Mcl-1 and induces apoptosis in chronic lymphocytic leukemia B cells.* Leukemia, 2009. **23**(4): p. 686-97.
68. Naylor, T.L., et al., *Protein kinase C inhibitor sotrastaurin selectively inhibits the growth of CD79 mutant diffuse large B-cell lymphomas.* Cancer Res, 2011. **71**(7): p. 2643-53.
69. Moremen, K.W., M. Tiemeyer, and A.V. Nairn, *Vertebrate protein glycosylation: diversity, synthesis and function.* Nat Rev Mol Cell Biol, 2012. **13**(7): p. 448-62.
70. Ohtsubo, K. and J.D. Marth, *Glycosylation in cellular mechanisms of health and disease.* Cell, 2006. **126**(5): p. 855-67.
71. Reily, C., et al., *Glycosylation in health and disease.* Nat Rev Nephrol, 2019. **15**(6): p. 346-366.
72. Freeze, H.H. and M. Aebi, *Altered glycan structures: the molecular basis of congenital disorders of glycosylation.* Curr Opin Struct Biol, 2005. **15**(5): p. 490-8.
73. Drake, P.M., et al., *Sweetening the Pot: Adding Glycosylation to the Biomarker Discovery Equation.* Clinical Chemistry, 2010. **56**(2): p. 223-236.
74. Alley, W.R., Jr., B.F. Mann, and M.V. Novotny, *High-sensitivity analytical approaches for the structural characterization of glycoproteins.* Chem Rev, 2013. **113**(4): p. 2668-732.
75. Suzuki, O., *Glycosylation in lymphoma: Biology and glycotherapy.* Pathol Int, 2019. **69**(8): p. 441-449.
76. Pang, X., et al., *Multiple Roles of Glycans in Hematological Malignancies.* Front Oncol, 2018. **8**: p. 364.
77. Suzuki, O., M. Abe, and Y. Hashimoto, *Sialylation by beta-galactoside alpha-2,6-sialyltransferase and N-glycans regulate cell adhesion and invasion in human anaplastic large cell lymphoma.* Int J Oncol, 2015. **46**(3): p. 973-80.
78. Suzuki, O. and M. Abe, *Galectin-1-mediated cell adhesion, invasion and cell death in human anaplastic large cell lymphoma: regulatory roles of cell surface glycans.* Int J Oncol, 2014. **44**(5): p. 1433-42.
79. Deeb, S.J., et al., *N-linked Glycosylation Enrichment for In-depth Cell Surface Proteomics of Diffuse Large B-cell Lymphoma Subtypes.* Molecular & Cellular Proteomics, 2014. **13**(1): p. 240-251.
80. Shatnyeva, O.M., et al., *Modulation of the CD95-induced apoptosis: the role of CD95 N-glycosylation.* PLoS One, 2011. **6**(5): p. e19927.
81. Vuillier, F., et al., *Lower levels of surface B-cell-receptor expression in chronic lymphocytic leukemia are associated with glycosylation and folding defects of the mu and CD79a chains.* Blood, 2005. **105**(7): p. 2933-40.
82. Mizuochi, T., et al., *Structural and numerical variations of the carbohydrate moiety of immunoglobulin G.* J Immunol, 1982. **129**(5): p. 2016-20.
83. Wasim, L., et al., *N-Linked Glycosylation Regulates CD22 Organization and Function.* Front Immunol, 2019. **10**: p. 699.
84. Cao, A., et al., *Galectin-9 binds IgM-BCR to regulate B cell signaling.* Nat Commun, 2018. **9**(1): p. 3288.
85. Tomaszewska, R., et al., *Sialic acid concentration in different stages of malignant lymphoma and leukemia in children.* Acta Paediatr Jpn, 1997. **39**(4): p. 448-50.
86. Suzuki, O., et al., *Alpha-2,6-sialylation of L-PHA reactive oligosaccharides and expression of N-acetylglucosaminyltransferase V in human diffuse large B cell lymphoma.* Oncol Rep, 2003. **10**(6): p. 1759-64.

87. Suzuki, O., et al., *UDP-GlcNAc2-epimerase regulates cell surface sialylation and cell adhesion to extracellular matrix in Burkitt's lymphoma.* International Journal of Oncology, 2002. **20**(5): p. 1005-1011.
88. Levene, P. and C. Alsberg, *The cleavage products of vitellin.* Journal of Biological Chemistry, 1906. **2**(1): p. 127-133.
89. Ardito, F., et al., *The crucial role of protein phosphorylation in cell signaling and its use as targeted therapy (Review).* Int J Mol Med, 2017. **40**(2): p. 271-280.
90. Dal Porto, J.M., et al., *B cell antigen receptor signaling 101.* Mol Immunol, 2004. **41**(6-7): p. 599-613.
91. Ciechanover, A., *The ubiquitin-proteasome proteolytic pathway.* Cell, 1994. **79**(1): p. 13-21.
92. Hershko, A. and A. Ciechanover, *The ubiquitin system.* Annu Rev Biochem, 1998. **67**: p. 425-79.
93. Kirisako, T., et al., *A ubiquitin ligase complex assembles linear polyubiquitin chains.* EMBO J, 2006. **25**(20): p. 4877-87.
94. Nijman, S.M., et al., *A genomic and functional inventory of deubiquitinating enzymes.* Cell, 2005. **123**(5): p. 773-86.
95. Thys, A., T. Douanne, and N. Bidere, *Post-translational Modifications of the CARMA1-BCL10-MALT1 Complex in Lymphocytes and Activated B-Cell Like Subtype of Diffuse Large B-Cell Lymphoma.* Front Oncol, 2018. **8**: p. 498.
96. Yang, Y.B. and L.M. Staudt, *Protein ubiquitination in lymphoid malignancies.* Immunological Reviews, 2015. **263**(1): p. 240-256.
97. Aebersold, R. and M. Mann, *Mass spectrometry-based proteomics.* Nature, 2003. **422**(6928): p. 198-207.
98. Karas, M., D. Bachmann, and F. Hillenkamp, *Influence of the Wavelength in High-Irradiance Ultraviolet-Laser Desorption Mass-Spectrometry of Organic-Molecules.* Analytical Chemistry, 1985. **57**(14): p. 2935-2939.
99. Fenn, J.B., et al., *Electrospray ionization for mass spectrometry of large biomolecules.* Science, 1989. **246**(4926): p. 64-71.
100. Makarov, A., *Electrostatic axially harmonic orbital trapping: a high-performance technique of mass analysis.* Anal Chem, 2000. **72**(6): p. 1156-62.
101. Scigelova, M. and A. Makarov, *Orbitrap mass analyzer - Overview and applications in proteomics.* Proteomics, 2006: p. 16-21.
102. Scheltema, R.A., et al., *The Q Exactive HF, a Benchtop mass spectrometer with a pre-filter, high-performance quadrupole and an ultra-high-field Orbitrap analyzer.* Mol Cell Proteomics, 2014. **13**(12): p. 3698-708.
103. Makarov, A., E. Denisov, and O. Lange, *Performance evaluation of a high-field Orbitrap mass analyzer.* J Am Soc Mass Spectrom, 2009. **20**(8): p. 1391-6.
104. Kelstrup, C.D., et al., *Performance Evaluation of the Q Exactive HF-X for Shotgun Proteomics.* J Proteome Res, 2018. **17**(1): p. 727-738.
105. Wells, J.M. and S.A. McLuckey, *Collision-induced dissociation (CID) of peptides and proteins.* Biological Mass Spectrometry, 2005. **402**: p. 148-185.
106. Roepstorff, P. and J. Fohlman, *Proposal for a common nomenclature for sequence ions in mass spectra of peptides.* Biomed Mass Spectrom, 1984. **11**(11): p. 601.
107. Schwartz, J.C., M.W. Senko, and J.E. Syka, *A two-dimensional quadrupole ion trap mass spectrometer.* J Am Soc Mass Spectrom, 2002. **13**(6): p. 659-69.
108. Olsen, J.V., et al., *Higher-energy C-trap dissociation for peptide modification analysis.* Nat Methods, 2007. **4**(9): p. 709-12.

109. Good, D.M., et al., *Performance characteristics of electron transfer dissociation mass spectrometry.* Mol Cell Proteomics, 2007. **6**(11): p. 1942-51.
110. Coon, J.J., et al., *Tandem mass spectrometry for peptide and protein sequence analysis.* Biotechniques, 2005. **38**(4): p. 519, 521, 523.
111. Syka, J.E., et al., *Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry.* Proc Natl Acad Sci U S A, 2004. **101**(26): p. 9528-33.
112. Medzihradszky, K.F., *Peptide sequence analysis.* Methods Enzymol, 2005. **402**: p. 209-44.
113. Cox, J., et al., *Andromeda: a peptide search engine integrated into the MaxQuant environment.* J Proteome Res, 2011. **10**(4): p. 1794-805.
114. Bhargava, M., et al., *Application of clinical proteomics in acute respiratory distress syndrome.* Clin Transl Med, 2014. **3**(1): p. 34.
115. Aebersold, R. and M. Mann, *Mass-spectrometric exploration of proteome structure and function.* Nature, 2016. **537**(7620): p. 347-55.
116. King, R., et al., *Mechanistic investigation of ionization suppression in electrospray ionization.* J Am Soc Mass Spectrom, 2000. **11**(11): p. 942-50.
117. Dongre, A.R., et al., *Influence of peptide composition, gas-phase basicity, and chemical modification on fragmentation efficiency: Evidence for the mobile proton model.* Journal of the American Chemical Society, 1996. **118**(35): p. 8365-8374.
118. Schilling, B., et al., *Platform-independent and Label-free Quantitation of Proteomic Data Using MS1 Extracted Ion Chromatograms in Skyline APPLICATION TO PROTEIN ACETYLATION AND PHOSPHORYLATION.* Molecular & Cellular Proteomics, 2012. **11**(5): p. 202-214.
119. Choi, H., D. Fermin, and A.I. Nesvizhskii, *Significance analysis of spectral count data in label-free shotgun proteomics.* Mol Cell Proteomics, 2008. **7**(12): p. 2373-85.
120. Webb-Robertson, B.J., et al., *Review, evaluation, and discussion of the challenges of missing value imputation for mass spectrometry-based label-free global proteomics.* J Proteome Res, 2015. **14**(5): p. 1993-2001.
121. Gevaert, K., et al., *Stable isotopic labeling in proteomics.* Proteomics, 2008. **8**(23-24): p. 4873-85.
122. Fenselau, C. and X. Yao, *18O2-labeling in quantitative proteomic strategies: a status report.* J Proteome Res, 2009. **8**(5): p. 2140-3.
123. Gygi, S.P., et al., *Quantitative analysis of complex protein mixtures using isotope-coded affinity tags.* Nature Biotechnology, 1999. **17**(10): p. 994-999.
124. Ross, P.L., et al., *Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents.* Molecular & Cellular Proteomics, 2004. **3**(12): p. 1154-1169.
125. Thompson, A., et al., *Tandem mass tags: A novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS.* Analytical Chemistry, 2003. **75**(8): p. 1895-1904.
126. Ting, L., et al., *MS3 eliminates ratio distortion in isobaric multiplexed quantitative proteomics.* Nat Methods, 2011. **8**(11): p. 937-40.
127. McAlister, G.C., et al., *MultiNotch MS3 Enables Accurate, Sensitive, and Multiplexed Detection of Differential Expression across Cancer Cell Line Proteomes.* Analytical Chemistry, 2014. **86**(14): p. 7150-7158.
128. Ong, S.E. and M. Mann, *Stable isotope labeling by amino acids in cell culture for quantitative proteomics.* Methods Mol Biol, 2007. **359**: p. 37-52.

129. Ong, S.E. and M. Mann, *A practical recipe for stable isotope labeling by amino acids in cell culture (SILAC).* Nat Protoc, 2006. **1**(6): p. 2650-60.
130. Ong, S.E., et al., *Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics.* Mol Cell Proteomics, 2002. **1**(5): p. 376-86.
131. Olsen, J.V. and M. Mann, *Status of large-scale analysis of post-translational modifications by mass spectrometry.* Mol Cell Proteomics, 2013. **12**(12): p. 3444-52.
132. Hennrich, M.L. and A.C. Gavin, *Quantitative mass spectrometry of posttranslational modifications: keys to confidence.* Sci Signal, 2015. **8**(371): p. re5.
133. Alvarez-Manilla, G., et al., *Tools for glycoproteomic analysis: Size exclusion chromatography facilitates identification of tryptic glycopeptides with N-linked glycosylation sites.* Journal of Proteome Research, 2006. **5**(3): p. 701-708.
134. Zielinska, D.F., et al., *Precision mapping of an in vivo N-glycoproteome reveals rigid topological and sequence constraints.* Cell, 2010. **141**(5): p. 897-907.
135. Kaji, H., et al., *Lectin affinity capture, isotope-coded tagging and mass spectrometry to identify N-linked glycoproteins.* Nat Biotechnol, 2003. **21**(6): p. 667-72.
136. Palmisano, G., et al., *Selective enrichment of sialic acid-containing glycopeptides using titanium dioxide chromatography with analysis by HILIC and mass spectrometry.* Nature Protocols, 2010. **5**(12): p. 1974-1982.
137. Larsen, M.R., et al., *Exploring the sialiome using titanium dioxide chromatography and mass spectrometry.* Molecular & Cellular Proteomics, 2007. **6**(10): p. 1778-1787.
138. Chen, J., P. Shah, and H. Zhang, *Solid Phase Extraction of N-Linked Glycopeptides Using Hydrazide Tip.* Analytical Chemistry, 2013. **85**(22): p. 10670-10674.
139. Zhang, H., et al., *Identification and quantification of N-linked glycoproteins using hydrazide chemistry, stable isotope labeling and mass spectrometry.* Nature Biotechnology, 2003. **21**(6): p. 660-666.
140. Mysling, S., et al., *Utilizing ion-pairing hydrophilic interaction chromatography solid phase extraction for efficient glycopeptide enrichment in glycoproteomics.* Anal Chem, 2010. **82**(13): p. 5598-609.
141. Hagglund, P., et al., *A new strategy for identification of N-glycosylated proteins and unambiguous assignment of their glycosylation sites using HILIC enrichment and partial deglycosylation.* Journal of Proteome Research, 2004. **3**(3): p. 556-566.
142. Yang, Z.P. and W.S. Hancock, *Monitoring glycosylation pattern changes of glycoproteins using multi-lectin affinity chromatography.* Journal of Chromatography A, 2005. **1070**(1-2): p. 57-64.
143. Zhang, C., et al., *Evaluation of Different N-Glycopeptide Enrichment Methods for N-Glycosylation Sites Mapping in Mouse Brain.* J Proteome Res, 2016. **15**(9): p. 2960-8.
144. Fang, P., et al., *In-depth mapping of the mouse brain N-glycoproteome reveals widespread N-glycosylation of diverse brain proteins.* Oncotarget, 2016. **7**(25): p. 38796-38809.
145. Dodds, E.D., *Gas-phase dissociation of glycosylated peptide ions.* Mass Spectrom Rev, 2012. **31**(6): p. 666-82.
146. Nilsson, J., *Liquid chromatography-tandem mass spectrometry-based fragmentation analysis of glycopeptides.* Glycoconj J, 2016. **33**(3): p. 261-72.
147. Chen, Z., J. Huang, and L. Li, *Recent advances in mass spectrometry (MS)-based glycoproteomics in complex biological samples.* Trends Analyt Chem, 2019. **118**: p. 880-892.
148. Thaysen-Andersen, M., N.H. Packer, and B.L. Schulz, *Maturing Glycoproteomics Technologies Provide Unique Structural Insights into the N-glycoproteome and Its Regulation in Health and Disease.* Mol Cell Proteomics, 2016. **15**(6): p. 1773-90.

149. Turiak, L., et al., *Site-specific N-glycosylation of HeLa cell glycoproteins.* Sci Rep, 2019. **9**(1): p. 14822.
150. Hinneburg, H., et al., *The Art of Destruction: Optimizing Collision Energies in Quadrupole-Time of Flight (Q-TOF) Instruments for Glycopeptide-Based Glycoproteomics.* Journal of the American Society for Mass Spectrometry, 2016. **27**(3): p. 507-519.
151. Stadlmann, J., et al., *Comparative glycoproteomics of stem cells identifies new players in ricin toxicity.* Nature, 2017. **advance online publication**.
152. Zeng, W.F., et al., *pGlyco: a pipeline for the identification of intact N-glycopeptides by using HCD- and CID-MS/MS and MS3.* Sci Rep, 2016. **6**: p. 25102.
153. Sun, S., et al., *Comprehensive analysis of protein glycosylation by solid-phase extraction of N-linked glycans and glycosite-containing peptides.* Nat Biotechnol, 2016. **34**(1): p. 84-8.
154. Liu, M.Q., et al., *pGlyco 2.0 enables precision N-glycoproteomics with comprehensive quality control and one-step mass spectrometry for intact glycopeptide identification.* Nat Commun, 2017. **8**(1): p. 438.
155. Bollineni, R.C., et al., *Large-scale intact glycopeptide identification by Mascot database search.* Sci Rep, 2018. **8**(1): p. 2117.
156. Riley, N.M., et al., *Capturing site-specific heterogeneity with large-scale N-glycoproteome analysis.* Nat Commun, 2019. **10**(1): p. 1311.
157. Glover, M.S., et al., *Characterization of intact sialylated glycopeptides and phosphorylated glycopeptides from IMAC enriched samples by EThcD fragmentation: Toward combining phosphoproteomics and glycoproteomics.* International Journal of Mass Spectrometry, 2018. **427**: p. 35-42.
158. Yu, Q., et al., *Electron-Transfer/Higher-Energy Collision Dissociation (EThcD)-Enabled Intact Glycopeptide/Glycoproteome Characterization.* J Am Soc Mass Spectrom, 2017.
159. Brown, C.J., et al., *Glycoproteome Analysis of Human Serum and Brain Tissue.* 2019.
160. Ranzinger, R., et al., *GlycomeDB - integration of open-access carbohydrate structure databases.* Bmc Bioinformatics, 2008. **9**.
161. Woodin, C.L., et al., *GlycoPep grader: a web-based utility for assigning the composition of N-linked glycopeptides.* Anal Chem, 2012. **84**(11): p. 4821-9.
162. Chandler, K.B., et al., *Exploring Site-Specific N-Glycosylation Microheterogeneity of Haptoglobin Using Glycopeptide CID Tandem Mass Spectra and Glycan Database Search.* Journal of Proteome Research, 2013. **12**(8): p. 3652-3666.
163. Wu, S.W., et al., *Sweet-Heart - an integrated suite of enabling computational tools for automated MS2/MS3 sequencing and identification of glycopeptides.* J Proteomics, 2013. **84**: p. 1-16.
164. Lynn, K.S., et al., *MAGIC: an automated N-linked glycoprotein identification tool using a Y1-ion pattern matching algorithm and in silico MS(2) approach.* Anal Chem, 2015. **87**(4): p. 2466-73.
165. Zhu, Z.K., et al., *GlycoPep Detector: A Tool for Assigning Mass Spectrometry Data of N-Linked Glycopeptides on the Basis of Their Electron Transfer Dissociation Spectra.* Analytical Chemistry, 2013. **85**(10): p. 5023-5032.
166. Zhu, Z.K., et al., *New Glycoproteomics Software, GlycoPep Evaluator, Generates Decoy Glycopeptides de Novo and Enables Accurate False Discovery Rate Analysis for Small Data Sets.* Analytical Chemistry, 2014. **86**(18): p. 9212-9219.
167. Mayampurath, A., et al., *Computational Framework for Identification of Intact Glycopeptides in Complex Samples.* Analytical Chemistry, 2014. **86**(1): p. 453-463.

168. He, L., et al., *GlycoMaster DB: Software To Assist the Automated Identification of N-Linked Glycopeptides by Tandem Mass Spectrometry.* Journal of Proteome Research, 2014. **13**(9): p. 3881-3895.
169. Bern, M., Y.J. Kil, and C. Becker, *Byonic: advanced peptide and protein identification software.* Curr Protoc Bioinformatics, 2012. **Chapter 13**: p. Unit13 20.
170. Toghi Eshghi, S., et al., *GPQuest: A Spectral Library Matching Algorithm for Site-Specific Assignment of Tandem Mass Spectra to Intact N-glycopeptides.* Anal Chem, 2015. **87**(10): p. 5181-8.
171. Park, G.W., et al., *Integrated GlycoProteome Analyzer (I-GPA) for Automated Identification and Quantitation of Site-Specific N-Glycosylation.* Sci Rep, 2016. **6**: p. 21175.
172. An, Z., et al., *N-Linked Glycopeptide Identification Based on Open Mass Spectral Library Search.* Biomed Res Int, 2018. **2018**: p. 1564136.
173. Liu, G., et al., *A comprehensive, open-source platform for mass spectrometry based glycoproteomics data analysis.* Mol Cell Proteomics, 2017.
174. Mayampurath, A., et al., *Label-free glycopeptide quantification for biomarker discovery in human sera.* J Proteome Res, 2014. **13**(11): p. 4821-32.
175. Rebecchi, K.R., et al., *Label-free quantitation: a new glycoproteomics approach.* J Am Soc Mass Spectrom, 2009. **20**(6): p. 1048-59.
176. Ye, Z., et al., *Glyco-DIA: a method for quantitative O-glycoproteomics with in silico-boosted glycopeptide libraries.* Nat Methods, 2019. **16**(9): p. 902-910.
177. Parker, B.L., et al., *Terminal Galactosylation and Sialylation Switching on Membrane Glycoproteins upon TNF-Alpha-Induced Insulin Resistance in Adipocytes.* Molecular & Cellular Proteomics, 2016. **15**(1): p. 141-153.
178. Budnik, B., et al., *SCoPE-MS: mass spectrometry of single mammalian cells quantifies proteome heterogeneity during cell differentiation.* Genome Biol, 2018. **19**(1): p. 161.
179. Robinson, R.C., N.A. Poulsen, and D. Barile, *Multiplexed bovine milk oligosaccharide analysis with aminoxy tandem mass tags.* PLoS One, 2018. **13**(4): p. e0196513.
180. Zhu, H., et al., *A LC-MS All-in-One Workflow for Site-Specific Location, Identification and Quantification of N-/O- Glycosylation in Human Chorionic Gonadotropin Drug Products.* AAPS J, 2017. **19**(3): p. 846-855.
181. Lee, H.J., et al., *Abundance-ratio-based semiquantitative analysis of site-specific N-linked glycopeptides present in the plasma of hepatocellular carcinoma patients.* J Proteome Res, 2014. **13**(5): p. 2328-38.
182. Savitski, M.M., et al., *Measuring and managing ratio compression for accurate iTRAQ/TMT quantification.* J Proteome Res, 2013. **12**(8): p. 3586-98.
183. McAlister, G.C., et al., *MultiNotch MS3 enables accurate, sensitive, and multiplexed detection of differential expression across cancer cell line proteomes.* Anal Chem, 2014. **86**(14): p. 7150-8.
184. Hogrebe, A., et al., *Benchmarking common quantification strategies for large-scale phosphoproteomics.* Nat Commun, 2018. **9**(1): p. 1045.
185. Keeley, T.S., S. Yang, and E. Lau, *The Diverse Contributions of Fucose Linkages in Cancer.* Cancers (Basel), 2019. **11**(9).
186. Herrera, H., et al., *Core-Fucosylated Tetra-Antennary N-Glycan Containing A Single N-Acetyllactosamine Branch Is Associated with Poor Survival Outcome in Breast Cancer.* Int J Mol Sci, 2019. **20**(10).
187. Becker, D.J. and J.B. Lowe, *Fucose: biosynthesis and biological function in mammals.* Glycobiology, 2003. **13**(7): p. 41R-53R.

188. Neelamegham, S., et al., *Updates to the Symbol Nomenclature for Glycans guidelines.* Glycobiology, 2019. **29**(9): p. 620-624.
189. Tu, Z., Y.N. Lin, and C.H. Lin, *Development of fucosyltransferase and fucosidase inhibitors.* Chem Soc Rev, 2013. **42**(10): p. 4459-75.
190. Belcher, J.D., et al., *The fucosylation inhibitor, 2-fluorofucose, inhibits vaso-occlusion, leukocyte-endothelium interactions and NF-kB activation in transgenic sickle mice.* PLoS One, 2015. **10**(2): p. e0117772.
191. Rillahan, C.D., et al., *Global metabolic inhibitors of sialyl- and fucosyltransferases remodel the glycome.* Nature Chemical Biology, 2012. **8**(7): p. 661-668.
192. Zhou, Y., et al., *Inhibition of fucosylation by 2-fluorofucose suppresses human liver cancer HepG2 cell proliferation and migration as well as tumor formation.* Scientific Reports, 2017. **7**.
193. Liu, Y.C., et al., *Sialylation and fucosylation of epidermal growth factor receptor suppress its dimerization and activation in lung cancer cells.* Proceedings of the National Academy of Sciences of the United States of America, 2011. **108**(28): p. 11332-11337.
194. Wang, Y.Q., et al., *Loss of alpha 1,6-fucosyltransferase inhibits chemical-induced hepatocellular carcinoma and tumorigenesis by down-regulating several cell signaling pathways.* Faseb Journal, 2015. **29**(8): p. 3217-3227.
195. Olsen, J.V., et al., *Global, in vivo, and site-specific phosphorylation dynamics in signaling networks.* Cell, 2006. **127**(3): p. 635-48.
196. Ke, M., et al., *Identification, Quantification, and Site Localization of Protein Posttranslational Modifications via Mass Spectrometry-Based Proteomics.* Adv Exp Med Biol, 2016. **919**: p. 345-382.
197. Mann, M. and O.N. Jensen, *Proteomic analysis of post-translational modifications.* Nat Biotechnol, 2003. **21**(3): p. 255-61.
198. Boersema, P.J., S. Mohammed, and A.J. Heck, *Phosphopeptide fragmentation and analysis by mass spectrometry.* J Mass Spectrom, 2009. **44**(6): p. 861-78.
199. Villen, J., S.A. Beausoleil, and S.P. Gygi, *Evaluation of the utility of neutral-loss-dependent MS3 strategies in large-scale phosphorylation analysis.* Proteomics, 2008. **8**(21): p. 4444-52.
200. Jedrychowski, M.P., et al., *Evaluation of HCD- and CID-type Fragmentation Within Their Respective Detection Platforms For Murine Phosphoproteomics.* Molecular & Cellular Proteomics, 2011. **10**(12).
201. Akimov, V., et al., *UbiSite approach for comprehensive mapping of lysine and N-terminal ubiquitination sites.* Nat Struct Mol Biol, 2018.
202. Satpathy, S., et al., *Systems-wide analysis of BCR signalosomes and downstream phosphorylation and ubiquitylation.* Mol Syst Biol, 2015. **11**(6): p. 810.
203. Wagner, S.A., et al., *Proteomic analyses reveal divergent ubiquitylation site patterns in murine tissues.* Mol Cell Proteomics, 2012. **11**(12): p. 1578-85.
204. Wagner, S.A., et al., *A proteome-wide, quantitative survey of in vivo ubiquitylation sites reveals widespread regulatory roles.* Mol Cell Proteomics, 2011. **10**(10): p. M111 013284.
205. Kim, W., et al., *Systematic and Quantitative Assessment of the Ubiquitin-Modified Proteome.* Molecular Cell, 2011. **44**(2): p. 325-340.
206. Nielsen, M.L., et al., *Iodoacetamide-induced artifact mimics ubiquitination in mass spectrometry.* Nat Methods, 2008. **5**(6): p. 459-60.
207. Porras-Yakushi, T.R., M.J. Sweredoski, and S. Hess, *ETD Outperforms CID and HCD in the Analysis of the Ubiquitylated Proteome.* J Am Soc Mass Spectrom, 2015. **26**(9): p. 1580-7.

208. Tyanova, S., T. Temu, and J. Cox, *The MaxQuant computational platform for mass spectrometry-based shotgun proteomics.* Nat Protoc, 2016. **11**(12): p. 2301-2319.
209. Cox, J. and M. Mann, *MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification.* Nat Biotechnol, 2008. **26**(12): p. 1367-72.
210. Szklarczyk, D., et al., *STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets.* Nucleic Acids Res, 2019. **47**(D1): p. D607-D613.
211. Tyanova, S., et al., *The Perseus computational platform for comprehensive analysis of (prote)omics data.* Nat Methods, 2016. **13**(9): p. 731-40.
212. Cox, J. and M. Mann, *1D and 2D annotation enrichment: a statistical method integrating quantitative proteomics with complementary high-throughput data.* BMC Bioinformatics, 2012. **13 Suppl 16**: p. S12.
213. Wisniewski, J.R., et al., *Universal sample preparation method for proteome analysis.* Nat Methods, 2009. **6**(5): p. 359-62.
214. Batth, T.S., et al., *Protein aggregation capture on microparticles enables multi-purpose proteomics sample preparation.* Mol Cell Proteomics, 2019.
215. Hughes, C.S., et al., *Single-pot, solid-phase-enhanced sample preparation for proteomics experiments.* Nat Protoc, 2018.
216. Holman, J.D., D.L. Tabb, and P. Mallick, *Employing ProteoWizard to Convert Raw Mass Spectrometry Data.* Curr Protoc Bioinformatics, 2014. **46**: p. 13 24 1-9.
217. Yuan, Z.F., et al., *pParse: a method for accurate determination of monoisotopic peaks in high-resolution mass spectra.* Proteomics, 2012. **12**(2): p. 226-35.
218. Kovalchik, K.A., et al., *RawTools: Rapid and Dynamic Interrogation of Orbitrap Data Files for Mass Spectrometer System Management.* J Proteome Res, 2019. **18**(2): p. 700-708.
219. Batth, T.S., C. Francavilla, and J.V. Olsen, *Off-line high-pH reversed-phase fractionation for in-depth phosphoproteomics.* J Proteome Res, 2014. **13**(12): p. 6176-86.
220. Yang, F., et al., *High pH reversed-phase chromatography with fraction concatenation as an alternative to strong-cation exchange chromatography for two-dimensional proteomic analysis.* Expert Review of Proteomics, 2012. **9**(2): p. 129-134.
221. Noble, W.S., *Mass spectrometrists should search only for peptides they care about.* Nature Methods, 2015. **12**(7): p. 605-608.
222. Bantscheff, M., et al., *Quantitative mass spectrometry in proteomics: a critical review.* Analytical and Bioanalytical Chemistry, 2007. **389**(4): p. 1017-1031.
223. Deeb, S.J., et al., *Super-SILAC allows classification of diffuse large B-cell lymphoma subtypes by their protein expression profiles.* Mol Cell Proteomics, 2012. **11**(5): p. 77-89.
224. Ji, Y., et al., *Integrated proteomic and N-glycoproteomic analyses of doxorubicin sensitive and resistant ovarian cancer cells reveal glycoprotein alteration in protein abundance and glycosylation.* Oncotarget, 2017. **8**(8): p. 13413-13427.
225. Zoller, M., *CD44: can a cancer-initiating cell profit from an abundantly expressed molecule?* Nature Reviews Cancer, 2011. **11**(4): p. 254-267.
226. Shaffer, A.L., et al., *BCL-6 represses genes that function in lymphocyte differentiation, inflammation, and cell cycle control.* Immunity, 2000. **13**(2): p. 199-212.
227. Yu, Y.Q., M. Gilar, and J.C. Gebler, *A complete peptide mapping of membrane proteins: a novel surfactant aiding the enzymatic digestion of bacteriorhodopsin.* Rapid Communications in Mass Spectrometry, 2004. **18**(6): p. 711-715.

228. Seddon, A.M., P. Curnow, and P.J. Booth, *Membrane proteins, lipids and detergents: not just a soap opera.* Biochim Biophys Acta, 2004. **1666**(1-2): p. 105-17.
229. Kachuk, C., et al., *Automated SDS Depletion for Mass Spectrometry of Intact Membrane Proteins though Transmembrane Electrophoresis.* J Proteome Res, 2016. **15**(8): p. 2634-42.
230. Xia, S., et al., *Integrated SDS removal and protein digestion by hollow fiber membrane based device for SDS-assisted proteome analysis.* Talanta, 2015. **141**: p. 235-8.
231. Hu, H., K. Khatri, and J. Zaia, *Algorithms and design strategies towards automated glycoproteomics analysis.* Mass Spectrom Rev, 2017. **36**(4): p. 475-498.
232. Nilsson, J., et al., *Enrichment of glycopeptides for glycan structure and attachment site identification.* Nat Methods, 2009. **6**(11): p. 809-11.
233. Kalli, A. and S. Hess, *Effect of mass spectrometric parameters on peptide and protein identification rates for shotgun proteomic experiments on an LTQ-orbitrap mass analyzer.* Proteomics, 2012. **12**(1): p. 21-31.
234. Sigrist, C.J.A., et al., *New and continuing developments at PROSITE.* Nucleic Acids Research, 2013. **41**(D1): p. E344-E347.
235. Li, L., et al., *B-cell receptor-mediated NFATc1 activation induces IL-10/STAT3/PD-L1 signaling in diffuse large B-cell lymphoma.* Blood, 2018. **132**(17): p. 1805-1817.
236. Walls, A.C., et al., *Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein.* Cell, 2020. **181**(2): p. 281-292 e6.
237. Wright, G.W., et al., *A Probabilistic Classification Tool for Genetic Subtypes of Diffuse Large B Cell Lymphoma with Therapeutic Implications.* Cancer Cell, 2020. **37**(4): p. 551-568 e14.
238. Bull, C., et al., *Targeting Aberrant Sialylation in Cancer Cells Using a Fluorinated Sialic Acid Analog Impairs Adhesion, Migration, and In Vivo Tumor Growth.* Molecular Cancer Therapeutics, 2013. **12**(10): p. 1935-1946.
239. Michel, L., et al., *Activated leukocyte cell adhesion molecule regulates B lymphocyte migration across central nervous system barriers.* Sci Transl Med, 2019. **11**(518).
240. Bod, L., et al., *IL-4-Induced Gene 1: A Negative Immune Checkpoint Controlling B Cell Differentiation and Activation.* J Immunol, 2018. **200**(3): p. 1027-1038.
241. Simonetti, G., et al., *SIGLEC-G deficiency increases susceptibility to develop B-cell lymphoproliferative disorders.* Haematologica, 2014. **99**(8): p. 1356-64.
242. Ji, Y., et al., *The Sel1L-Hrd1 Endoplasmic Reticulum-Associated Degradation Complex Manages a Key Checkpoint in B Cell Development.* Cell Rep, 2016. **16**(10): p. 2630-2640.
243. Witkowska, M., P. Smolewski, and T. Robak, *Investigational therapies targeting CD37 for the treatment of B-cell lymphoid malignancies.* Expert Opin Investig Drugs, 2018. **27**(2): p. 171-177.
244. Young, R.M. and L.M. Staudt, *Targeting pathological B cell receptor signalling in lymphoid malignancies.* Nature Reviews Drug Discovery, 2013. **12**(3): p. 229-243.
245. Ryan, D.H., et al., *Vascular cell adhesion molecule-1 and the integrin VLA-4 mediate adhesion of human B cell precursors to cultured bone marrow adherent cells.* J Clin Invest, 1991. **88**(3): p. 995-1004.
246. Dang, Q., et al., *LAMP1 Overexpression Predicts for Poor Prognosis in Diffuse Large B-cell Lymphoma.* Clin Lymphoma Myeloma Leuk, 2018. **18**(11): p. 749-754.
247. Mello, A.S., et al., *LAMP2 as a marker of EBV-mediated B lymphocyte transformation in the study of lysosomal storage diseases.* Molecular and Cellular Biochemistry, 2014. **385**(1-2): p. 1-6.
248. Yada, M., et al., *Phosphorylation-dependent degradation of c-Myc is mediated by the F-box protein Fbw7.* Embo Journal, 2004. **23**(10): p. 2116-2125.

9. Appendices

A. All the Supplementary data are stored on a CD-ROM that can be found attached at the back of the hard copy of this thesis.

Supplementary Data S1: Protein databases used for the comparison in SugarQuant

Supplementary Data S2: Total identified glycoforms, glycosites and glycans in 2FF-treated DG75 cells using SugarQuant

Supplementary Data S3: Single-shot proteome data in 2FF-treated DG75 cells

Supplementary Data S4: Annotation of altered glycoproteins in 2FF-treated DG75 cells

Supplementary Data S5: Quantitative phosphoproteome data in 2FF-treated DG75 cells

Supplementary Data S6: Single-shot proteome data in N-glycoproteomics analysis of 5 ABC and 4 GCB cell lines

Supplementary Data S7: Total identified glycoforms, glycosites and glycans in N-glycoproteomics analysis of 5 ABC and 4 GCB cell lines

Supplementary Data S8: Total quantified glyco-events in deglycoproteomics analysis of 5 ABC and 4 GCB cell lines

Supplementary Data S9: Single-shot proteome data in N-glycoproteomics analysis of TMD8 after 2FF treatment and FUT8 KO

Supplementary Data S10: Total identified glycoforms, glycosites and glycans in N-glycoproteomics analysis of TMD8 after 2FF treatment and FUT8 KO

Supplementary Data S11: Single-shot proteome data in GPome and pYome analysis of DLBCL

Supplementary Data S12: Quantitative phosphoproteome data from GPome and pYome analysis of DLBCL

Supplementary Data S13: Significantly regulated phospho-events in GPome analysis of DLBCL

Supplementary Data S14: Significantly regulated phospho-events in pYome analysis of DLBCL

Supplementary Data S15: Single-shot proteome data in ubiquitinome analysis of DLBCL

Supplementary Data S16: Quantitative ubiquitinome data in DLBCL

Supplementary Data S17: Significantly regulated ubi-events in ubiquitinome analysis of DLBCL

B. Supplementary Table S1: Detailed parameter settings for the Glyco-SPS-MS3 and optimized MS2 methods for TMT6-labelled N-glycopeptide analysis.

		MS2	**Glyco-SPS-MS3**
MS1	MS instrument	**Orbitrap Fusion or Lumos**	
	Detector Type	**Orbitrap**	
	Orbitrap Resolution	**120 k**	
	Mass Range (m/z)	**350-2000**	
	Maximum injection time (ms)	**50**	
	AGC target	**$5e^5$**	
	RF Lens	**60%**	
	DataType	**Profile**	
	Precursor selection range (m/z)	**700-2000**	
MS2	Isolation mode	Quadrupole	**Quadrupole**
	Isolation window (m/z)	2	**2**
	Scan range mode	Auto normal	**Auto normal**
	First mass (m/z)	120	**132**
	Activation type	HCD	**HCD**
	Normalized collision energy (%)	25/35/45	**25**
	Detector type	Orbitrap	**Orbitrap**
	Orbitrap resolution	15 K	**15 K**
	Maximum injection time (ms)	500	**150**
	AGC target	$5e^5$	**$5e^5$**
	Data type	Profile	**Profile**
	Precursor selection range (m/z)		**700-2000**
MS3	Number of Notches		**10**
	Isolation mode		**Quadrupole**
	Isolation window (m/z)		**2**
	MS2 isolation window (m/z)		**2**
	First mass (m/z)		**120**
	Scan range mode		**Auto normal**
	Activation type		**HCD**
	Collision energy (%)		**35**
	Detector type		**Orbitrap**
	Orbitrap resolution		**15 K**
	Maximum injection time (ms)		**350**
	AGC target		**$5e^5$**

Note: For N-glycopeptides quantitative analysis with TMT10plex or 11plex labeling, all the parameters were the same except for an adjustment of orbitrap resolution in the MS3 step from 15 K to 60 K.

C. List of abbreviations

2FF	2-fluoro-L-fucose
ABC	Ammonium bicarbonate
ABC DLBCL	Activated B-cell-like DLBCL
ACN	Acetonitrile
AGC	Automated gain control
AID	Activation-induced deaminase
ALL	Acute lymphocytic leukemia
AML	Acute myeloid leukemia
AP	Acetone-precipitation
Asn	Asparagine
ATP	Adenosine triphosphate
BCR	B-cell receptor
BL	Burkitt's lymphoma
BRP	Basic reversed-phase high performance liquid chromatography
BSA	Bovine serum albumin
CAA	Chloroacetamide
CID	Collision induced dissociation
CLL	Chronic lymphocytic leukemia
CML	Chronic myeloid leukemia
CSR	Class-switch recombination
CSS	CRISPR screen score
Ctr	Control
CV	Coefficient of variation
Da	Dalton
DAG	Diacylglycerol
DDA	Data-dependent acquisition
DIA	Data-indendent acquisition
DLBCL	Diffuse large B-cell lymphoma
DMSO	Dimethylsulfoxide
DPA	Differential phosphopeptides analysis
DP-events	Differential phospho-events
DTT	Dithiothreitol
DUB	Differential ubiquitylated peptide analysis
DUBs	Deubiquitinating enzymes
eBL	Endemic BL
EBV	Epstein-Barr virus
ECD	Electron capture dissociation
ERLIC	Electrostatic repulsion hydrophilic interaction chromatography
ETD	Electron transfer dissociation
EThcD	Electron-transfer/higher-energy collision dissociation
FA	Formic acid
FASP	Filter-aided sample preparation
FCS	Fetal calf serum
FDA	US Food and Drug Administration
FDR	False discovery rate

FT-ICR	Fourier-transform ion cyclotron
Fuc	Fucose
FUK	Fucokinase
Gal	Galactose
GCB DLBCL	Germinal center B-cell-like DLBCL
GO	Gene Ontology
GP	Global phosphoproteome
GPSM	N-glycopeptide-spectrum-match
GTP	Guanosine triphosphate
H_3PO_4	Meta-phosphoric acid
HCD	Higher energy collisional dissociation
HEPES	4-(2-Hydroxyethyl)piperazine-1-ethanesulfonic acid
HIV	Human immunodeficiency virus
HL	Hodgkin's lymphoma
HPO_3	Phosphoric acid
HRS	Hodgkin-Reed-Sternberg
HSCs	Hematopoietic stem cells
IAA	Iodoacetamide
IBR	Ibrutinib
ICAT	Isotope-coded affinity tags
IG	Immunoglobulin
IgH	Immunoglobulin heavy
IgL	Immunoglobulin light
IMAC	Immobilized metal ion affinity chromatography
IP	Immunoprecipitation
IP3	Inositol triphosphate
IR	Infrared photoactivation
IRM	Ion-routing multipole
IT	Injection time
iTRAQ	Isobaric tags for relative and absolute quantitation
KO	Knock-out
LC	Liquid chromatography
M	Molar
MALDI	Matrix-assisted laser desorption ionization
MOAC	Metal oxide affinity chromatography
mRNA	Messenger RNA
MS/MS	Tandem mass spectrometry
MSA	Multistage activation
Na_3VO_4	Sodium orthovanadate
NaCl	Sodium chloride
NaF	Sodium fluoride
NaOH	Sodium hydroxide
NCE	Normalized collision energy
NeuAC	N-acetylneuraminic acid
NeuGC	N-glycolylneuraminic acid
NF	Normalization factor
NH_4Ac	Ammonium acetate

NH_4OH	Ammonium hydroxide
NHL	Non-Hodgkin lymphomas
NP-40	Nonidet P-40
PAC	Protein aggregation capture
PBS	Phosphate buffered saline
PCA	Principal component analysis
ppm	Parts per million
PRT	PRT-060318/-062617
pS	Phosphorylated serine
pT	Phosphorylated threonine
PTM	Post-translational modification
pY	Phosphorylated tyrosine
R	Arginine
RNA	Ribonucleic acid
rpm	Rounds per minute
RT	Room temperature
SA	Sialic acid
SAX	Strong anion exchange
sBL	Sporadic BL
SCX	Strong cation exchange
SDS	Sodium dodecyl sulphate
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
Ser	Serine
sgRNA	Small guide ribonucleic acid
SH2	SRC homology 2
SILAC	Stable isotope labeling with amino acids in cell culture
SLL	Small lymphocytic lymphoma
sNCE	Stepped collision energy
SP3	Single-pot solid-phase enhanced sample preparation
SPS	Synchronous precursor selection
ST6Gal1	β-galactoside α-2, 6-sialyltransferase
TEAB	Triethylammonium bicarbonate buffer
TFA	Trifluoroacetic acid
Thr	Threonine
TiO_2	Titanium dioxide
TLR	Toll-like receptor
TMT	Tandem mass tags
TOF	Time-of-flight
Tris	Trisamine
U	Unit
XICs	Extracted ion currents