Dissertation
submitted to the
Combined Faculties for the Natural Sciences and for Mathematics
of the Ruperto-Carola University of Heidelberg, Germany
for the degree of
Doctor of Natural Sciences

Presented by: M.Sc. Fei Ye
Born in: Ningbo, Zhejiang, P.R.China
Oral-examination: September 18th, 2017

Cell-type Dependent Changes of Protein Post-translational Modifications in The Ageing Human Bone Marrow Proteome

Referees:
Dr. Christoph Müller
Prof. Dr. Joachim Wittbrodt

Acknowledgment

I would like to thank my supervisor Dr. Anne-Claude Gavin, for giving me the opportunity to participate in the SyStemAge project, for the scientific freedom she gave me and for her comments on this thesis. It has been a memorable experience, through which I have learned to believe in myself and to dream about the impossible.

I am grateful to my thesis advisory committee members Dr. Christoph Müller, Prof. Dr. Joachim Wittbrodt and Prof. Dr. Julio Saez-Rodriguez for their support throughout my PhD. I would also like to take this opportunity to thank all my previous supervisors, for nurturing my scientific curiosity and inspiring me to do research.

During my PhD, Marco Hennrich taught me mass spectrometry from the basic. I benefited from discussion with Denes Turei on data analysis. Suggestions from current and past members of the Gavin group were always helpful. The comradeships with Noorie Karimbocus and Ximing Ding were indispensable.

For this thesis, I would like to thank Cihan Erkut and Charlotte Gehin for their suggestions, Ximing Ding for taking his time to translate the abstract, and Henryk Edlund, *le renard du Petit Prince*, for helping me proofread.

All people I befriended with at EMBL coloured the days of my PhD. I would like to thank these wonderful people, including Bingqing Zhao, Chenchen Zhu, Olga Mikhaylichenko, Vladislava Milchevskaya, Serge Dmitrieff, Benjamin Lang, Nan Li, Kun Qu and Zhengge Zhong, and many others who are not mentioned here.

I am very grateful to my mentor Dr. Gregor Obernosterer for always encouraging me to find meanings and pursue dreams in life, and captain Dr. Christian Frese for giving me simple and effective suggestions during my crisis.

Especially, I would like to express my deepest gratitude to my friend Lin Gen, for his unreserved help and support throughout the last year of my PhD, for the knowledge he shared with me, for the priceless lessons I learned from him, and for going through all sorts of difficulties with me. Without him, I would not make it this far and this thesis certainly would not be the same.

Last but most importantly, I would like to thank my parents for the undoubted support and trust they gave me during the past years.

I am in debt to all of you.

Contents

1 Summary

Ageing is a highly complexed and regulated interplay between genetic and environmental factors in living organisms. At cellular level, ageing is accompanied with numerous alternations including changes on protein post-translational modifications (PTMs). Using quantitative proteomics approach, my phD project aimed at extending our current understanding on ageing in the human bone marrow, with a focus on the role of PTMs. The study stratified human bone marrow into six main sub-populations of cells and examined changes of the proteome for 59 donors whose age differences span 40 years. Modified proteins were identified through a combination of mass-tolerant and standard database search method. The analysis showed that PTMs are prevalent in the bone marrow proteome and displayed cell type dependencies. Proteins differentially modified include glycolytic proteins and proteins from the heterogeneous nuclear ribonucleoprotein family. Furthermore, analysis on the association between modified proteins and age showed that the level of modification for the majority of the modified proteins remained stable during ageing. Acetylated proteins in the haematopoietic progenitor cells were found to bias towards negative correlation with age, which may be linked with reduced differentiation potential of the stem cells. Co-regulations of modified proteins during ageing were also observed in the bone marrow cells, especially in the haematopoietic progenitor cells. The results from this study could serve as the basis for elucidating the biological mechanisms by which these PTMs regulate age-associated processes in different bone marrow derived cells.

2 Zusammenfassung

Alterung ist ein hoch komplexes und streng reguliertes Zusammenspiel zwischen genetischen und umweltlichen Faktoren in lebenden Organismen. Auf zellulärer Ebene ist der Alterungsprozess begleitet von zahlreichen Modifizierungen, einschliesslich Veränderungen der posttranslationalen Proteinmodifikationen (PTM). Mit einem quantitativen proteomischen Ansatz ist das Ziel dieser Arbeit unser jetziges Verständnis von Alterung in der Knochenmarknische zu erweitern, mit einem Fokus auf die Rolle der PTM. In der folgenden Arbeit wird das menschliche Knochenmark in sechs Zellsubpopulationen unterteilt. Veränderungen des Proteoms wurden für 59 Spender untersucht, deren Altersunterschied 40 Jahre beträgt. Modifizierte Proteine wurden identifiziert durch Kombinierung von Masse-tolerant und Standard Datenbank Suchmethoden. Die Analyse zeigte, dass PTM im Knochenmarkproteom weitverbreitet und abhängig vom Zelltyp sind. Zu den unterschiedlichen modifizierten Proteinen gehören Proteine der Glykolyse und Proteine der heterogenen Ribonukleinfamilie. Zusätzlich zeigte die altersabhängige Assoziierungsanalyse der modifizierten Proteine, dass das Modifizierungsniveau für die Mehrheit der modifizierten Proteine während des Alterungsprozesses stabil blieben. Die acetylierten Proteine in den hämatopoetischen Progenitoren korrelierten negativ mit dem Alter, was vielleicht auf das reduzierte Differenzierungspotential der Stammzellen zurückzuführen war. Koregulierungen der modifizierten Proteine während des Alterungsprozesses wurden ebenfalls in den Knochenmarkzellen, vor allem aber in den hämatopoetischen Progenitoren beobachtet. Die Ergebnisse dieser Arbeit könnten als Basis dienen für die Aufklärung biologischer Mechanismen, in denen die beschriebenen PTM die altersassoziierten Prozesse in unterschiedlichen Knochenmarkzellen regulieren.

3 Abbreviation

ERP	erythroid progenitor cell
FACS	fluorescent activated cell sorting
FDR	false discovery rate
GRA	granulocyte progenitor cell
HPLC	high-performance liquid chromatography
HPC	haematopoietic progenitor cell
hnRNP	heteronuclear binding protein
HSC	haematopoietic stem cell
LYM	lymphoid progenitor cell
LC-MS	liquid chromatography-mass spectrometry
MON	monocyte progenitor cell
MS	mass spectrometry
MS/MS	tandem mass spectrometry
MSC	mesenchymal stromal cell
PCA	principal component analysis
p.p.m	parts per million
PSM	peptide spectrum matching
PTM	post-translational modification
RNA	ribonucleic acid
RRM	RNA recognition motif
RP-HPLC	reverse phase high-performance liquid chromatography
TMT	tandem mass tagging

4 List of Figures and Tables

List of Figures

List of Tables

List of Supplementary Figures

5 Introduction

5.1 Protein post-translational modifications are crucial for the regulation of cellular activities

Proteins are the basis of cellular function of a cell [Alberts, 2008,Lehninger et al., 2005]. It is an ensemble of polypeptides synthesized inside cells from amino acids through covalent linkage of the carboxyl group of one amino acid to amino group of another amino acid. The primary structure of a protein refers to chains of amino acids in a specific sequence. The order of the amino acids in the sequence is determined by the genetic information encoded in the genome. The backbone chain of a polypeptide is arranged based on the distance of hydrogen-bond between the main-chain peptide groups. This conformational arrangement gives rise to the secondary structure of a protein, such as α-helixes and β-sheets. Packing of these secondary structures via non-specific hydrophobic interaction and specific interaction, such as salt bridges and disulfide bonds, gives rise to a three-dimensional conformation. The overall three-dimensional conformation is the tertiary structure of a protein, and is essential for its biological function. The conserved part of the tertiary structure of a protein is also known as a protein domain [Apic et al., 2001]. Protein domains are present in nearly all proteins (except disordered proteins, reviewed in [Dyson and Wright, 2005, Wright and Dyson, 2015]), and can evolve differently from the rest of the amino acid sequences [Thornton et al., 1999]. Duplication and combination of protein domains create a plethora of protein structures, which ultimately determine the

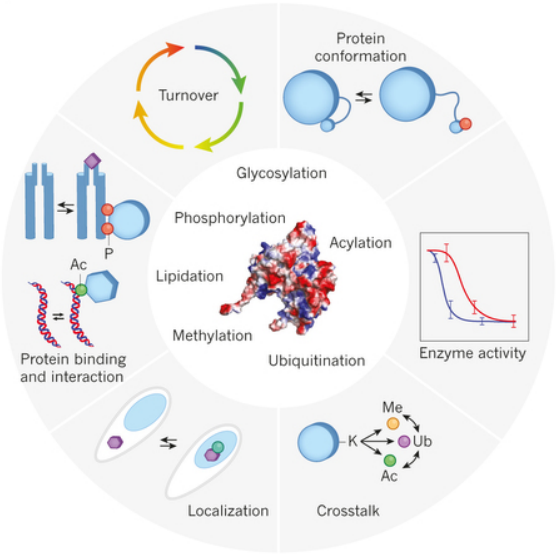


Figure 1: Protein post-translational modifications and their functions inside cells. Cells regulate various functions through protein modifications. Proteins can be modified through attachment of various chemical moieties by controlled process using specific sets of enzymes. PTMs often have multiple functions inside the cell. Phosphorylation can change conformation of proteins through allosteric modulation or change protein activity; lipidation and glycosylation can direct sub-cellular location of proteins; methylation and acylation can affect protein-protein interactions whereas ubiquitination can control the lifetime of proteins. Different PTMs can crosstalk with each other on the same protein through competing for the same amino-acid residue, or residues in close proximity. Some commonly studied modifications and their functions are illustrated here. Abbreviations: Ac, acetyl; Me, methyl; P, phosphate; K, lysine; Ub, ubiquitin. Adapted from [Aebersold and Mann, 2016].

function of proteins. In addition, many proteins function as assemblies of multiple polypeptide chains, which are often referred to as protein subunits. The number and arrangement of subunits of a multi-unit protein complex is also known as the quaternary structure of the protein.

Proteins actively participate in all cellular activities, from regulating gene transcription, protein translation, to responding to internal and external stimuli. They are multifunctional molecules: as enzymes they catalyse chemical reactions inside the cellular environment; as structural proteins they provide mechanical support and enable movement for cells; as transporters they carry molecules required by cells to locations where they are needed. Because of their indispensable roles inside cells, several layers of mechanisms have evolved to regulate proteins [Alberts, 2008]: regulation of protein synthesis and degradation rates can control the abundance level of proteins; allosteric transitions such as releasing the catalytic subunit of a protein or cooperative binding of ligands can induce structural changes of proteins; segregation of proteins in different cellular compartments is used to create physical barriers for protein-protein interactions; post-translational modifications (PTMs) alter the chemical properties of proteins and can modulate activities of proteins.

Protein PTM is regulatory mechanism in addition to gene expression and translational control [Lehninger et al., 2005, Walsh, 2006]. Inside a cell, PTMs are continuously regulated via endogenous enzymes, or can be acquired from internal and external environment. Essentially, protein PTMs are chemical modifications to the side-chains of amino acids of proteins, which can be a reversible or irreversible process. It can occur at any stage during the lifespan of a protein, to both fine-tune the level of protein expression and modulate protein function (Figure 1): direct modification such as acetylation and glycosylation during and after completion of translation affects folding and stability of proteins; phosphorylation can induce protein conformational change and control protein activity; ubiquitination, the attachment of a 76-amino-acid polypeptide ubiquitin to a peptide, can target proteins for proteasomal degradation; lipidation and glycosylation can direct proteins to specific cellular compartments. There are about 20,000 annotated protein-coding genes in the human genome, but protein PTMs greatly extend the diversity and complexity of the resulting human proteome.

PTMs do not act in isolation. The crosstalk between PTMs further increases the complexity by which protein functions can be regulated. There is an increasing amount of evidence of extensive crosstalks between different protein PTMs. Broadly, as suggested by Hunter, crosstalk between protein PTMs can be classified into positive and negative interactions [Hunter, 2007]. In the positive form, the initial modification serves as an activating point for additional PTM or the removal of a second PTM. This is best illustrated by the phosphorylation-dependent ubiquitination as well as phosphorylation-dependent

SUMOylation [Koepp et al., 2001,Nguyen et al., 2013,Hietakangas et al., 2006]. In contrast, in the negative crosstalk two modifications have a direct competition for the same amino acid or indirectly inhibit each other through competing in three-dimensional space.

5.2 Mass spectrometry-based proteomics as a mean for studying protein post-translational modifications

Classical approaches for detecting protein PTM can be traced back to the early structural and functional studies on individual proteins [Walsh, 2006]. Methods such as Edman degradation and Bergmann degradation were developed in the early 20th century. Based on the chemical properties of peptide bonds, those methods can precisely define amino acid sequence and the position of the amino acid being modified in a peptide sequence. However, these methods require use of radioactive substances, which is inconvenient and hazardous. Subsequently, methods such as protein immunno-blotting or enzyme-linked immunosorbent assay (ELISA) were developed in the mid-20th century and provided alternatives for determining protein PTMs. These methods rely on the binding interactions between antigen and antibody to identify modified proteins. Accuracy and reproducibility of these antibody-based assays are highly dependent on the specificity of the antibody. It still remains as a challenging task nowadays to obtain antibodies with high specificity for PTMs with small chemical moieties and minimal dependence on flanking amino acid sequences. Peptide and protein arrays were developed in more recent years. Based on the specificity of PTM-specific enzymes, the methods have been used to profile binding affinity and for identification of modification residues [Zhu et al., 2000, Rathert et al., 2008].

The aforementioned biochemical approaches are restricted to studying the structure and function of a small group of proteins. In contrast, mass spectrometry (MS)-based proteomics methods aim to provide a global understanding of the entire collection of proteins in a biological sample. Such collection of proteins in a system, whether it is an organism, a tissue or an organelle, is known as a proteome. Information extracted from proteomics data is able to provide key insights regarding the dynamics of protein functions of the chosen biological system. Proteomics-based methods are also suitable for studying protein PTMs, because any modifications on a protein will lead to a significant mass shift of the molecule. The change of mass is reflected in the mass spectrometer analysis as the modified peptide has a different mass-over-charge (m/z) ratio than that of the unmodified peptide. The type of modification can be identified and the site of modification can be precisely located to a single amino acid through analyzing the fragmentation patterns of modified peptides [Aebersold and Mann, 2016].

Furthermore, proteomics-based methods are able to address some of the bottlenecks encountered in the biochemical approaches [Zhao and Jensen, 2009]. There are several advantages of using proteomics-based methods to study protein PTMs. Apart from global

identification of substrate proteins and the possibility to study the dynamics of protein PTMs in living systems, MS-based proteomics methods are also highly sensitive due to the advanced MS instruments. The use of high-performance liquid chromatography (HPLC) systems and high-resolution mass spectrometers has enabled measurement up to attomolar level concentration [Forsgard et al., 2010]. Labelling techniques as well as the label-free quantification methods developed in the last decade further enable us to perform accurate quantification on abundance and changes of protein PTMs. It is entirely feasible to perform quantitative analysis on the stoichiometry of a protein modification, i.e. the proportion of modified proteins versus the unmodified proteins in a biological system. All of these advantages, coupled with the use of automated robotic platforms and software for data analysis facilitate the study of PTMs in an unbiased and high throughput manner.

MS-based proteomics methods have been applied by researchers to study various protein PTMs, such as phosphorylation (phosphoproteome), acetylation (acetylome), methylation (methylome), ubiquitination (diglycine remnant proteomics) and oxidation (redox proteomics). System-wide PTM studies have consistently identified large numbers of amino acid sites that are modified. Phosphorylation is the most studied protein PTM using proteomics methods. Recent large scale and in-depth analysis on the human phosphoproteome has identified from 20,443 to 38,000 phosphorylation sites in human cells [Choudhary and Mann, 2010,Sharma et al., 2014]. Other modifications such as methylation and acetylation, in contrast, are less studied and characterized than phosphorylation due to the difficulties in generating specific antibodies that recognize various modification sites at proteome level. For acetylation, the first proteome-wide analysis in human cells on lysine acetylation revealed 3,600 acetylation sites in Jurkat cells [Choudhary et al., 2009]. Proteomics analysis on substrates of deacetylase SIRT3 also identified 3,174 acetylation sites in U2O2 cells [Sol et al., 2012]. Recently, the first high-resolution mass spectrometry analysis on arginine methylation revealed 8,030 arginine methylation sites within 3,300 human proteins in human embryonic kidney 293 cells [Larsen et al., 2016]. This high number suggests that the occurrence of protein methylation may be comparable to that of phosphorylation.

5.3 Protein identification and quantification in mass spectrometry-based proteomics

Proteins can be studied either as intact molecules or as a collection of peptide forms using mass spectrometer. The former one, known as top-down proteomics, though technically and computationally challenging, has the advantage that, in principle, all modifications on a given protein, can be examined all at the same time [Tran et al., 2011]. The latter one, known as the bottom-up proteomics, is more widely used by researchers for it is experimentally more tractable and computationally more established. To date, there are three main approaches used in bottom-up proteomics (Figure 2A): the data-dependent acquisition (DDA) method aims at unbiasedly profiling the complete proteome; the data-independent

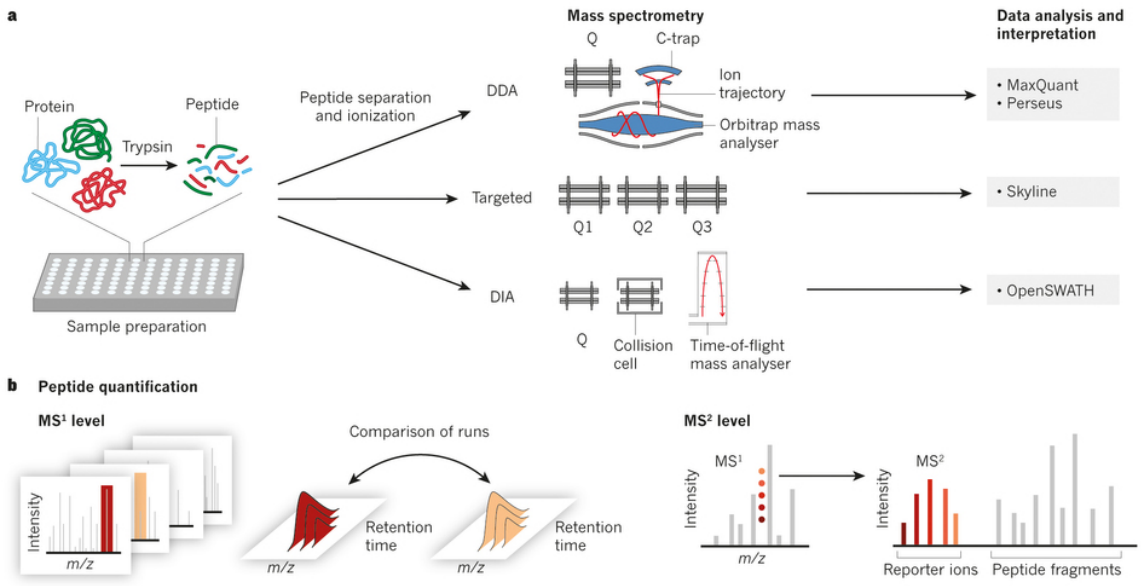


Figure 2: Protein identification and quantification in bottom-up proteomics. A) bottom-up proteomics workflow can be divided into three stages, sample preparation, measurement by mass spectrometer and in silico data analysis. During the sample preparation stage, proteins are extracted from source material and digested to peptides by sequence-specific enzymes. Often the digested tryptic peptides are fractionated by chromatography before being injected into the mass spectrometer. Afterwards, fractionated peptides are electro-sprayed into a mass spectrometer. For mass spectrometry analysis, there are three widely used methods: the data-dependent acquisition (DDA) methods acquire a full spectrum of peptides at MS1 level, and subsequently collect as many fragmentation spectra at MS2 level as possible; targeted methods such as selected reaction monitoring (SRM) selectively capture peptides of known mass-to-charge ratio (m/z) at MS1 level, fragment the captured peptides and then monitor several of the fragmented peptides over time; the data-independent acquisition(DIA) method such as the sequential window acquisition of all theoretical fragment-ion spectra (SWATH-MS) is able to select peptides with a range of m/z values. The selected peptides are fragmented and analyzed in a time-of-flight (TOF) mass spectrometer. Data analysis of proteomics data usually requires implementation of software suited for each method. For data obtained from the DDA method, commercial software such as Proteome Discoverer and free software such as MaxQuant as well as its downstream analysis software Perseus are being widely used. For targeted studies software such as SkyLine is often used. Studies using DIA method require interpretation of multiplexed fragment spectra and known fragment spectra from large spectral libraries is often used to help with the spectra identification. Software such as OpenSWATH has been developed. B) peptide quantification can be achieved at both MS1 level and MSn level. For MS1 quantification (left), elution profile of precursor ions from the high-performance liquid chromatography (HPLC) column are integrated, enabling quantitative comparison of isotopic clusters of the same peptide from two different mass spectrometer runs. Absolute quantities can be estimated by adding up peak volumes of all peptides that identify a particular protein and then determining the proportion to the (known) total proteome mass that has been analysed. Peptides can also be subjected to label-free quantification at the MS2 level (right). The intensities of fragment-ion that are unique to a specific peptide are used for quantification, in a way that is analogous to the use of precursor-ion signal intensities for quantification using MS1-level data. For multiplexed shotgun proteomics, the reporter ions required for quantification are released at MS2 or MS3 stage during the analysis. For DIA methods the intensities of fragments that belong to the same precursor ion are extracted to yield a measure of peptide abundance. Adapted from [Aebersold and Mann, 2016].

acquisition(DIA) method aims at producing an in-depth, more comprehensive fragmentation information for a given sample [Chapman et al., 2014]; the targeted proteomics method, which uses selected reaction monitoring (SRM), aims at selectively and reproducibly acquiring peptides of interests [Picotti and Aebersold, 2012].

In general, protein identification in bottom-up proteomics begins with extraction and purification of proteins of interests, cleaving proteins into short oligopeptides by sequence-specific enzymes (Figure 2A) [Zhang et al., 2013]. Subsequently, the enzymatically digested peptides are separated to reduce the complexity of the sample for MS. Different types of fractionation methods have been developed, including electrophoretic technique-based fractionation, differential-centrifugation-based fractionation and multi-dimensional chromatography-based fractionation. Combinations of orthogonal fractionation methods are often preferred as peptides are separated based on different physiochemical properties, hence the complexity of the mixture can be further reduced.

Directly before or after the fractionation, an enrichment step is usually incorporated to select the modified proteins or peptides from the sample mixture. This step is the main difference between a generic bottom-up proteomics study and a PTM study. The enrichment procedures differ according to the nature of each modification. Antibody-based affinity purification is widely used for PTMs with high affinity antibodies available, such as tyrosine phosphorylation, arginine methylation and lysine acetylation [Zhang et al., 2009, Ong et al., 2004, Rikova et al., 2007]. A protein PTM can also be chemically modified in vitro for affinity labelling to enable further affinity purification, as in the case of β-elimination of O-phosphorylated residues and histidine tag or HA tag of N-terminus of ubiquitin [Oda et al., 2001, Peng et al., 2003]. Ionic interaction-based enrichment methods utilise the chemical properties of modification groups. Methods such as immobilized metal affinity chromatography (IMAC) and titanium dioxide-based solid matrix take advantage of the negatively charged phosphate group and its ability to interact with ion exchange beads [Swaney et al., 2009, Pinkse et al., 2004]. Other methods such as strong cation exchange (SCX) explore the charge-state differences between tryptic phosphopeptides and tryptic unphosphorylated peptides at low pH for peptide separation [Beausoleil et al., 2004]. After the enrichment step, samples containing the PTM-enriched peptides are introduced into the vacuum of a mass spectrometer via electro-spray. Information of the mass-over-charge (m/z) of each peptide on the MS level and MSn levels are collected by the mass analyser.

The acquired MS/MS spectra are analyzed using software developed according to the type of mass spectrometry analysis (Figure 2B). The software can deduce the sequence of the peptide and determine which protein in the protein database or the protein sequence from a corresponding nucleic acid sequence gives the best match to the spectrum. Protein

sequences in databases are usually cleaved in silico based on the specificity of the enzyme used in a given experiment. The masses of the digested peptides are calculated from the in silico digestion. If the calculated peptide mass matches with the peptide mass acquired at MS level, the masses of the corresponding fragment ions will be predicted by in silico fragmentation. The masses of the predicted fragment ions are then compared with the masses of the acquired fragment ions. For a given spectrum, database search algorithms also evaluate quality of the matching between the in silico sequences and the experimentally acquired spectrum. The scores or ratings of the matching are used by software to decide which predicted peptide sequence is the most likely one among all.

Over the last few decades mass spectrometry has matured from a technique that produces purely descriptive information on protein to a technique that can generate accurate and quantitative information that can be used to gain insights on protein abundance and dynamics. Several quantification strategies have been established for MS-based proteomics. They can be broadly categorised into two types, the absolute and relative quantification. Peptide quantification methods for relative quantification can be achieved at MS-level or MSn-level. The relative quantification of MS-based proteomics can be further divided into two types, label-free quantification and quantification methods that use isotope labelling (Figure 3A). In shotgun proteomics studies, there are mainly two techniques that are used in isotope labelling, namely the metabolic labelling and the chemical labelling.

In metabolic labelling, stable isotopes are incorporated into proteins during protein synthesis. Cells or organisms used in this method are typically auxotroph for the source of the isotope used for isotopic labelling [Wu et al., 2004]. Alternatively, stable isotopes by amino acids in cell culture (SILAC) offers a different approach, in which an amino acid deficient culture medium is supplemented with stable isotope enriched amino acids [Ong et al., 2002]. After a few cell cycles the heavy isotopes will be integrated into the proteome due to protein turnover. Complete isotopic labelling is often achieved after 5-10 cell cycles. One main advantage of using metabolic labelling is that the isotopes are introduced during protein synthesis, hence making it possible to differentiate proteins that are synthesised at different stages during an experiment.

Chemical labelling, on the other hand, introduces chemical labels after proteins are extracted from the source material and purified. Chemical labelling utilises the reactive groups present on peptides and proteins, and can be achieved at both peptide and protein levels. The thiol group of the cysteine residues and the primary amine groups are often the targeted function groups. Linkers containing deuterium atoms can conjugate with a thiol-specific reactive group to label the free cysteine residues, as exemplified by isotope coded affinity tags (ICAT) [Gygi et al., 1999]. Labelling on primary amine groups is often achieved through derivatization of N-hydroxysuccinimide (NHS) chemistry or re-

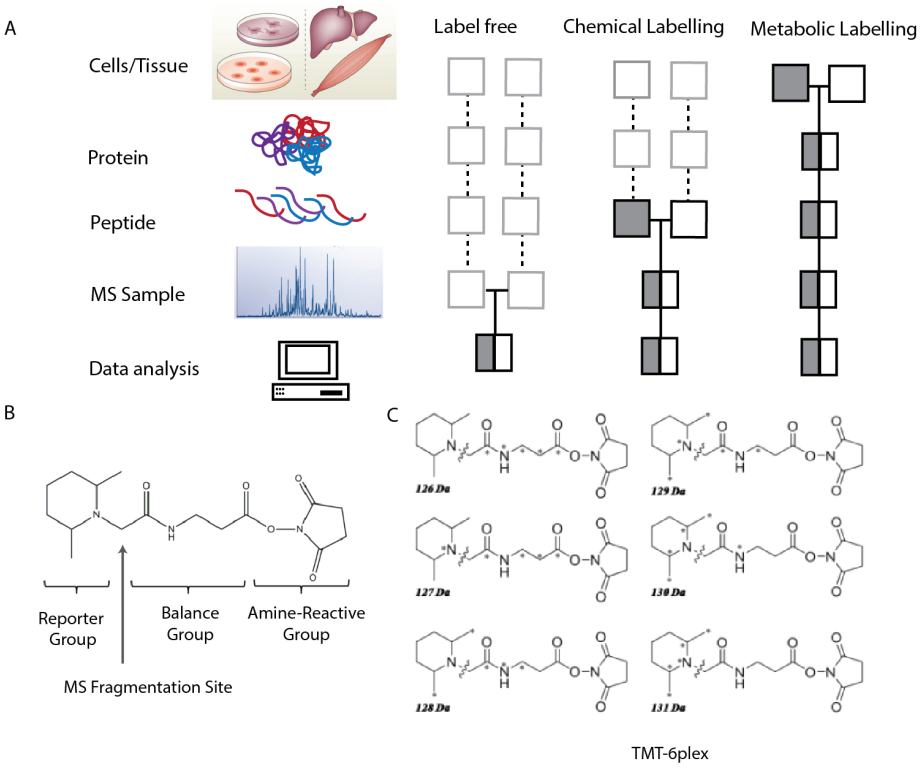


Figure 3: Labelling strategies in quantitative proteomics. A) workflows of quantitative proteomics. Step noted dashed lines shows where the samples are treated separately during experiment whereas steps noted with solid lines shows where the samples are combined. Labels are introduced at different stages depending on the choice of method. In metabolic labelling, samples are labelled and combined at early stage, where as in chemical labelling samples are labelled and combined at a later stage. Label-free quantification does not introduce any chemical tag but requires every sample to be processed separately. B) structure of isobaric mass tags used in chemical labelling. Chemical structure of a generic TMT reagent is shown here: an amine-reactive group that labels N-terminus and amine group of lysine in peptides, a mass normalization (balance) group that neutralises the mass difference contributed by reporter ion to ensure overall mass of tag is the same, and a reporter group that provides information on abundance of peptide upon MS/MS in individual samples. Cleavage site (indicated with arrow) enables release of reporter ion from the whole tag upon MS/MS. C) chemical structure of TMT 6-plex reagents with 13C and 15N heavy isotope positions (asterisks). The tags are isobaric, with a different distribution of isotopes between the reporter and balance groups. Adapted from [Rauniyar and Yates III, 2014].

ductive amination reaction. Di-methyl labelling on amine groups relies on the reaction between amine with formaldehyde followed by reduction by cyanoborohydride [Hsu et al., 2003]. Two cycles of these reactions will convert primary amines into tertiary amines with two methyl groups. Isomers of formaldhyde and cyanoborohydride are used to introduce stable isotopes. Alternatively, the isobaric mass tagging method offers another choice for chemical labelling. The isobaric mass tagging methods also rely on the reaction between the NHS group and the free amine groups on the peptides [Thompson et al., 2003, Ross et al., 2004, Rauniyar and Yates III, 2014]. Different from other labelling methods that aim at creating mass differences between samples, isobaric mass tags introduce chemical labels that have the identical total mass. The generic structure of an isobaric mass tag is illustrated in Figure 3B. Generally, those tags are made up by three parts: a reactive group, a reporter group and a mass balancer group. The total number of isotopes in the mass reporters and mass balancers remains the same, which ensures that the same peptide from differentially labelled samples will have the same profile during full MS scan. The heavy isotopes are distributed between mass tags and mass balancers, therefore the same peptide from differentially labelled samples can be distinguished from its mass tag profiles during MS/MS scan (Figure 3C). One unique advantage of isobaric mass tagging method is that it provides multiplexing capabilities for MS analysis, because the labelling neither affects the LC separation nor interferes with the MS spectra acquisition. The most common isobaric mass tagging methods allow multiplexing from 2 samples and can be scaled up to 10 samples during one MS analysis. In addition, quantification of samples labelled with isobaric mass tags such as TMT or iTRAQs uses information of the cleaved reporter ions from MSn spectra. Compared with MS level quantification, MSn level quantification also gives higher signal-to-noise ratio.

5.4 Mass-tolerant database search provides an alternative way to unbiasedly profile protein post-translational modifications for shotgun proteomics data

Protein identification in shotgun proteomics involves handling large volumes of data collected from MS analysis. The tandem spectra are matched to the corresponding peptides in the database, which leads to the protein identification. As described in the previous section, this starts with analysing the mass of the intact enzymatically cleaved peptides acquired during full MS scan, followed by comparisons of the mass between the predicted fragment peptide ions and the ones acquired experimentally during MSn scans. Peptide sequences are assigned through cross-correlation of in silico predicted spectra with those acquired by MS/MS, using database searching engines, such as SEQUEST and Mascot. However, it has been noted that the majority of the MS/MS spectra acquired in a typical shotgun proteomics experiment remain unidentified [Keller et al., 2002]. The underlying sources for low identification rate in shotgun proteomics is still under debate, but can be partially explained by unspecific enzymatic activity resulting unexpected cleavage sites,

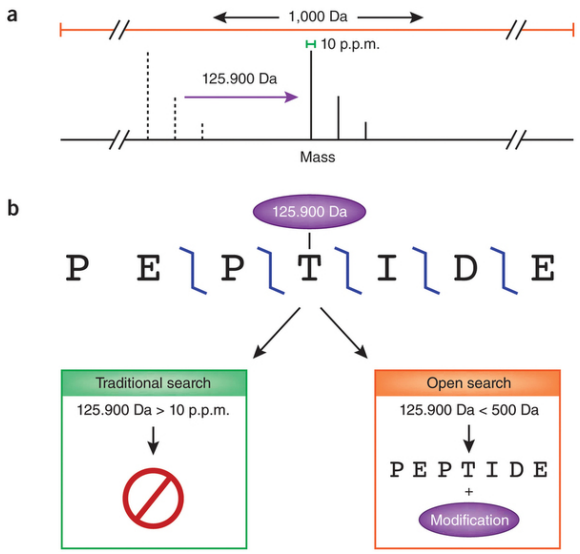


Figure 4: Schematic illustration of mass tolerant database method. A) with the standard database search method intact mass spectrum of a peptide is measured at a mass accuracy of <10 p.p.m. However, often a peptide carries unknown modifications, which will lead to a change of the detected mass value for both its precursor peptide ions (dashed lines to solid lines in the spectrum) and its fragment peptide ions. B) unexpected modifications on a peptide result in a mass shift of its molecular weight. Standard database search (green box) is not able to identify such peptide due to its stringent precursor peptide tolerance. In the mass-tolerant database search method (orange box), algorithms accept a larger deviation of detected precursor ion mass to theoretical peptide mass, enabling peptide with modification to be identified. The type of modification could be inferred from the value of the mass deviation, and the site of the modification could be further deduced from the MSn spectra. Adapted from [Skinner and Kelleher, 2015].

peptide isoforms not being present in sequence databases as well as unknown protein PTMs.

Proteins are frequently being modified, both in vivo inside cells and in vitro during sample handling or during measurement. It is estimated that for each unmodified tryptic peptide detected in a shotgun proteomics study there are on average 8-12 modified forms of the same peptide [Nielsen et al., 2006]. Normally, the modified peptides and proteins are studied in a targeted approach by adding specific enzyme inhibitors and including additional enrichment steps during experiment. However, protein PTMs that are not of interest in a given study are usually omitted during database search. Spectra originating from those modified proteins hence remain unidentified. Another major reason underlying the unmatched spectra in shotgun proteomics data is sequence variation not being present in databases. Underlying reasons contributing to the amino acid sequence variation include genetic polymorphisms, peptide isoform and also sequencing errors. Sequence variation from genetic polymorphisms may reduce the lifetime of proteins or cause protein malfunction. The sequence variation has been investigated through applying de novo sequencing algorithms and variant peptide databases in proteomics studies.

Many solutions have been created to address those issues in peptide identification. For example, the peptide sequence tagging methods first infer a short peptide sequence, i.e. the tag from the tandem mass spectrum, and then search in the database for the complete peptide sequences that match with the tag as well as the sequence masses flanking it. The sequence tag methods can identify a peptide in the presence of an unknown PTM or an amino acid substitution between an entry in the sequence database and the measured peptide, but this method is limited to spectra where the partial sequences can be derived either manually or by simple algorithms. A different approach to address the spectrum identification problem is to use de novo algorithms. The de novo methods in proteomics deduce peptide sequences directly from deciphering the spectrum without resourcing to a sequence database. Open source software such as SeqMS and commercial software such as Peaks are examples of this technique. However, the accuracy of the de novo methods are restricted by the accuracy of the MS instruments and the peak intensity of the MS results.

As a result of the rapid improvement on the MS instrumentation, such as the invention of the Orbitrap mass analyzer, the peptide mass can be determined with an accuracy of up to sub-p.p.m level. With the wide application of tandem MS/MS, it has become a routine to obtain such accuracy even for the fragmented peptides. Such high mass accuracy has enabled researchers to apply a wide tolerance during database search for the precursor ions and to unbiasedly detect peptides with modifications. Large precursor tolerance searches have been used in top-down proteomics for protein identification. It has also been used in bottom-up proteomics and tested on a dataset of around 10,000 peptides [Nielsen et al., 2006]. Chick et al. adapted the fragment ion mass-tolerant database search method and

extended it to identify modified proteins from human HEK 293 cell line [Chick et al., 2015]. The principle of the mass-tolerant database search method is illustrated in Figure 4. In total, 184,000 additional peptides were identified with the mass-tolerant database search method, resulting in an increase of 46% peptide identification. The modifications that the newly identified peptides carried were mostly from three sources, endogenous modification, chemical modifications introduced during experiments, and modification introduced by MS. It was estimated by Chick et al. that approximately 20% of the newly identified peptides have modifications that are biologically relevant, such as phosphorylation, methylation, acetylation as well as many amino acid variations.

5.5 Protein post-translational modifications play a pivotal role in ageing

Proteins are susceptible to numerous changes which occur during ageing, either through a spontaneous reaction or as a consequence of physiologic or pathologic processes [Rattan et al., 1992]. Alterations in the rate and fidelity of protein synthesis, modification status and protein turnover are one of the main features of cellular ageing. Cellular ageing is affected by many intrinsic and extrinsic pathways, which have cross-talks at many different levels and are regulated through different protein modifications [Oh et al., 2014]. Despite the vast number of protein modifications that proteins can carry, only a few well-known PTMs have been studied in relation to ageing in individual studies, such as phosphorylation, methylation, oxidation, ADP-ribosylation and glycation. Those modifications often occur on proteins that are involved in key cellular processes such as DNA synthesis, protein synthesis and degradation, energy metabolism, cytoskeletal organisation, and the components of the extracellular matrix. Hence, it is difficult to pinpoint any particular pathway or protein modification as the leading cause of ageing of a cell, though some appear to be more involved and may be the key regulators at the crossroad of different pathways.

5.5.1 Phosphorylation

Phosphorylation is one of the most widespread protein PTMs, through which functions of protein are regulated in response to external and internal stimuli. The dynamic of protein phosphorylation is actively regulated by protein kinases and phosphatases, two groups of enzymes that catalyse the transfer of a phosphate group to a substrate and the removal of a phosphate group from a substrate. Proteins can be phosphorylated on nine amino acids, namely tyrosine, serine, threonine, cysteine, arginine, lysine, aspartic acid, glutamic acid and histidine. Among the nine modifiable amino acids, serine, threonine and tyrosine phosphorylation are predominant in eukaryotic cells [Lehninger et al., 2005]. The cycle of substrate phosphorylation by kinase and phosphatase starts from binding of ATP to the active site of a kinase, followed by binding of the substrate to the active site of the kinase. The γ phosphate of ATP is then transferred to serine, threonine or tyrosine on the substrate from the kinase, and then substrate is released from the kinase. Subsequently, binding

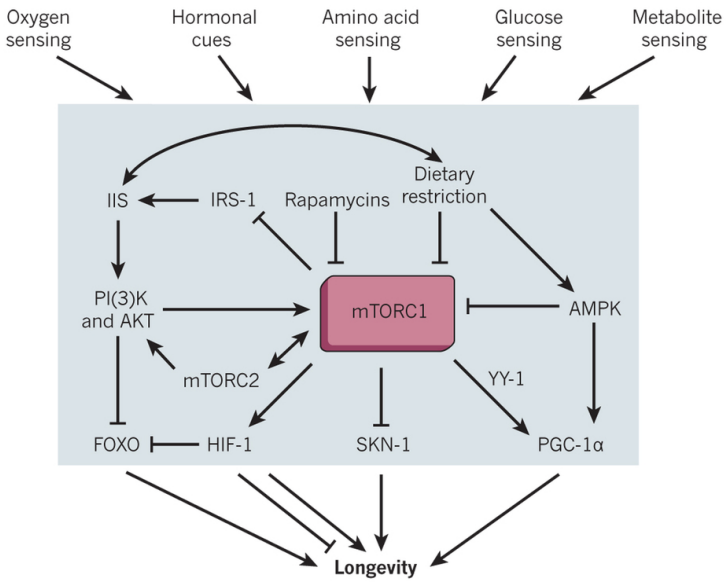


Figure 5: Interplay between mTOR and longevity pathways through phosphorylation. The serine/threonine protein kinase mTORC1 is one of the central hubs of several longevity pathways. Activity of mTORC1 is subjected to various stimulations such as changes of nutrient and hormone levels. The activity of mTORC1 is directly related to longevity as inhibition of mTORC1 by dietary restriction and rapamycin, or by activation of AMP-activated protein kinase (AMPK) results in increased lifespan in model organisms, such as *S.cerevisiae*, *C.elegans*, *D.melanogaster* and mice. Inhibition of mTORC1 promotes stress response transcription factor SKN-1 and leads to an increased lifespan. mTORC1 also negatively regulates longevity-related phosphoinositide 3-kinase(PI(3)K), protein kinase B and PI(3)K-AKT pathway through inhibition of insulin receptor substrate 1(IRS-1). In addition, activation of hypoxia-inducible factors 1(HIF-1) by mTORC1 under hypoxic condition also contributes to longevity through inhibiting FOXO family proteins, the PI(3)K and AKT mediated transcription factors. The relationship between mTORC2 and mTORC1 remains unclear, but it is known that mTORC2 contributes to longevity through inhibiting FOXO3A via PI(3)K-AKT pathway. Adapted from [Johnson et al., 2013].

of phosphatase to the phosphorylated substrate catalyses the hydrolysis of a phosphomonoester, resulting in the removal of the phosphate moiety from the substrate.

Based on the information from human genome, Manning et al. identified in total 518 protein kinases, constituting about 1.7% of all human genes. The human protein kinases can be categorised into seven major types based on their sequence conservation: AGC, which includes protein kinase A (PKA), protein kinase G (PKG) and protein kinase C (PKC) families; CMGC, which includes cyclin-dependent kinases (CDKs), mitogen-activated protein kinases (MAPKs), glycogen synthase kinase (GSK) and CDK-like kinases; Ca2+/calmodulin-regulated kinases (CAMK), casein kinase 1 family (CK1), related to yeast sterile kinases (STE) as well as tyrosine kinases and tyrosine kinase-like

(TKL) [Manning et al., 2002b, Manning et al., 2002a]. Kinases that are serine-threonine specific can be sub-divided into three classes: the basophilic kinases (e.g. PKA, PKD, AKT, CAMK2, AURORA and CHK) favour basic residues around the phosphorylation site; acidophilic kinases (e.g. CKI and CKII) favour acidic residues near modification site; proline-directed kinases (e.g. Cdks and MAPKs) require a proline residue directly at the C-terminal to the phosphorylation site [Kreegipuu et al., 1998]. Similarly, based on the human genome, recently Chen et al. identified a human phosphatome of 264 phosphatases, representing 1.2% of the human proteome [Chen et al., 2017]. Based on substrate specificity, phosphatases can be divided into three families: the phosphoprotein phosphatase (PPP) family (e.g. PP1, PP2A and PP2B) and the metallo-dependent protein phosphatase family (e.g. PP2C), which are responsible for the dephosphorylation of the majority of phospho-serine and phospho-threonine; and the protein tyrosine phosphatase (PTP) family, which dephosphorylate phospho-tyrosine. A subfamily of PTP is able to dephosphorylate all three phospho-amino acids [Barford et al., 1998, Johnson, 2009].

Signalling pathways regulated by phosphorylation span from but are not limited to metabolism, proliferation, differentiation, motility, organelle trafficking, membrane transport, immunity and memory formation. One example that has direct implication in ageing is the serine/threonine-protein kinase mTOR signalling pathway (Figure 5). mTORC1 directly and indirectly regulates the phosphorylation of at least 800 proteins. The mTOR pathway is activated by growth factors and also in states of high nutrition [Johnson et al., 2013]. Phosphorylated mTOR complex 1 (mTORC1) up-regulates protein synthesis by phosphorylating components of mRNA translation and ribosome synthesis machinery. Mutation or RNA interference knockdown of mTOR and direct inhibition of mTOR activity through rapamycin or calorie restriction (CR) have all been demonstrated to extend the lifespan of several model organisms, including yeast *S.cerevisiae*, nematode *C.elegans*, fruitfly *D.melanogaster* and mice [Fabrizio et al., 2001, Vellai et al., 2003, Kapahi et al., 2004, Harrison et al., 2009, Fontana et al., 2010]. Down-regulation of mTORC1 activity through activation of AMP-activated protein kinase (AMPK) has also been shown to promote longevity. In contrast to the mTOR pathway, the AMPK pathway is activated in response to stress conditions, such as low glucose, hypoxia, ischemia, and heat shock, during which the supply of cellular ATP is depleted [Mihaylova and Shaw, 2011, Carling et al., 2011]. AMPK phosphorylates tuberous sclerosis protein 2 (TSC2), an upstream inhibitor of mTORC1, resulting in down-regulation of mTORC1 activity. Furthermore, the regulatory protein of mTORC1 raptor can also be inactivated through direct phosphorylation by AMPK [Inoki et al., 2003, Gwinn et al., 2008]. In addition, it has also been reported that AMPK can extends the lifespan of *C.elegans* through CREB-regulated transcriptional coactivator 1 (CRTC-1), which is independent of mTORC1 activity [Mair et al., 2011].

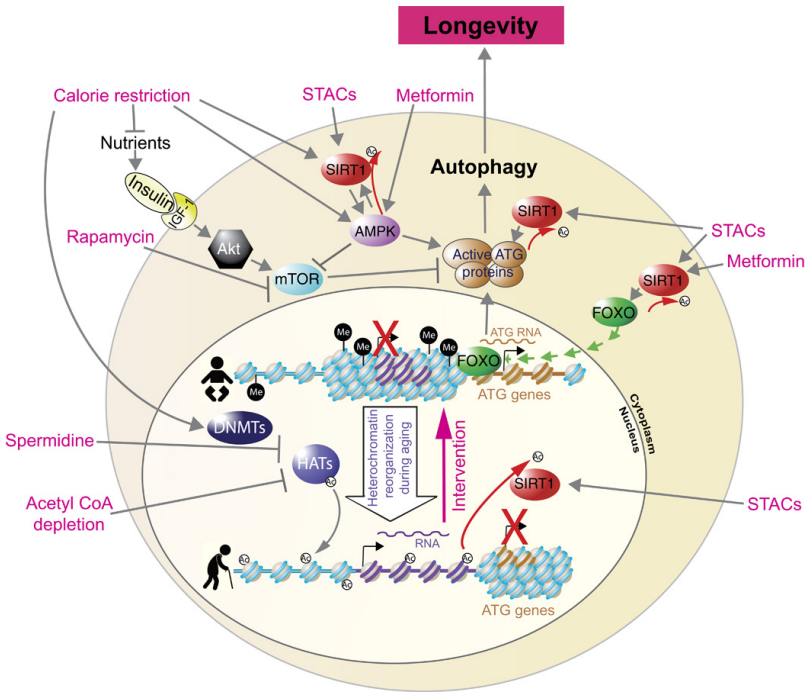


Figure 6: Intervening longevity pathways through acetylation. Interventions such as calorie restriction(CR), acetyl-CoA depletion, inhibition of HATs and activation of SIRTs are known to increase lifespan. One speculation is that different external stimuli for lifespan extension partially act through acetylation (presumably on histones) to stimulate autophagy. SIRTs induce FOXO-mediated expression of autophagy pathway components through deacetylating major players of the autophagy induction pathway. In addition, down-regulation of SIRT1 mediated deacetylation of histone H4K16, serves as an important negative feedback regulatory loop during autophagy to prevent overstimulation of autophagic flux [Füllgrabe et al., 2013]. The pink writing highlights interventions that are shown to extend life span in model organisms, arrows indicate stimulating effects, and blocked lines indicate inhibitory effects. Adapted from [Pal and Tyler, 2016].

5.5.2 Acetylation

Acetylation is a protein PTM of equal importance in cellular ageing. Analogous to phosphorylation, protein acetylation is also tightly regulated by enzymes. Protein can be acetylated on its N-terminal residue but also lysine, serine, threonine and cysteine. Apart from the co-translational acetylation at N-termini, lysine is the most commonly detected residue for acetylation [Drazic et al., 2016]. Lysine acetyltransferases (KATs) transfer acetyl group to the substrates, whereas lysine deacetylases (KDACs) remove acetyl group from the substrates [Roth et al., 2001, Lee and Workman, 2007]. To date, there are 22 different KATs from three families identified in human and mouse [Berndsen and Denu, 2008]. The KATs can be categorised into three types based on their structures, GCN5, CBP/p300 and MYST11. All KATs utilise acetyl-coenzyme A (acetyl-CoA) as an essential cofactor to donate an acetyl group to the target lysine residue on substrate. Acetyl-CoA is required as a cofactor for KAT-mediated lysine acetylation, but can also acetylate lysine non-enzymatically. For KDACs there have been 18 discovered in human and mouse so far. The KDACs can be classified into two families with distinct catalytic mechanisms,Zinc dependent histone deacetylases (HDAC1-11) and nicotinamide adenine dinucleotide (NAD+) dependent sirtuin deacetylases (SIRT1-7) [Haberland et al., 2009, Finkel et al., 2009]. The HDACs are predominantly present in nucleus and cytoplasm. The catalytic activities of HDAC1, HDAC2 and HDAC3 increase upon binding of inositol-1,4,5,6- tetraphosphate, and are inhibited by metabolites such as D-β-hydroxybutyrate (βOHB), L-carnitine, pyruvate and sphingosine-1- phosphate (S1P) [Shimazu et al., 2013, Huang et al., 2012,Hait et al., 2009,Thangaraju et al., 2009,Millard et al., 2013]. The SIRT family deacetylases are mostly detected in nucleus. Deacetylation by SIRTs is dependent on its co-enyzme NAD+ [Imai et al., 2000]. Sirtuin-catalysed deacetylation converts NAD+ to nicotinamide, and conversely nicotinamide acts as a competitive inhibitor for the catalytic activities of SIRTs [Bitterman et al., 2002, Avalos et al., 2005].

Studies on protein acetylation have suggested that KATs and KDACs play multiple roles in regulating ageing process (Figure 6). Using HAT inhibitor spermidine or reducing the HAT co-factor acetyl-CoA, it has been demonstrated that reduction of global histone acetylation increases lifespan in *D.melanogaster*, *C.elegans* and mammalian cells, and also increase resistance to stress in *S.cerevisiae* and mammalian cells [Eisenberg et al., 2009,Eisenberg et al., 2014]. Similarly, deletion genes encoding HDACs that remove histone acetylation markers such as H3K56Ac, results in a shortened life span in *S.cerevisiae*. In mammalian cells, changes in acetylation status often induce autophagy, which modulates longevity [Mariño et al., 2014]. It is known that autophagy-induced lifespan extension is often mediated through AMPK and TOR pathways [Gelino and Hansen, 2012, Morselli et al., 2010]. Changes of acetylation induce autophagy but increase lifespan independently of mTORC1 and p53, as it has been observed that SIRT1 was not required for the induction of autophagy by rapamycin or p53 inhibition both in mammalian cells and in *C.elegans*.

This suggests that acetylation affects different autophagy components than mTORC1 and p53 for the extension of lifespan. In agreement with the observations, it has been found that induction of autophagy through restricted level of acetyl-CoA alone can also extend lifespan in *D.melanogaster*. It has been found that SIRT1 stimulates autophagy through deacetylating key components of the autophagy pathway, such as autophagy-related protein 5 (Atg5), 7 and 8 [Ng and Tang, 2013]. The nucleus SIRT1 also induces FOXO-mediated expression of autophagy pathway components. In addition, histone acetylation is suggested to be involved in the inhibition of autophagy when a high concentration of acetyl-CoA is present [Mariño et al., 2014]. Down-regulation of SIRT1 mediated deacetylation of histone H4K16 provides a negative feedback regulatory loop, which prevents overstimulation of autophagy [Füllgrabe et al., 2013].

5.5.3 Methylation

Methylation is another common PTM that has wide implications for protein functions. Recent proteomics studies on human cells by Larsen et al. estimated that protein methylation to be as wide-spread as phosphorylation and ubiquitination [Larsen et al., 2016]. Protein methylation mostly occurs on arginine and lysine residues [Comb et al., 1966,Aletta et al., 1998]. Protein methyltransferases are enzymes that catalyse the transfer of methyl group to substrate proteins. They are classified into two types, lysine methyltransferases (PKMTs) and arginine methyltrasferases (PRMTs). PKMTs can be further divided into two different groups, the SET-domain protein methyltransferase superfamily, including proteins known to methylate histones such as SUV39, SET1, SET2, EZ, RIZ, and non-SET-domain proteins, such as DOT1-like histone H3K79 methyltransferase (DOT1L), methyltransferase-like 10 (METTL10) and METTL21A [Feng et al., 2002,Hamamoto et al., 2015]. The SET-domain proteins transfer a methyl group from S-adenosyl-L-methionine (AdoMet) to the amino group of a lysine residue on the histone or other substrates. The non-SET-domain protein catalyse the transfer of methyl group in a similar mechanism despite lacking a SET domain. For arginine methylation, the PRMTs catalyze the transfer of a methyl group from S-adenosylmethionine (SAM) to the guanidino nitrogen atoms of arginine for the formation of methylarginine. Based on structure, human PRMTs are classified in three groups, type I, II and III PRMTs [Blanc and Richard, 2017]. The type I (PRMT1, PRMT3, PRMT4 and PRMT6 and PRMT8) and type II enzymes (PRMT5 and PRMT9) all have central cavities and two active sites opposing each other to form a homodimer structure [Zhang and Cheng, 2003, Jain et al., 2016]. In contrast, type III enzyme PRMT7 does not possess a central cavity and acts as a monomer [Jain et al., 2016, Debler et al., 2016]. Both the type I and II PRMTs catalyse the formation of monomethyl arginine. In addition, the type I PRMTs can further add a methyl group to the monomethyl arginine to form an asymmetric dimethyl arginine whereas the type II enzymes forms an symmetric dimethyl arginine. Type III enzyme PRMT7 forms monomethylarginine. The added methyl groups can be removed by protein demethylases such as lysine-specific demethy-

lase 1 (LSD1 or KDM1A) and Jumonji C (JmjC)-domain-containing proteins [Shi et al., 2004, Tsukada et al., 2006].

Protein lysine methylation is known for its regulatory functions in gene transcription as PKMTs are the main enzymes catalyse histone methylation, whereas arginine methylation has been reported to regulate cellular activities such as RNA processing, gene transcription, DNA damage repair, and signal transduction [Hamamoto et al., 2004]. Enzymes regulating protein methylation and demethylation have been implicated in disease progression, such as deregulation of PKMTs (SUV39H1 and EZH2) in tumorgenesis of lung and prostate cancer [Hamamoto et al., 2015].

Among the substrates targeted by methyltransferases and demethyases, histones are most frequently linked to ageing. Histone methylation is known to have an impact on longevity and this epigenetic trait is shown to be heritable in *C.elegans* [Greer et al., 2011]. Studies in model organisms showed an increase of activating histone methylation sites and a corresponding decrease in repressive histone marks during ageing, suggesting a global alteration of chromatin structure during the ageing process. In addition to histone methylation, which is mostly regulated by PKMTs on lysines, it is also known that arginine methylation plays a role in delaying ageing. Arginine methylation is required to maintain the differentiation capacity of various stem cells. The loss of PRMTs in transgenic mice, including PRMT1, CARM1, PRMT5, PRMT6 and PRMT7, have all been linked to cellular senescence and premature ageing phenotype [Gao et al., 2016, Pang et al., 2013, Banasavadi-Siddegowda et al., 2016, Neault et al., 2012, Blanc et al., 2016]. PRMTs are also associated with ageing for their function in cellular senescence. In muscle stem cells and hematopoietic stem cells, loss of arginine methylation is linked to reduced differentiation capacity and diminished stem cell population, implying a link between epigenetic modulation of arginine methylation and ageing [Blanc and Richard, 2017, Liu et al., 2015].

5.6 Bone marrow as a critical component of the immune system and its role in ageing

The immune system is a multi-layered defence mechanism intensely shaped by evolutionary forces in the organism's defence against pathogenic factors. In human, the immune system is an interactive network composed of lymphoid organs, immune cells, various humoral factors and cytokines. There are two types of immune responses, the innate and the adaptive. During an event of pathogenic invasion, elements that comprise the innate immune system, namely the neutrophils, monocytes, macrophages and the complement system will provide rapid response. In contrast, the adaptive immune system takes days and weeks to develop and respond. The adaptive immune response involves antigen-specific reactions mediated by T lymphocytes (T cells) and B lymphocytes (B cells). As the major

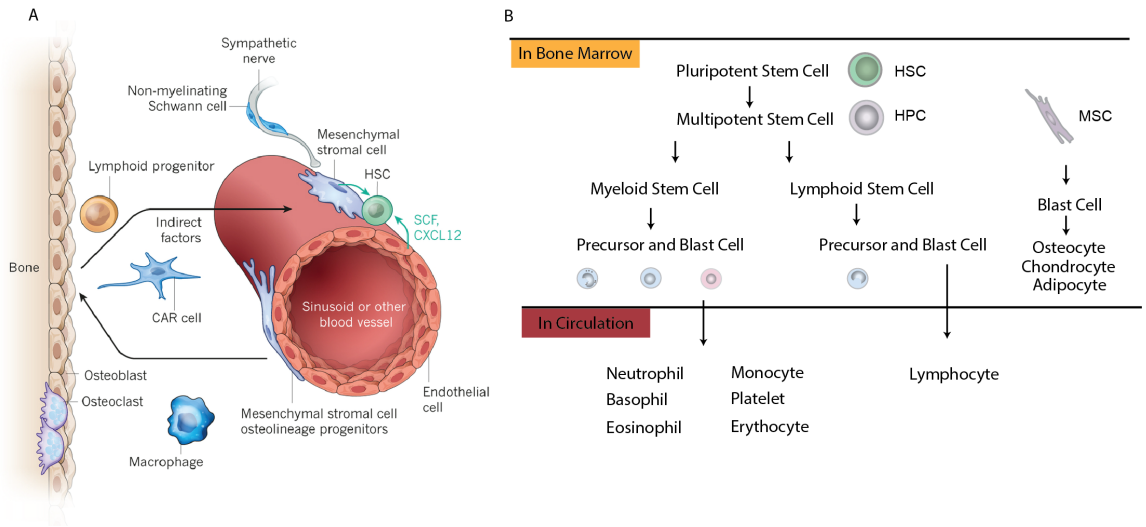


Figure 7: Bone marrow microenvironment and its composition. A) the bone marrow microenvironment is composed of parenchyma and stroma. The parenchyma contains haematopoietic stem cells (HSCs) and progenitor cells at various differentiation stages. The HSCs are found to preferentially locate near the sinusoids. The stroma contains MSCs, which promote maintenance of HSCs by secreting soluble factors such as SCF64, CXCL12. Perivascular part of the stroma contains cells including CXCL12-abundant reticular (CAR) cells, which are destined to form bone in vivo. Other cells and structures that inhabit the bone marrow niche include sympathetic nerves, non-myelinating Schwann cells and osteoclasts. Osteoblasts do not directly promote the maintenance of HSCs but promote the maintenance and differentiation of subpopulations of lymphoid progenitors by secreting factors such as CXCL12. Early lymphoid restricted progenitors thus reside in an endosteal niche that is spatially and cellularly distinct from HSCs. Adapted from [Morrison and Scadden, 2014]. B) classical model of haematopoiesis predicts a dichotomous lineage restriction of multi-potent haematopoietic stem cell to lymphoid stem cell and common myeloid stem cell. The myeloid and lymphoid cells will further differentiate into various immune cells. The precursor cells of erythrocytes, granulocytes, monocytes, thrombocytes, and lymphocytes are all formed in the bone marrow, whereas the matured cells circulate in the blood stream. In addition, the bone marrow MSCs are precursor cells that differentiate to constitute the non-haematopoietic part of the bone marrow. Progenitor cells of MSCs include osteocyte, chondrocyte and adipocyte.

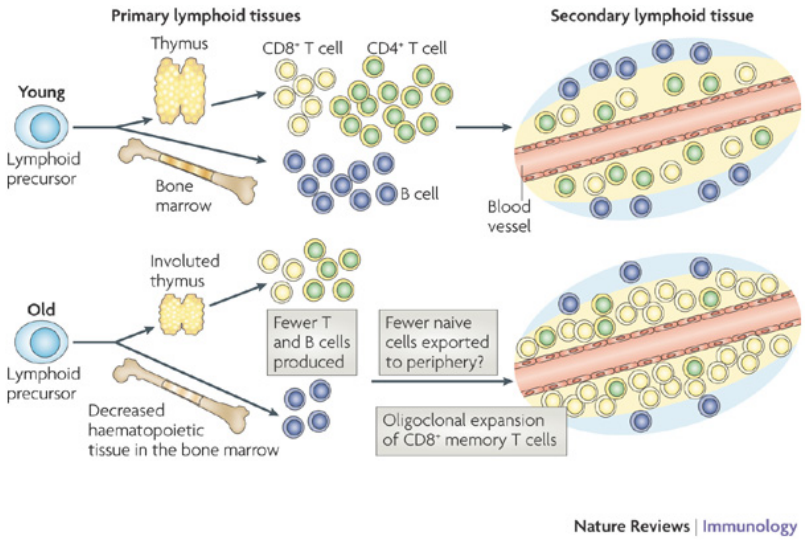


Figure 8: Effects of ageing on lymphocyte production and distribution of cells in secondary lymphoid tissues. Several events contribute to the decreased number of immune cells and the declined of function of adaptive immune system during ageing. Production of T cell is decreased as a result of thymic involution, which entails decreased mass of thymic cortical and medullary and an increased amount of fat. Production of B cells is also likely to decrease with age as content of haematopoietic bone marrow reduces with age. In addition, in the elderly population it is often observed that the CD8+ T cells clonally expand their population, resulting in a skewed proportion of T-cell repertoire and an increased number of terminally differentiated memory CD8+ T cells in peripheral blood. Adapted from [Dorshkind et al., 2009].

players in the cell-mediated immunity, both T cells and B cells are created inside the primary lymphoid organ, the bone marrow through differentiation of the haematopoietic stem cells (HSCs). In adults, upon differentiation from HSCs into lymphoid precursor cells, the lymphoid precursor cells migrate from the bone marrow to the thymus, where they complete the differentiation process and become mature T cells. In contrast, the precursor B cells are matured inside the bone marrow.

The human bone marrow is an haematopoietic organ composed of cortical and trabecular bone, cartilage, haematopoietic and connective tissues (Figure 7A) [Zhao et al., 2012, Morrison and Scadden, 2014]. Transcriptomics data shows that 58% of all human proteins are expressed in the bone marrow. In total 300 of those genes have elevated expression levels (at least five-fold) in bone marrow compared to other tissue types [Uhlén et al., 2015]. The parenchyma part of bone marrow harbours HSCs and various blood cells. As the most primitive cell type inside bone marrow, the HSCs differentiate into stem cells of the myeloid and lymphoid lineages. The myeloid and lymphoid stem cells will further differentiate and eventually become blood cells and platelets. The precursor cells of erythrocytes, granulocytes, monocytes, thrombocytes, and lymphocytes are all formed in the bone marrow. Therefore, the bone marrow contains blood cells in various stages (Figure 7B). On the other hand, stroma is the non-haematopoietic structural tissue of the bone marrow. The stroma does not directly participate in haematopoiesis but instead it provides the microenvironment and growth factors needed during haematopoiesis of the parenchymal cells. The stroma consists of mesenchymal stromal cells (MSCs) and various multi-potent cells, including osteoblasts, adipocytes, fibroblasts and endothelial cells. The bone marrow MSCs are the precursor cells that differentiate into various progenitor cell types, which constitute the bone marrow niche. In addition, the bone marrow MSCs are also known to have immunoregulatory functions [Chen et al., 2006]. They are able to modulate immune responses through direct interactions with different immune cells as well as through secreting soluble factors. In vitro, the bone marrow MSCs have shown to inhibit proliferation of T cells, B cells, natural killer (NK) cells and dendritic cells (DCs) [Devine et al., 2003, Ghannam et al., 2010, P De Miguel et al., 2012]. They are also able to inhibit functions of immune cells, such as secretion of cytokines and antibodies as well as antigen presentation. It has been suggested that the bone marrow MSCs function as immune modulators. They suppress the functions of the immune cells and play a role in transplantation tolerance, autoimmunity and cancer metastasis [Chen et al., 2006].

The immune system undergoes a series of alternations upon ageing, affecting the health condition and survival of the individual [Simon et al., 2015]. Overall upon ageing, the immune system displays declined fidelity and efficiency, the feature known as immunosenescence [Goronzy and Weyand, 2013, Oishi and Manabe, 2016]. For the adaptive immunity, the number of peripheral B cells does not decline with age but the number of T cells in-

creases upon ageing [Weiskopf et al., 2009, Fülöp et al., 2013]. During ageing, deterioration of the ability to establish immunological memory in response to new antigens accompanied by decreased production of cytokines by CD4 and CD8 T cells was also observed (Figure 8) [Walker and Slifka, 2010, Akue et al., 2012]. The innate immunity also experiences profound changes upon ageing. The lineage differentiation in haematopoiesis has been reported to skew towards myeloid populations in the elderly [Cho et al., 2008, Tang et al., 2013, Geiger et al., 2013]. Furthermore, the levels of pro-inflammatory cytokines IL-1β, IL-6, IL-18 and TNFα in the blood circulation were reported to increase in the aged immune system [Van Duin et al., 2007, Franceschi et al., 2007]. In the elderly persistent inflammation contributes to a chronic low-grade inflammation, which leads to the further development of other conditions such as cancer, cardiovascular disease, stroke, Alzheimer's disease and dementia [Chung et al., 2009, Childs et al., 2015, Oishi and Manabe, 2016]. In addition to the changes to the immune cells, immunosenescence may also evoke changes in the microenvironment. It has been suggested that interactions between immune cells and senescent non-immune cells with gut microbiota may play a role in immunosenescence [Hotamisligil, 2017, Smith et al., 2017]. Thus, intertwined mechanisms at cellular, tissue and systemic levels contribute together to the development of immunosenescence during ageing.

6 Aim and Study Design

6.1 Aim of the study

Ageing, which can be broadly defined as a progressive decline of functions of a living organism, is the results of a highly complex and regulated interplay between genetic and environmental factors. It involves dysfunctions at the cellular, tissue and organ levels [López-Otín et al., 2016]. Several candidate hallmarks for the primary causes of ageing, the antagonistic and integrative responses to ageing have been discussed and reviewed [López-Otín et al., 2013]. They are, DNA damage, telomere erosion, epigenetic drift, loss of proteostasis, stem cell exhaustion, altered intercellular communication and deregulated nutrient sensing. Due to its complex nature, there are still numerous questions that need to be addressed, in order to reach a comprehensive understanding on this biological phenomena. Advanced technologies such as next-generation sequencing accompanied with in vitro experiments and animal model studies can elucidate on the genetic and epigenetic changes associated with ageing. In addition to focusing on individual gene, protein or pathway, systematic approaches will provide integral insights on the mechanisms that link different biological processes to ageing. Using the human bone marrow as a model system, my PhD project aims at extending our current understanding of ageing in the human system. In particular, the role of protein PTMs in ageing in different human bone marrow derived cells are examined. Questions specifically asked are:

- How is the human bone marrow proteome modified?
- Which proteins or modifications are differentially regulated within the human bone marrow across different cell types?
- Is there any proteins and modifications that are associated with advanced age?
- Is there any proteins or modifications that are differentially regulated within the human bone marrow during ageing?

6.2 Study design

To address the questions, six different cell types residing in the bone marrow, namely haematopoietic progenitors (HPCs), lymphoid progenitors (LYMs), monocyte progenitors (MONs), erythroid progenitors (ERPs), granulocyte progenitors (GRAs) and mesenchymal stromal cells (MSCs) were analysed (Figure 9). Cells were taken from donated bone marrow aspirate at Heidelberg University Hospital under the approval by the ethics committee. Different cell populations were obtained via fluorescent activated cell sorting (FACS) using antibodies against different haematopoietic lineage markers. In total, 59 healthy donors were recruited for this study. The ages of the donors are in between 20 to 60. High-resolution liquid-chromatography coupled tandem mass spectrometry(LC-MS/MS)

was used to measure the proteome of the bone marrow niche. All samples were labelled with tandem mass tagging (TMT) for quantitative comparisons between individuals in the ageing analysis.

The acquired MS data was analysed using both standard and mass-tolerant database search methods. The standard shotgun proteomic database search method uses narrow mass tolerance windows for both precursor ions (MS1 search window) and fragment ions (MS2 search window). The narrow mass tolerance window ensures efficient and accurate prediction of the identity of the peptide and consequently the identity of the protein. Hence the standard database search method was employed to identify unmodified peptides and proteins. On the other hand, the mass-tolerant database search method uses an extended mass tolerance window for precursor ions and a narrow mass tolerance window for fragment ions. The increased precursor ion search window allows spectrum from peptides with unknown PTMs to be analyzed. Therefore, the mass-tolerant data base search was applied to discover abundant PTMs within the human bone marrow proteome. Subsequently, for confident identification of modified peptides, the standard database search was performed again, including PTM identified from mass-tolerant searches in the search parameter. This combined strategy of standard and mass-tolerant database search methods hence enabled me to unbiasedly capture abundant PTMs inside the human bone marrow as well as identify proteins with those modifications. Together with the quantitative labelling method, proteins differentially modified during ageing across all six cell types were analysed.

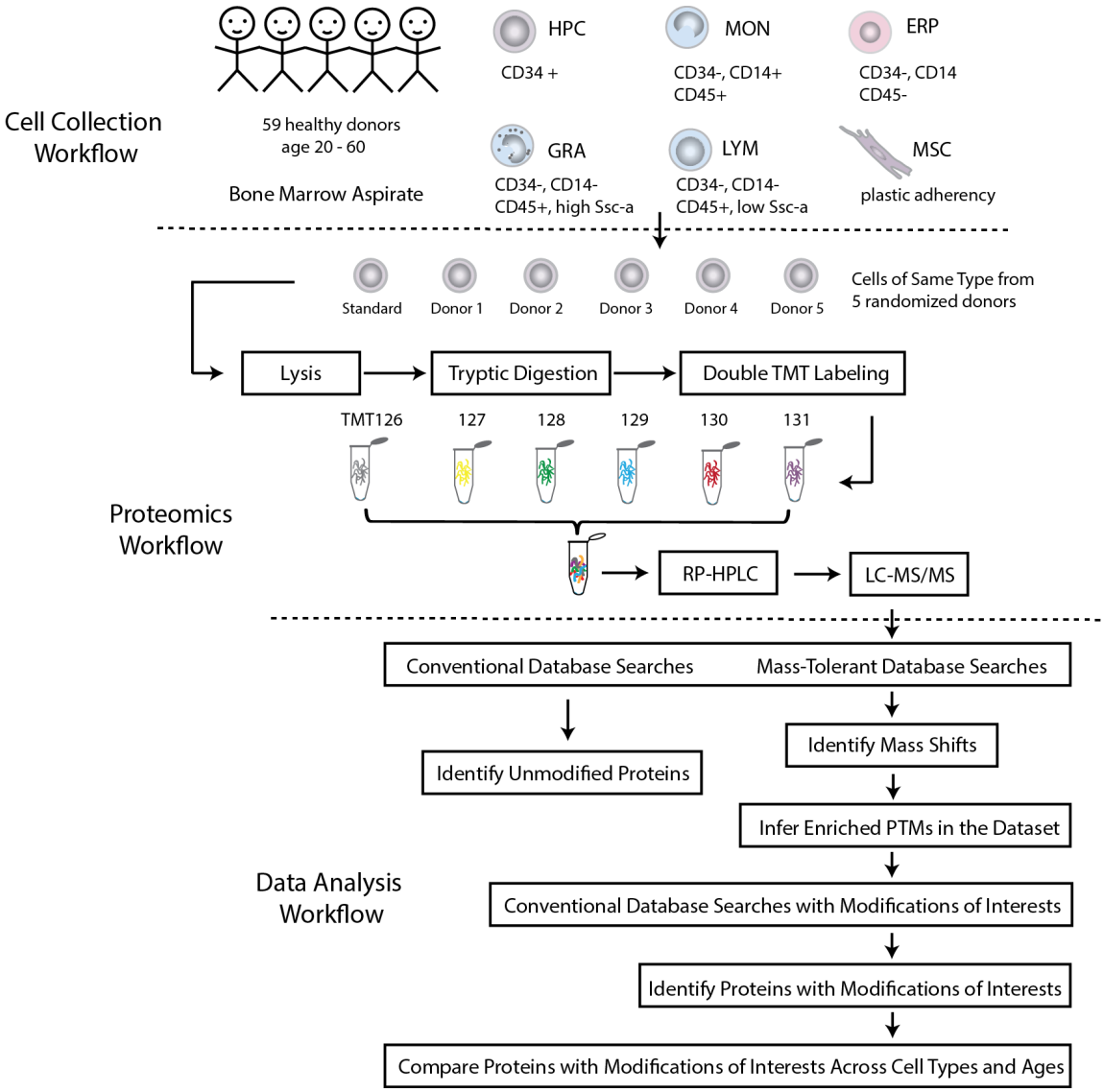


Figure 9: Project design. Cells were obtained from donated bone marrow material. Haematopoietic lineage cells were sorted via fluorescent activated cell sorting (FACS) whereas mesenchymal stromal cells were separated by in vitro seeding. Cells were lysed, digested and labelled with TMT-6plex labels. Cell-type specific internal standards were labelled with TMT 126, whereas age-randomised samples were labelled with TMT 127 to TMT 131. TMT sample mixtures were fractionated using orthogonal HPLC separation and measured by tandem MS/MS. Data was analysed using the combined mass-tolerant and standard database searches. Mass-tolerant search was applied to identify spectra with significant mass shifts, with which prominent protein PTMs were deduced. Proteins with modifications were then identified using standard database searches including PTMs of interests in the dynamic search parameter. Modified proteins were validated in-silico and investigated for differences across cell types as well as age-associated changes in human bone marrow. Abbreviations: HPC, haematopoietic progenitor cell; LYM, lymphoid progenitor cell; MON, monocyte progenitor cell; ERP, erythroid progenitor cell; GRA, granulocyte progenitor cell; MSC, mesenchymal stromal cell

7 Profile Protein Post-translational Modifications Inside Human Bone Marrow With Mass-tolerant Search Method

7.1 Summary

The project aimed at charting the human bone marrow proteome, and extending the understanding on ageing with the human bone marrow as a model. For the project, I performed the proteomics experiments and acquired the dataset. I further used a combined strategy to use both standard and mass-tolerant database searches to identify the frequently occurred PTMs in the human bone marrow proteome. Proteins with the frequently occurred PTMs were then identified and compared. In this chapter, overview of the human bone marrow proteome are presented. Optimisation steps for the parameters used in the mass-tolerant database searches are presented next, followed by the profiles of the most frequent occurred PTMs of the human bone marrow proteome. Interplay and cross-talks between different PTMs inside the human bone marrow are also analysed.

7.2 Result

7.2.1 Standard database search identified over half of the proteins from the expressed genes in the human bone marrow

In this study, the proteome of the human bone marrow was analysed using tandem mass spectrometry, coupled with TMT tandem tagging as quantification strategy. To resolve the human bone marrow proteome, bone marrow material was collected from 59 healthy donors with ages between 20 to 60 (Supplementary Figure 1). Cells from six major cell types inside the human bone marrow were analyzed. The six cell types analysed were haematopoietic progenitor cell (HPC), lymphoid progenitor cell (LYM), monocyte progenitor cell (MON), erythroid progenitor cell (ERP), granulocyte progenitor cell (GRA) and mesenchymal stromal cell (MSC). Cells from the haematopoietic lineage were isolated using fluorescence-activated cell sorting (FACS) whereas MSCs were obtained via in-vitro culture. Proteins obtained from the cell lysate were digested to tryptic peptides, which were further fractionated with orthogonal liquid-chromatography and measured by liquid chromatography-tandem mass spectrometry (LC-MS/MS) with higher energy collisional dissociation (HCD) in a quadrupole Orbitrap mass spectrometer.

The proteome of the six bone marrow derived cell types were analyzed using both standard and mass-tolerant database search methods. The standard database search method (with 15ppm precursor ion tolerance) was used to identify the unmodified proteins within the human bone marrow proteome. With a 1% protein false discovery rate (FDR), 7,322 proteins of the human bone marrow proteome were captured (Figure 10A). The number of the identified proteins represented 65% of the bone marrow transcriptome annotated in the Human Protein Atlas [Uhlén et al., 2015, Andersson et al., 2014]. Notably there

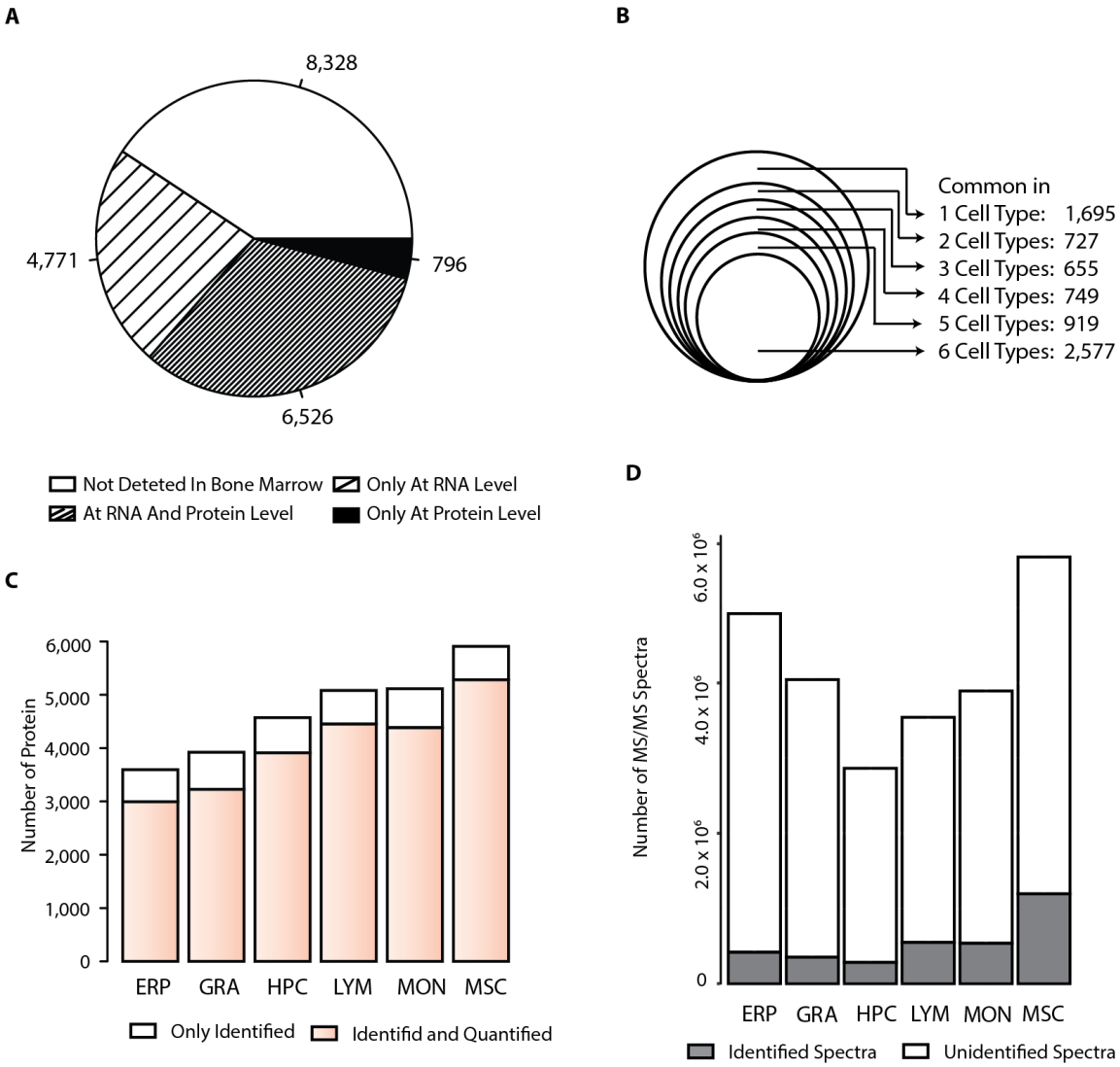


Figure 10: Overview of the human bone marrow proteome. A) 6,526 unmodified proteins from six different cell types were identified in this proteomic dataset, representing 65% of the bone marrow-expressed genes. The majority of the 796 proteins not detected in previous bone marrow transcriptomics and proteomics dataset, were extra-cellular matrix proteins expressed at high level in MSC population. B) in total, 2,577 proteins were identified in all six cell types, representing 35% of the bone marrow protein content. 1,695 proteins, 23% of the bone marrow proteome, were detected in only one cell types. C) the number of proteins that can be quantified for ERPs, GRAs, HPCs, LYMs, MONs and MSCs were 2,994, 3,229, 3,911, 4,451, 4,385 and 5,283, respectively. The number of quantified proteins represented 82 to 89% of the proteins identified in the study. The quantified proteins were analyzed for age-associated changes in the protein PTM in the human bone marrow. D) proteins identified in the study were from a fraction of the entire MS dataset. Using standard database search alone, 8 to 21% of the spectrum of the human bone marrow proteome was successfully mapped to protein database.

was a small set of proteins (n=796) within the human bone marrow that have not been previously identified in transcriptomics analysis. The 796 proteins are predominantly extracellular matrix proteins and membrane associated proteins, the majority of which were highly expressed in the MSC population. This presumably explains the discrepancy in the identification between ours and the previous study as MSC was not included in the study by Andersson et al.. For the individual cell types, 3,588 proteins were detected in ERPs, 3,917 proteins in GRAs, 4,568 proteins in LYMs, 5,077 proteins in MONs, 5,110 proteins in HPCs and 5,907 proteins in MSCs. 35% of the identified bone marrow proteins (n =2,577) belongs the core-set of the bone marrow proteome, i.e. proteins expressed in all six cell types, whereas 23% of the bone marrow proteome were detected in only one cell type (Figure 10B).

Depending on the cell type, 82 to 89% of the identified proteins were quantified based on the TMT 6-plex labels (Figure 10C). There were 2,994 proteins in ERPs, 3,229 proteins in GRAs, 3,911 proteins in HPCs, 4,451 proteins in LYMs, 4,385 proteins in MONs and 5,283 proteins in MSCs quantified based on TMT tandem mass tags. Proteins that can be quantified were used for quantitative comparisons of protein PTM levels across different ages in Section 9.

7.2.2 Unbiased identification of proteins with modifications in human bone marrow with mass-tolerant database search

The standard database search method successfully matched less than one fifth of the spectrum collected in this study (Figure 10D). Many of the unidentified spectra may originate from modified peptides, hence the mass-tolerant database search method was applied to uncover the modified fraction of the bone marrow proteome. The principle of mass-tolerant database search method is illustrated in Figure 4. In short, mass-tolerant database search method increases precursor ion tolerance during database search, which allows the search algorithm to be more tolerant to mass differences during peptide assignment. Therefore, peptides with unknown modification could be identified by comparing the mass shift of spectrum to the mass difference induced by known PTMs.

In order to unbiasedly uncover PTMs enriched inside human bone marrow, I first examined the range of the mass differences induced by known protein PTMs in Unimod. Unimod is a public domain database dedicated for protein modifications for mass spectrometry application, containing accurate values of mass differences introduced by natural and artificial modifications. 85% of the modifications annotated in Unimod database induce changes of mass value below 500 Da (Figure 11A). Since the majority of the known protein PTMs induce mass changes below 500Da, five different precursor ion tolerance windows (i.e.100 Da, 200 Da, 300 Da, 400 Da and 500 Da) were applied to capture as many potential modifications inside the bone marrow as possible. Noted that when a spectra was identified

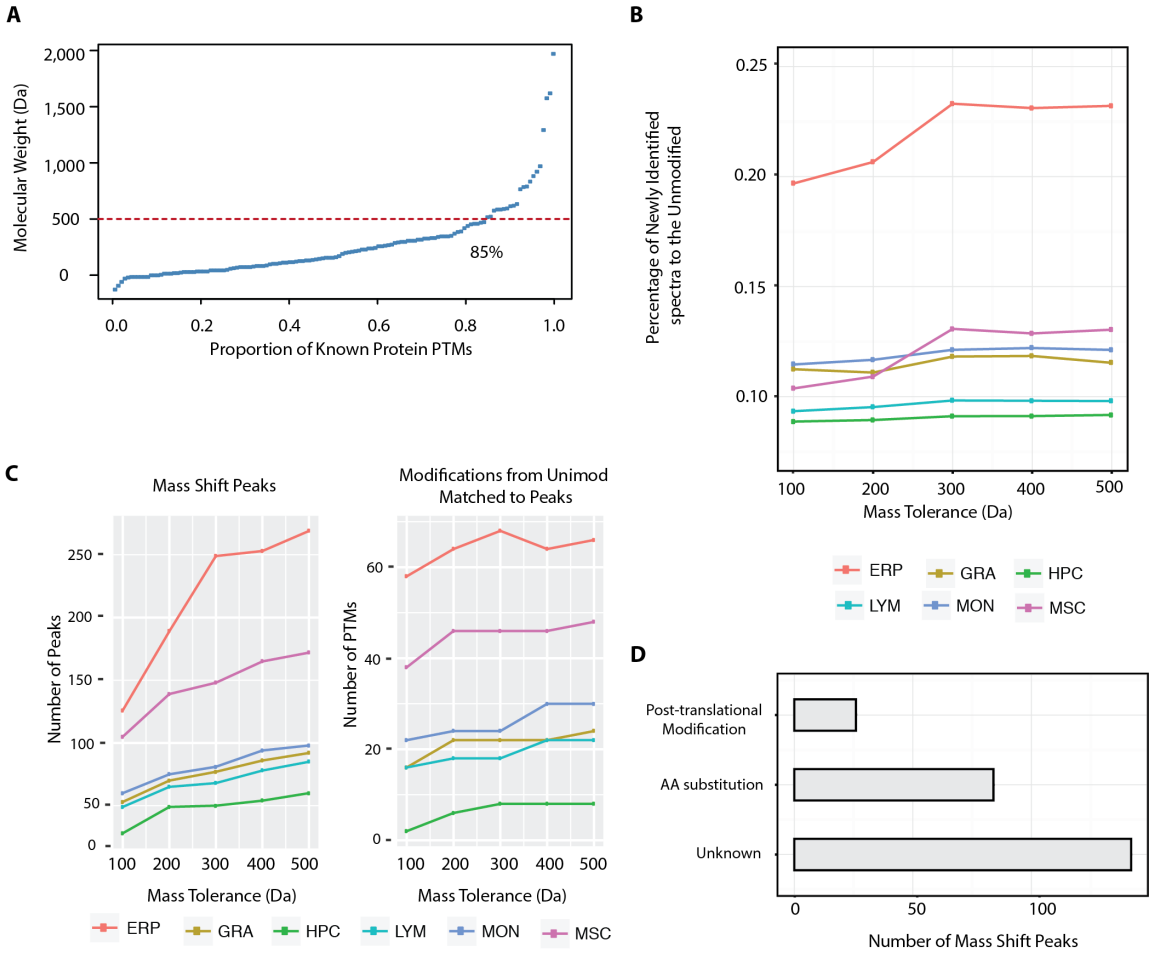


Figure 11: Identification of post-translational modifications enriched in human bone marrow proteome. A, 85% of all protein PTMs annotated in Unimod database introduce a mass shift less than 500 Da; B, the number of spectra matches to Uniprot database increased with precursor ion mass tolerance increasing from 100 Da to 500 Da; C, number of identified mass shift peaks increased with larger precursor ion mass tolerances, but the number of protein PTMs can be deduced through correlation with Unimod database saturated when the precursor ion mass tolerance reach 400 Da to 500 Da; D, 261 mass shift peaks were identified from the bone marrow proteome. 45 % of the mass shift peaks could be induced by protein modifications, including 26 protein PTMs and 84 amino acid (AA) substitutions.

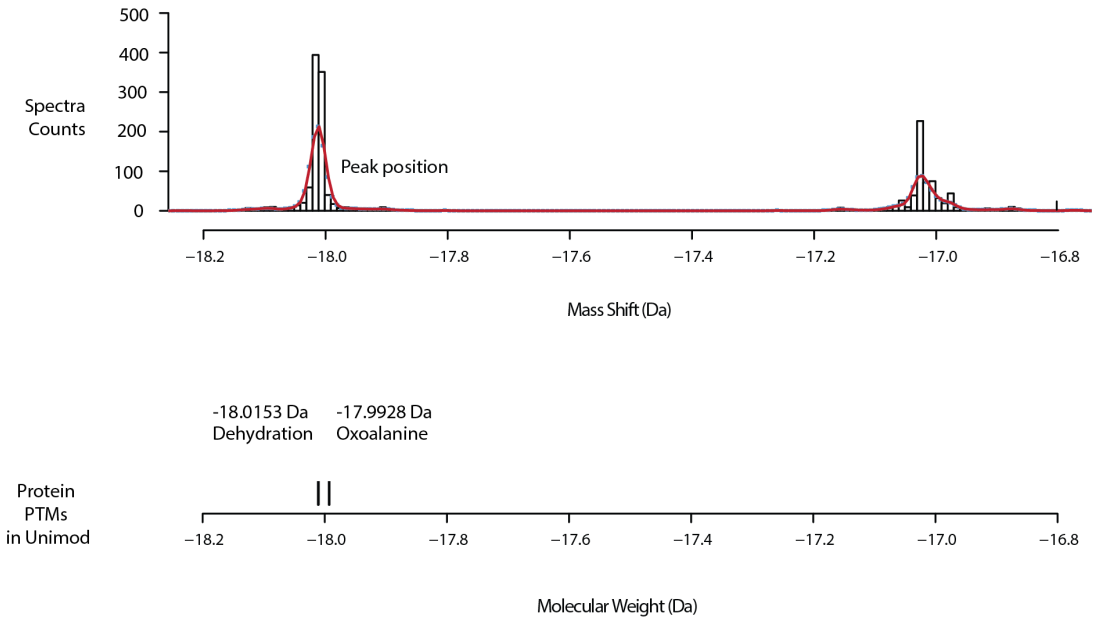


Figure 12: Deducing protein PTMs from the mass shift of spectrum. An example to demonstrate the peak calling of peptide mass shift and deduction of protein PTM from the peptide mass shift. The result of one TMT-6plex experiment from MSC using mass tolerant search with 500 Da precursor ion tolerance window is presented. Spectrum were clustered according to the mass shifts, i.e. the difference between the mass of the peptide captured by mass analyzer in mass spectrometer and the molecular weight of the peptide predicted to match the spectrum). Mass shift peaks were called out using Gaussian kernel function. Between the mass shift window -16.80 Da and -18.20 Da two distinct mass shift peaks were identified. The median of the mass shift of all spectrum within the two peaks was assigned as the mass shift values of the peaks. Given 0.01 Da mass deviation, known protein PTMs that are within the range of 0.01 Da of the mass shifts is dehydrated phosphorylation, which induces a molecular weight change of -18.0153 Da.

differently in standard and mass-tolerant database searches, only the identification result obtained from the standard database search was kept.

The number of spectrum matched to a peptide sequence in a protein database is known as peptide spectrum match (PSM). The number of PSM increased due to reduced stringency to the accuracy of the precursor ion mass during peptide spectrum matching (Figure 11B). For all six cell types, 8% to 23% additional spectrum were matched to protein database entries using mass-tolerant searches in comparison to the number of PSMs in standard database searches, suggesting many of the unassigned spectra are of good quality and may carry unknown modifications.

To determine the composition of the modified peptides within the dataset, peptides with mass shifts were collected and collated based on their mass shift values (bins of 0.01 Da) (Figure 12).The mass shift of a spectra is the difference between the molecular weight of the peptide calculated by the database search algorithm and that measured by the mass spectrometer. A collection of peptides with the same mass shift are most likely carrying the same PTM. Therefore, peaks were called from the spectrum with mass shifts. The average mass shift value of all spectrum within each peak was assigned as the mass shift value of the peak. The total number of peaks that passed the cutoff increased with precursor ion tolerance window size(Figure 11C).

To identify the modifications, the mass shift values of the peaks were compared to the mass changes induced by protein PTMs annotated in Unimod database. Two criteria were applied: 1) the mass difference between the mass shift value of a given mass shift peak is within 0.01 Da of the mass difference induced by the modification in Unimod database; and 2) for the peptides belonging to a given mass shift peak, the amino acid required for the known modification should be enriched in the peptide sequences. Similar to the number of peptide with mass shift, the number of matched Unimod modifications increased with precursor tolerance window size (Figure 11C). The most frequently occurred protein PTMs were summarized in Table 1. Modifications captured in all six cell types include various forms of oxidation, phosphorylation and methylation. Modifications detected in specific cell types were quinone modification (in ERPs) and glycosylation(in MSCs and GRAs). In total, there were 119 modifications deduced from the entire dataset, 26 of which are known protein PTMs (Figure 11D).

7.2.3 Validation of proteins identified with post-translational modifications

The identified modified peptides were further evaluated for accuracy. Being well studied and characterised PTMs, phosphorylation, acetylation and methylation were used for cross-checking for whether peptides identified with the three modifications have been previously detected in other studies. The identified phosphopeptides in bone marrow proteome

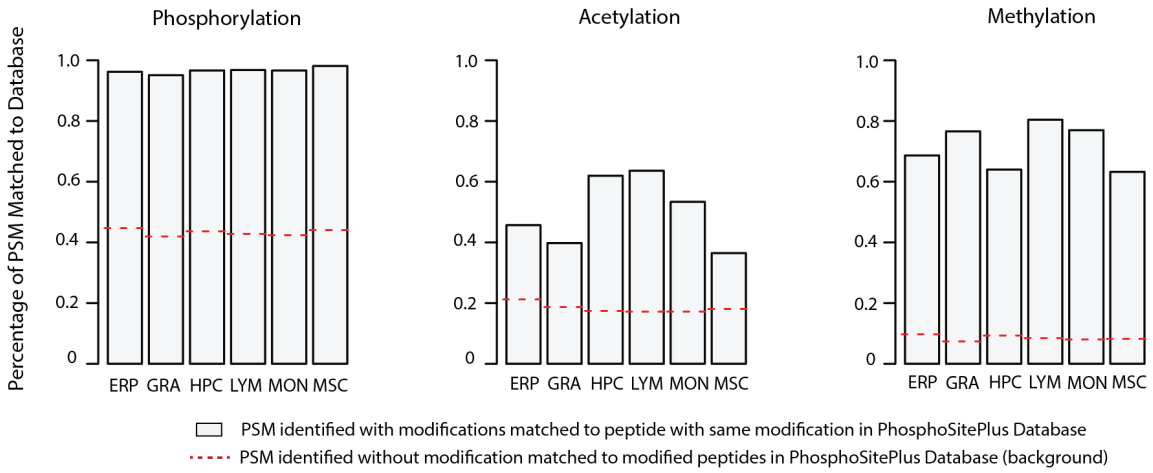


Figure 13: Validation of modified peptides with protein modification database. The identified phosphorylated, acetylated and methylated peptides were compared to curated protein phosphorylation, acetylation and methylation databases. 95% -98% of the phosphorylated peptides detected in this study have been previously identified in other PTM-targeted proteomics or biochemical analysis. Fractions of the acetylated and methylated peptides identified in the database were 36% -62% and 63 -82%. The peptide matching rates of all three modifications were approximately 2 times higher than the background peptide matching rates, which are 42 -45% for phosphorylation, 17% -21% for acetylation and 7% -10% for methylation, as indicated by the red dashed lines.

were compared with all phosphopeptides annotated in PhosphoSitePlus database (version of 2017.03.21) [Hornbeck et al., 2015]. Phosphorylation is by far the most extensively studied protein PTM. Expectedly, 95% to 98% of the phosphopeptides have been previously identified in other PTM-targeted proteomics and biochemical studies (Figure 13). In comparison, matching rate of unmodified peptides (background) to the phosphopeptides in PhosphoSitePlus database was only 42% to 45%. The peptide matching rates for acetylated peptides and methylated peptides were 36 to 62% and 63 to 82% (Figure 13). Though lower than phosphorylation, the peptide matching rates for the two modifications were still at least two fold higher than peptide matching rates of the unmodified peptides to the corresponding databases, which were 17% to 21% and 7% to10%, respectively. This indicated that the modified peptides identified through mass tolerant database search is of good quality.

7.2.4 Mass-tolerant database search identifies various forms of post-translational modifications

Various forms of phosphorylation and methylation were identified, which were dehydrated phosphorylation, dehydrated and oxidised phosphorylation, dimethylation, hydroxylated and deamidated methylation. The peptide matching rates for peptides with these modification were summarized in Supplementary Figure 2. The rates were lower than the classical phosphorylation and methylation but still more than 1.5 fold higher than background. Of note, the oxidized and dehydrated phosphorylation was inferred from the mass shift peak with a mass shift value of -2.02 Da. This modification was initially identified as 2-amino-3-oxo-butanoic acid modification using annotation from Unimod database. But most evidences showed that the modification that induces this mass shift is highly related to phosphorylation. Correlation analysis with other modified peptides identified in the study showed that peptides identified with 2-amino-3-oxo-butanoic acid modification have the closest relation with oxidised peptides and dehydrated phosphopeptides (with Pearson correlation coefficient of 0.42 for both). Comparison with PhosphoSitePlus database further showed that most of these spectrum have been previously identified in targeted phosphorylation studies (Supplementary Figure 2). The combination of oxidation and dehydration of phosphorylated residue would induce a monoisotopic mass shift -2.01565 Da on peptide, which is within 0.01Da tolerance of the mass shift value -2.02Da. In addition, this modification was reported to occur on tyrosine and serine residues, which overlap with residues for phosphorylation. Therefore, I reasoned that the modification that induces the mass shift of -2.02Da is oxidized and dehydrated phosphorylation.

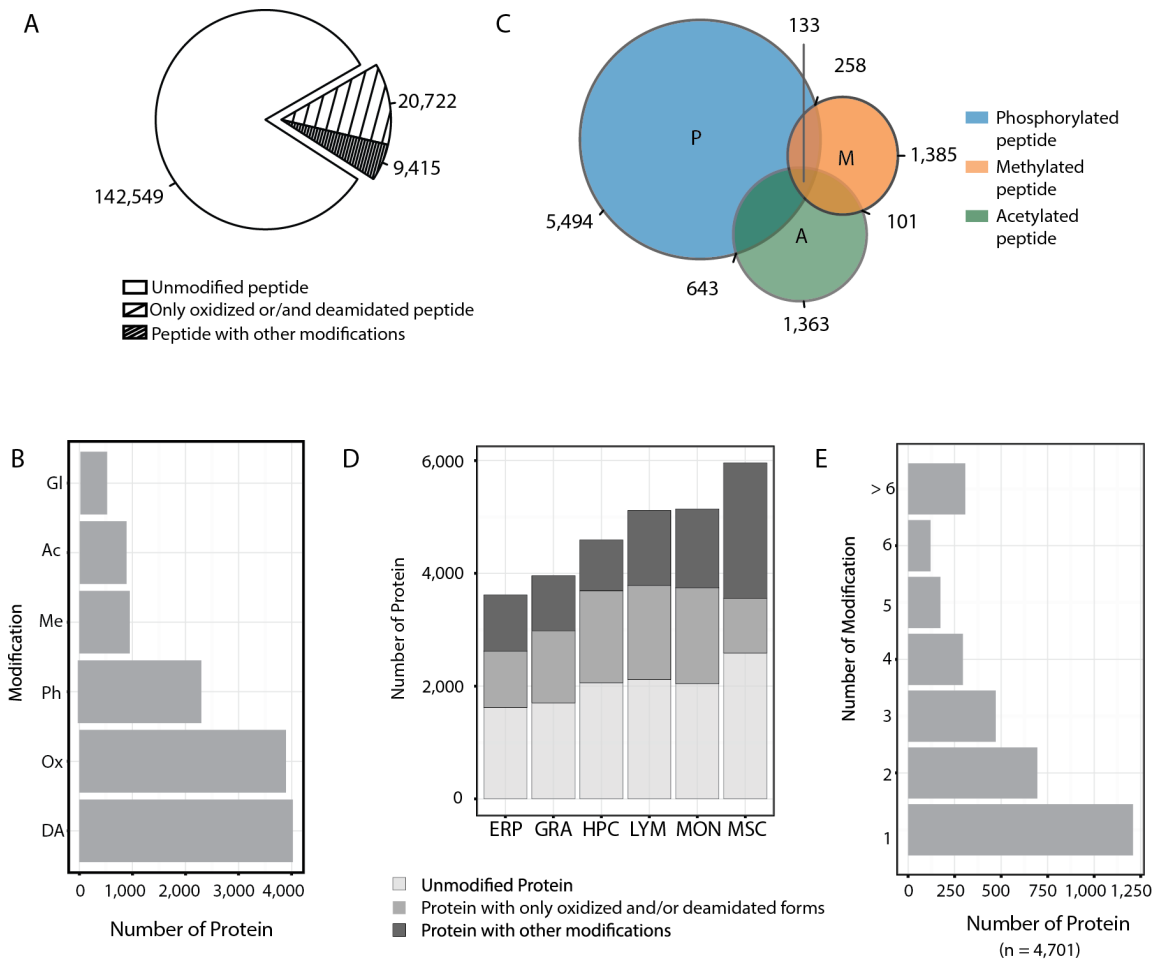


Figure 14: Common modifications inside human bone marrow. A) in addition to 142,549 unmodified peptides identified through standard database search, 30,137 modified peptides (16%) were identified through standard database search with modifications included in dynamic search parameter. A significant amount of modified peptides were oxidized and/or deamidated. B) deamidation (DA) and oxidation (Ox) were the most frequently identified PTMs in human bone marrow, followed by various forms of phosphorylation, methylation, acetylation and glycosylation. C) 6,528 phosphorylated peptides, 2,240 acetylated peptides and 1, 877 methylated peptides were identified across six cell types in human bone marrow. 5% of those modified peptides (n=1,534) were identified to have two and three different modified forms among the three modifications, which may originate from regions where proteins have PTMs cross-talks. D) the number of unmodified and modified proteins identified were at comparable levels across all six bone marrow-derived cell types. 1,997 proteins in ERPs, 2,256 proteins in GRAs, 2,995 proteins in LYMs, 3,093 proteins in MONs, 2,535 proteins in HPCs and 3,371 proteins in MSCs were modified, representing 55 to 60% of the proteome of each cell type. E) the majority of the proteins were identified with low frequency of modification. Abbreviations: Ac, acetylation; DA, deamidation; Gl, glycosylation; Ph, phosphorylation; Me, methylation; Ox, oxidation; P, phosphorylation; A, acetylation; M, methylation.

Table 1: Identification rates of the most frequent protein PTMs in human bone marrow

Modification	PSM	% of All PSMs	Peptide	% of All Peptides	Protein	% of All Proteins
Oxidation[1]	383,015	8.00	18,215	11.91	3,891	52.08
Deamidation	397,550	8.30	16,848	11.02	4,020	53.81
Phosphorylation[2]	34,118	0.71	6,528	4.27	2,324	31.11
Methylation[3]	21,221	0.44	2,407	1.57	1,157	15.498
Acetylation	11,429	0.24	2,240	1.46	879	11.77
Reduction	6,519	0.14	1,597	1.04	983	13.16
Glycosylation[4]	2,710	0.06	880	0.58	500	6.69
Carboxylation	1,364	0.03	331	0.22	224	3.00
Quinone	1,078	0.02	312	0.20	218	2.92

7.2.5 Post-translational modifications are prevalent across bone marrow cell types

From over 4.7 million high-confidence PSMs (with 1% protein FDR), 49,358 different modified peptides forms from 4,701 proteins were identified(Table 1). Overall 16% of the peptides in the bone marrow proteome were modified (Figure 14A). The majority of the modified peptides identified were oxidized or deamidated peptides. 31% of the modified peptides (n=9,415) from 3,269 proteins were modified by other common PTMs. Among those common PTMs, phosphorylation, acetylation and methylation were the most frequently detected modifications(Figure 14B). In addition to peptides modified by individual modifications, a fraction of the peptides (n=1,534) were modified with different types of PTMs (Figure 14C).

The proportion of modified proteins were at comparable levels across all six cell types, representing 55 to 60 % of the proteins identified in each cell type (Figure 14D). There were 1,997 modified proteins identified in ERPs, 2,256 proteins in GRAs, 2,995 proteins in LYMs, 3,093 proteins in MONs, 2,535 proteins in HPCs and 3,371 proteins in MSCs. Proteins modified by PTMs other than oxidation and deamidation were proportionally larger in MSC than in other cell types. In addition, proteins in bone marrow were observed to have low modification levels. The majority of the bone marrow proteins were identified with only one to two modifications (Figure 14E).

Phosphorylation, acetylation and methylation are the three most frequently detected modifications apart from oxidation and deamidation. Therefore, I focused on these three modifications in all the analysis presented in subsequent chapters.

[1] mono- and di-oxidation

[2] phosphorylation, dehydrated phosphorylation, dehydrated and oxidised phosphorylation

[3] mono- and di-methylation, hydroxylated and deamidated methylation

[4] fucose, hexose, acetylhexosamine and methylglyoxal

7.2.6 Modifications occur in specific sequence context

The same PTM occurs on different residues through various mechanisms and this could be reflected through the amino acid context surrounding the modification site. Therefore I further examined the amino acid preference flanking phosphorylation, acetylation and methylation sites.

The amino acid preference surrounding phospho-serine (Sp) and phospho-threonine(Tp) were highly similar (Figure 15A). Lysine (K) and arginine (R) were depleted in positions immediate to the modified residue and enriched at distant locations. The negatively charged aspartic acid (D) and proline (P) was highly enriched at adjacent positions. I observed an enrichment at +2 position for proline around phospho-tyrosine (Yp) residues, compared to the +1 preference in Sp and Tp. Most other amino acid distributions are similar between Yp and Sp/Tp .

The patterns of amino acid frequency observed for methyl-lysine (Kme) and methyl-arginine (Rme) were distinct from each other (Figure 15E). For peptides identified with Kme, S and R were highly enriched at -1 and -2 positions whereas K was enriched at both -2 and +2 positions. For peptides identified with Rme, methionine (M) was the preferred residue before the modification site (-1 and -2) whereas negatively charged residues such as K, R and histidine (H) appeared to be more frequent after the modification site (Figure 15F).

As each modified residues displayed a distinct pattern for the surrounding amino acids, this suggested that the regulations of these modification may be sequence specific. Along the same lines, I asked whether there is any specific enzyme generating the different PTMs across different sequence context. This is especially applicable for phosphorylation because the enzymes that target different phospho motifs are well studied.

Thus, I used pLogo to visualize the short linear pattern of amino acids surrounding phosphorylated sites. Results showed specific amino acid patterns surrounding different phosphorylation sites (Figure 16B). In agreement with the amino acid preference results presented before, P was significantly enriched at the +1 position of both Sp and Tp, in contrast to the preference at +2 for Yp. For all three phosphorylated amino acids, K was unfavored before the modification site.

7.2.7 Phosphorylation, acetylation and methylation occur frequently on functional domains in bone marrow

PTMs can occur on different regions of a protein to effect and modulate different biological functions. We can broadly understand these functions by 1) looking at which

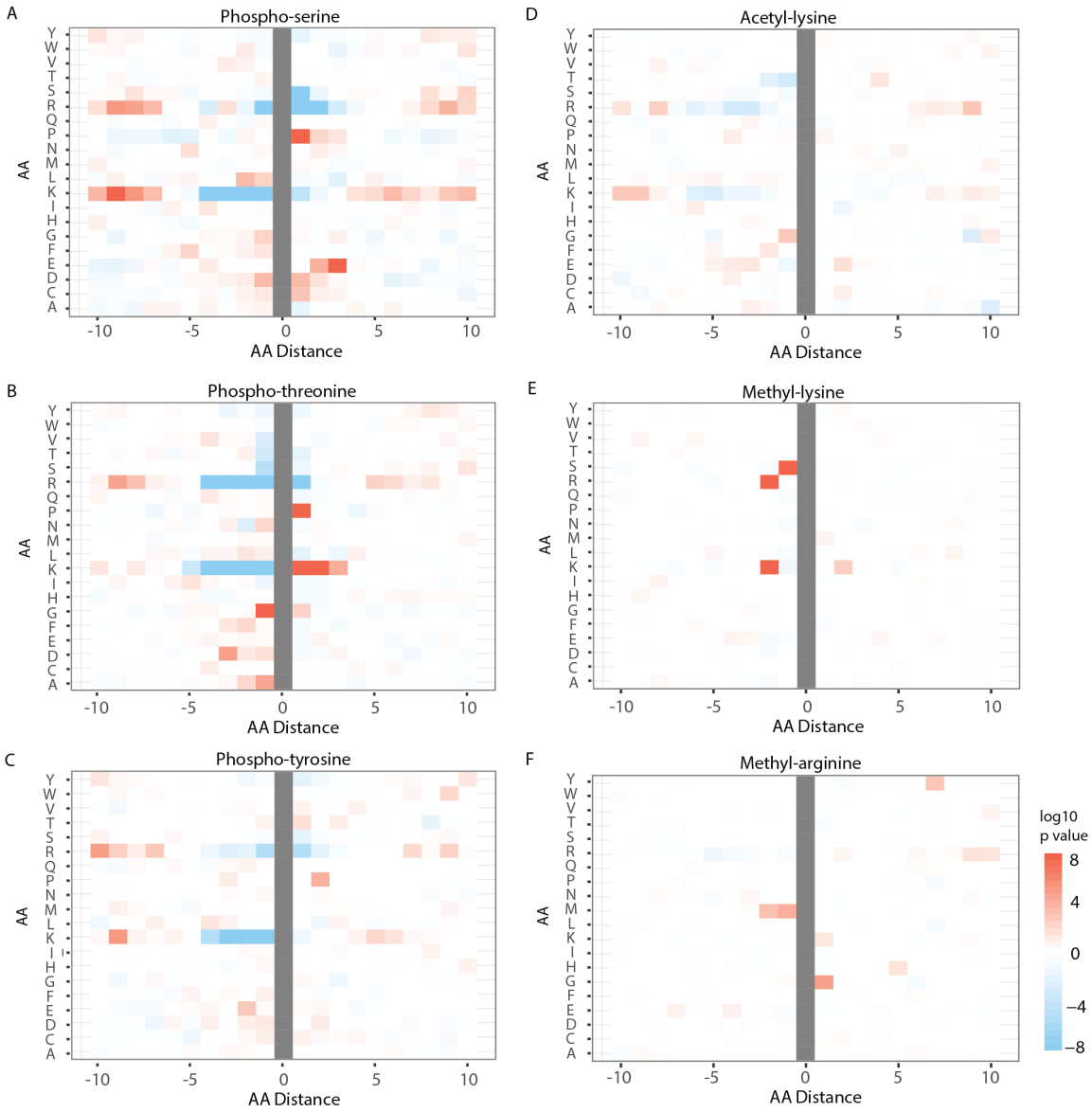


Figure 15: Amino acid preference neighbouring identified phosphorylation, methylation and acetylation sites. The relative frequencies of 20 amino acid residues flanking sites of phosphorylation (A-C), acetylation (D) and methylation(E-F) were shown. Each modified residues displayed a distinct preferences for the surrounding amino acids. Preference for proline close to modification sites, as well as depletion and enrichment of lysine and arginine were observed in phosphorylated peptides. Charged residues such as lysine, arginine and glutamic acid as well as hydrophilic residue threonine were preferred to be depleted before acetyl-lysine, whereas lysine and arginine were frequently located away from acetyl-lysine site. Methionine and glycine were highly enriched adjacent to a methyl-arginine but not methyl-lysine. Enrichment of hydrophilic residues such as histidine and tyrosine were observed in distant position from a methyl-arginine. Lysine and arginine as well as hydrophilic residue serine were either enriched directly before or after a methylation site.

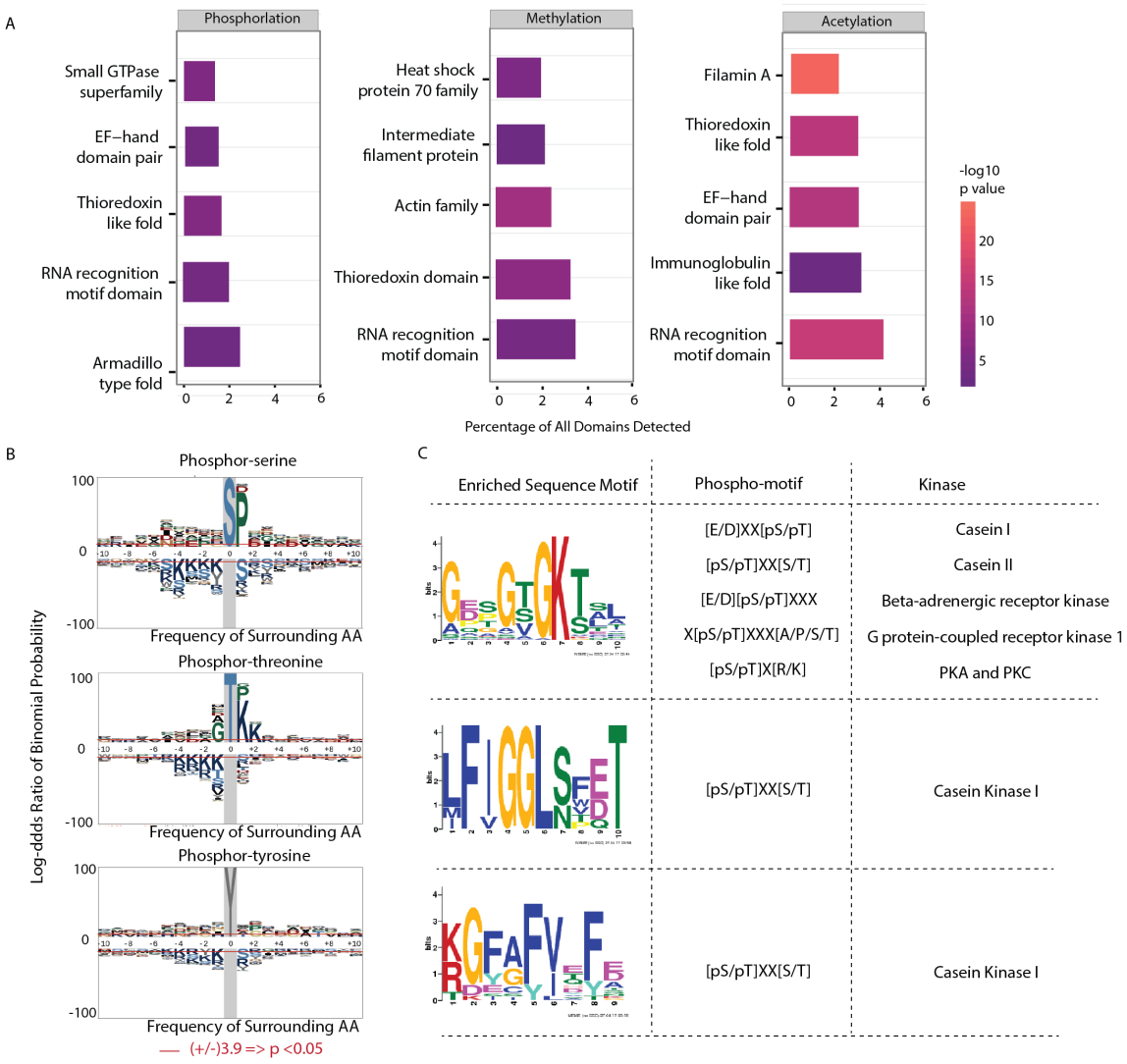


Figure 16: Functional analysis of phosphorylated, acetylated and methylated proteins. A) protein domains enriched in phosphorylated, methylated and acetylated peptides (top five most frequent modified domains, for structurally related domains, the most representative ones with the lower p values were chosen for display) were involved in various functions, including enzymatic activity, ion- and RNA binding and scalffolding. B) motif enriched revealed proline enrichment in peptides with phosphor-serine and phosphor-threonine. C) sequence enrichment of phosphorylated peptides uncovered known phospho-motif. The identified phosphorylated peptides were previously identified as substrate of known kinases.

functional domains the PTM occurs on and 2) which specific enzymes are utilised by the cells to generate these modifications.

I first examined protein domains that are enriched for phosphorylated, acetylated and methylated proteins. Among all phosphorylated, acetylated and methylated proteins, 412, 171 and 409 proteins domains were significantly enriched for each modification, respectively. The most frequently modified domains were shown in Figure 16A. Notably, RNA-binding motif (RRM) domain and thioredoxin domain/thioredoxin-like fold were enriched for all three modifications. EF-hand domain pair were enriched in both phosphorylated and acetylated proteins, indicating potential co-regulation between the two modifications.

For phosphorylated peptides, I further analyzed phospho-motif enrichment and used these motif sequences to find their upstream kinases. Proteins were found to be targeted by basophilic kinases and acidophilic kinases as basophilic kinase motif (PKA and PKC) and acidophilic kinase motif (CKI and CKII) were identified in the peptide sequence analysis (Figure 16C). Those kinases regulate a large repertoire of proteins with various activities inside cells, suggesting that phosphorylation may actively regulate the function of the bone marrow.

7.2.8 Hotspot of post-translational modifications in bone marrow showed association between phosphorylation and other modifications

Protein PTMs hotspots, regions where multiple protein PTMs occur within the protein are often the regulatory centre of a protein. The frequency of protein modification in the hotspot also provides hints on the co-regulation of protein PTMs on protein function. The identified phosphorylated, methylated and acetylated peptides from the bone marrow proteome enabled me to examine the co-occurrence of these modifications inside human bone marrow.

In total, 305 protein PTM hotspots from 166 proteins were observed in the bone marrow proteome(Figure 17A). The majority of the proteins with PTM hotspot contain 1 to 2 hotspots. Only 3% of the proteins with PTM hotspot have over 5 regions that were heavily modified. The 305 protein PTM hotspots can be further divided into subgroups according to the composition and frequency of the modifications detected within the hotspot (Figure 17B). 67 hotspots were identified to have phosphorylation, acetylation and methylation, representing 22% of all identified hotspots. The combination of phosphorylation and acetylation were the most predominant co-occurring PTMs inside bone marrow. Together they regulated 39% of the identified hotspots (n=119).

As an example of protein with PTM hotspots, heterogeneous nuclear ribonucleoproteins A2/B1 (hnRNP A2B1) was shown in Figure 17C. Three PTM hotspots were identified

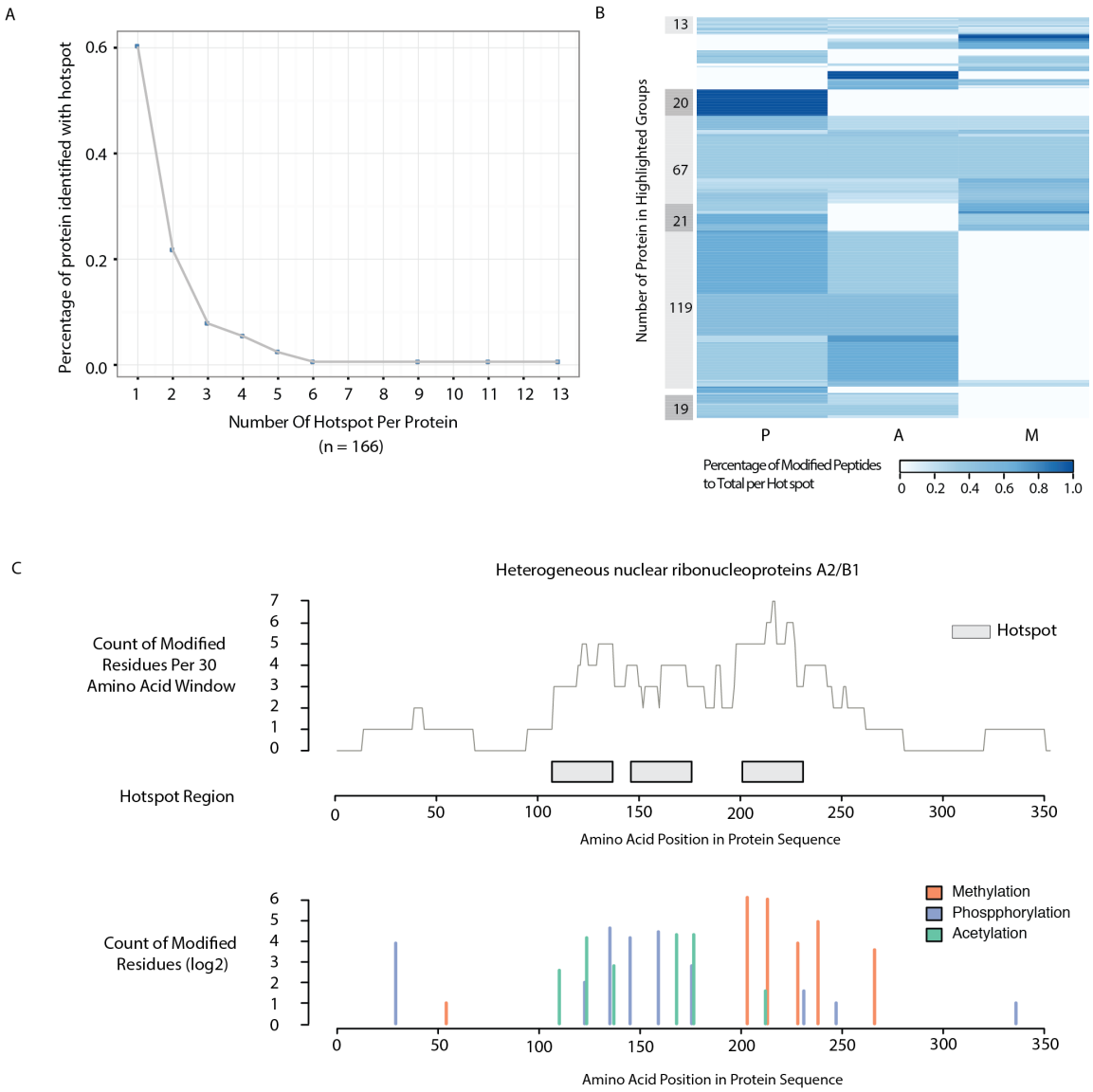


Figure 17: Co-occurrence of protein post-translational modifications inside bone marrow. A) 305 protein PTM hotspots were discovered from 166 proteins in human bone marrow. The majority of the proteins harboured singular PTM hotspot. 9 proteins (3% of the proteins identified with PTM hotspot) were found to have over 5 regions that were heavily modified. B) according to the modifications and the proportion of each modification in each hotspot, the 305 PTM hotspots were clustered and further divided into subgroups. 67 hotspots were regulated by phosphorylation, acetylation and methylation, 21 by phosphorylation and methylation, 119 by phosphorylation and acetylation. Co-regulations were observed in 19 hotspots. C) in total three PTM hotspots were identified in heterogeneous nuclear ribonucleoproteins A2/B1 (hnRNP A2B1). The three hotspots were located in close proximity, but the modifications harboured within each hotspot were distinctively different. The first two hotspots were predominantly regulated by phosphorylation and acetylation, whereas the third hotspot was regulated mostly through methylation. Abbreviations: P, phosphorylation; A, acetylation; M, methylation.

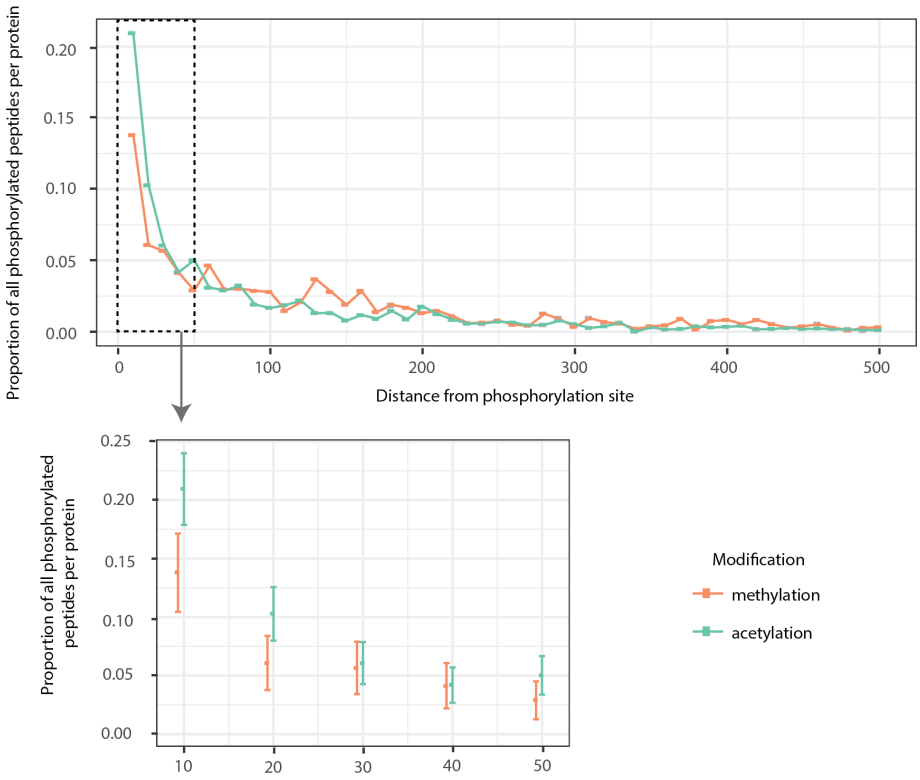


Figure 18: Cross-talks between phosphorylation and other modifications. Distances of acetylation and methylation to the nearest phosphorylation site (up to 500 amino acid distance) on the same protein were calculated. From the proportion of phosphorylation peptides containing a nearest secondary PTM sites it can be seen that majority of the co-occurring phosphorylation and other PTMs were concentrated within a distance 100 amino acids from the phosphorylation site. In addition, in the region 10 to 20 amino acids away from a phosphorylation site, acetylation was significantly more frequent in comparison to methylation.

in hnRNP A2B1. The three modification hotspots were located in close proximity, but showed distinct differences of the composition of modifications. The first two PTM hotspots were identified to contain mostly phosphorylation and acetylation sites, whereas the third hotspot appeared to be predominantly regulated by methylation.

As phosphorylation often appeared to occur together with other modifications, I further analysed the frequencies of phosphorylation in the vicinity of a second PTM. For a phosphorylated protein, I asked what proportion of phosphorylated residues are within a certain amino acid distance, given a methylation or acetylation site. Results showed that the co-occurrence of phosphorylation with other modifications were mostly concentrated within 100 amino acids away from a phosphorylation site (Figure 18). In addition, a high correlation of acetylation and phosphorylation in short distance was observed. 30% of all identified co-occurring acetylation and phosphorylation were located 10 to 20 amino acids apart from each other. Compare to other modifications, acetylation also co-occurred more frequently with phosphorylation within 20 amino acids from phosphorylation site.

7.3 Discussion

7.3.1 Bone marrow proteome is extensively modified

The data from the six cell types provided an overview of the proteome of human bone marrow microenvironment. As one of the major human tissue, the bone marrow proteome has been extensively studied using a combined approach of next generation RNA-sequencing and in situ affinity-based proteomics by human proteome atlas [Andersson et al., 2014, Uhlén et al., 2015]. This study is by far the first in-depth quantitative proteomics analysis on the human bone marrow niche. The spectrum identification rates of this study are comparable to the identification success rate of a typically LC/MS proteomic-type analysis, which is between 5 to 15% [Keller et al., 2002]. The captured proteome well-represented the content of the human bone marrow as 65% genes transcribed in the bone marrow were identified. Proteins that have been previously identified at mRNA level but not in this proteomics study may be explained by low transcription level or low translation efficiency.

The results further illustrated the scale and diversity of the modifications inside human bone marrow. With the combined strategy of using both standard and mass-tolerant database searches, I showed that the human bone marrow is modified by a variety of protein PTMs, including phosphorylation, methylation and acetylation. In addition, it is also observed that some modifications are specifically enriched in a few cell populations. For example, peptides with reduction were predominantly identified in MSCs whereas quinone modification were mostly found in ERPs. This suggests that cell-type specific modifications

exists inside bone marrow, which may fine-tune the functions of a subpopulation of the bone marrow cells.

7.3.2 Frequently modified proteins in bone marrow regulate diverse spectrum of functions

In order to further understand the functions of the common PTMs inside human bone marrow, domains enriched for phosphorylated, acetylated and methylated peptides were examined. Domains with the highest propensity to be modified across all domain detected were thioredoxin domain, RNA recognition motif (RRM) domain and EF-hand domain.

The majority of proteins containing thioredoxin domain are evolutionary related to thioredoxin [Hanschmann et al., 2013]. They act as major cellular protein disulfide reductases and are key regulators of redox-mediated signaling. Phosphorylation of thioredoxin-1 (Trx) T100 is reported to be mediated through PKC-dependent pathway [Chen et al., 2010]. The phosphorylation of Trx contributes to its nuclear location, and up-regulates the activity of NF-kappaB. In addition, phosphorylated Trx serve as an inhibitor of ASK1, a protein kinase responsive to stress stimuli and involved in cellular processes such as differentiation, apoptosis and autophagy [Yoon et al., 2009]. Other modifications on thioredoxin domain have been observed in other proteins mostly in proteomics studies, such as acetylation and nitrosylation on Trx and Trxr 1 [Rush et al., 2005, Weichsel et al., 2007, Choudhary et al., 2009]. However, the regulatory function of each modification remain to be characterized.

Similarly, protein with RRM domain have a variety of RNA-associated functions, from regulating alternative splicing to maintaining RNA stability and translation. Phosphorylation, methylation and SUMO modification of RRM have been described in previous studies and are known to regulate the conformation and stability of RRM proteins [Schullery et al., 1999, Glisovic et al., 2008].

Proteins with EF-hand domain are regulated through binding of calcium ions [Ermak and Davies, 2002, Mazumder et al., 2014]. Calmodulin and calcineurin are the best represented proteins among the EF-hand superfamily, which are known to indirectly regulate transcription factors in response to Ca2+ concentration. Modifications such as phosphorylation and methylation of EF-hand domain have been identified in mass spectrometry based studies in different human tissues, including liver cells and T lymphocytes [Raftery and Geczy, 1998, Carrascal et al., 2008, Olsen et al., 2010, Bian et al., 2014].

Taken together, domains enriched with phosphorylation, acetylation and methylation are involved in cellular processes covering from regulating gene transcription to cell differentiation and apoptosis. The extensive coverage of functions implies that PTMs may play an important part for sustaining the functions of bone marrow. In addition, the high

frequency of PTMs and potential co-regulation between PTMs on these domains further suggested that thioredoxin domain, RRM domain and EF-hand domain may serve as the centre hub of PTMs in human bone marrow.

7.3.3 Phosphorylation regulates bone marrow proteins through cross-talk with other PTMs

From an evolutionary point of view, combinations of different PTMs is an efficient strategy for cell to cope with stimuli as they provide immediate response, faster than controlling translation or transcription. Regions of amino acid sequences with high density of PTMs are often known as the PTM hotspot. PTM hotspot functions as regulatory centres and have been described in many proteins, such as histones and p53 [Bannister and Kouzarides, 2011,Brooks and Gu, 2003]. Therefore, I investigated the co-regulation of different PTMs by analyzing the PTM hotspots inside the human bone marrow.

The wide occurrence of phosphorylation in PTM hotspots highlights phosphorylation as a key PTM for bone marrow. The PTM hotspots identified in the bone marrow can be divided into two subgroups, the ones regulated by single PTM and the ones by multiple PTMs. Among the three common PTMs analysed in this study, phosphorylation dominated most of the single PTM hotspots and also participated in the majority of the hotspots of combined PTMs. Phosphorylation is known to often co-occur or compete with other PTMs, including acetylation, methylation, glycosylation and ubiquitylation [van Noort et al., 2012,Swaney et al., 2013,Zeidan and Hart, 2010,Estève et al., 2011,Fang et al., 2014, Biggar and Li, 2015]. In addition, phosphorylated proteins were found to reside in central network positions and have pairwise interactions with other proteins PTMs [Duan and Walther, 2015]. The observed high frequency of co-occurring modifications in the vicinity of phosphorylation suggested that the function of these hotspot regions can be modulated by PTM crosstalk. The observed high frequency of phosphorylation co-occurring with other PTMs is in agreement with the previous studies, and further underscore the central importance of phosphorylation in regulating human bone marrow proteome.

7.3.4 Interplay between phosphorylation and acetylation may be functionally important for bone marrow proteins

The strength of a PTM in the vicinity of a secondary PTM (based on the number of modified peptides detected) provides information on whether the two modifications could be physically interacting. Results on the distance of a phosphorylation site to its nearest modification showed that the majority of the modifications are distributed evenly across the protein sequence, whereas acetylation was more frequently located surrounding a phosphorylation site within 10 to 20 amino acid distance.

At the proteome level, acetylation has been reported to co-occur frequently with phosphorylation in genome-reduced bacterium *M.pneumoniae* [van Noort et al., 2012]. The cross-talk between phosphorylation and lysine acetylation often co-occurs within the same protein at the interaction interfaces as well as on multifunctional proteins. In addition, the conservation of phosphorylation near acetylation sites was found to be relatively high, further hinting functional co-regulation [Beltrao et al., 2012]. The pattern of co-occurrence for phosphorylation and acetylation on the same protein was also observed on human protein data sets [Duan and Walther, 2015]. Hence, the observed co-occurring phosphorylation and acetylation could be of functional importance to the human bone marrow.

At the protein level, co-regulation of phosphorylation and acetylation in short-distance has also been observed on individual proteins. For example, phosphorylation and acetylation co-regulate the intracellular location of FOXO. Acetylation of FOXO1 diminishes its binding to DNA, increasing its phosphorylation by AKT, which leads to association with 14-3-3 proteins and cytoplasmic retention [Matsuzaki et al., 2005]. In addition, co-occurring phosphorylation and acetylation in close proximity is also known to induce conformational changes necessary for the recognition of 'reader' proteins. Both phosphorylated and acetylated of CREB are required for its interaction with CBP and the formation of ternary complex with CBP KIX and BRD domains [Paz et al., 2014]. On the other hand, a negative regulation between acetylation and phosphorylation has also been reported in the case of p53 [Ou et al., 2005]. Therefore we propose that the proteins identified with phosphorylation and acetylation in close proximity may serve as good candidates for further co-regulation analysis.

7.3.5 Use mass-tolerant database search method to extend understanding on protein modifications

The combined method of standard and mass-tolerant database search addressed 17.86% of the spectra that were unassigned when only the standard database search was applied. A previous study estimated over one third of the unassigned spectra have substoichmetric modifications [Chick et al., 2015]. By this measure, I have recovered half of the potentially modified peptides. I validated most of the phosphorylated, methylated and acetylated peptides through in-silico methods. To provide additional confirmation for the identity of the modified peptides, one could compare the spectra information to those in the mass spectrometry peptide database such as PeptideAtlas.

The remaining unassigned spectra could be attributed to the following factors: combinations of modifications, peptides modified by lowly abundant PTMs that do not generate strong signals to be detected as peaks, and amino acid variants due to genetic polymorphisms. These modifications were not covered in the scope of this study, but can be

addressed with similar approaches. Of note, though computationally challenging for large-scale datasets, increasing the search windows in mass-tolerant database search and including lowly abundant PTMs in the second standard database search could potentially recover many of those modified peptides and provide more insights on the functions of those modifications.

7.4 Conclusion and perspective

The human bone marrow proteome was measured by quantitative high-resolution tandem mass spectrometry in this study. With the combination of standard and mass-tolerant database search method, protein PTMs within the human bone marrow were systematically analyzed. The human bone marrow proteome was found to be extensively modified. Common modifications such as phosphorylation, acetylation and methylation were found to be present in up to one third of the bone marrow proteins, and may modulate important cellular processes within the bone marrow. The specific patterns from phosphorylated, acetylated and methylated peptides revealed the sequence preferences for each of the modification. Analysis on the cross-talks between different PTMs further suggested phosphorylation plays an important role in regulating activities of bone marrow proteins, on its own and also in orchestration with other modifications. In addition, the short distance between phosphorylation and acetylation sites further suggested the presence of co-regulation of the two modifications is a common feature and may be functionally important to the human bone marrow.

Arguably the study gives a partial view of the entire modification landscape of the bone marrow proteome. Some modifications such as phosphorylation are unstable in-vitro and and may be easily lost during experiment. As the study aimed at unbiasedly profiling the protein modification landscape inside the human bone marrow, further PTM-oriented analysis can provide a more comprehensive understanding of the PTMs inside human bone marrow, either for a specific modification or for subpopulation of bone marrow derived cells. PTMs that induce molecular weight shift over 500Da, as well as combinations of different PTMs, which were omitted in the study could be explored via in-silico approaches. In addition, the identified PTMs as well as the modified proteins require further validation to confirm their presence of modification in-vitro and in-vivo. Additional analysis on spacial proximity in three-dimensional structure will provide insights on the functional relationships between phosphorylation and other modification inside the human bone marrow. Furthermore, for the co-regulation between phosphorylation and acetylation, functional analysis would help to understand whether the mode of co-operation is competitive or cooperative regulation.

8 Protein Phosphorylation, Acetylation and Methylation Display Differences Across Bone Marrow Cells

8.1 Summary

In this study, the proteome of human bone marrow was analysed using tandem mass spectrometry. Enriched protein PTMs were first deduced through mass-tolerant searches and re-identified using standard database searches. Following the establishment of the mass-tolerant database search method for the project and profiling the dynamics of protein PTMs of the bone marrow proteome, the differences between the protein modification status among haematopoietic progenitors (HPCs), lymphoid progenitors (LYMs), monocyte progenitors (MONs), erythroid progenitors (ERPs), granulocyte progenitors (GRAs) and mesenchymal stromal cells (MSCs) were examined. The three most frequently detected modifications (apart from oxidation and deamidation), phosphorylation, acetylation and methylation were further analyzed for their profiles across cell types. In this chapter, protein modifications that display cell-type dependency across the six cell types are presented.

8.2 Result

8.2.1 Profiles of phosphorylation, acetylation and methylation in bone marrow show cell-type dependencies

Bone marrow proteins were frequently modified. Phosphorylation, acetylation and methylation were the most frequently detected modifications and are the focus of my analysis in the following chapters. In total across six different bone marrow cell types, there were 2,324, 950 and 879 proteins identified to be phosphorylated, methylated and acetylated, respectively.

To understand the distribution of the modified proteins in different bone marrow cells, we analyzed the similarities between the occurrences of phosphorylation, methylation and acetylation on modified proteins. For the modified proteins that were detected across all six cell types, I used the Jaccard index to estimate how similar are the modifications (whether it is phosphorylated, methylated or acetylated across any of the six cell types) between those proteins. This pairwise distance measure was calculated, subjected to hierarchal clustering, and the results are plotted together with the number of modified peptides (Figure 19A). Phosphorylation was more wide-spread than acetylation and methylation and also co-occurred frequently with acetylation and methylation. In addition, there were 201 proteins detected to be solely phosphorylated in MSCs. These proteins were expressed in all six cell types, however no phosphorylated nor acetylated nor methylated forms were identified in the other five cell types. Hierarchical clustering of proteins identified to be phosphorylated, acetylated or methylated also revealed that profiles of modification in haematopoietic lineage cells were more closely related. In particular, MONs and LYMs

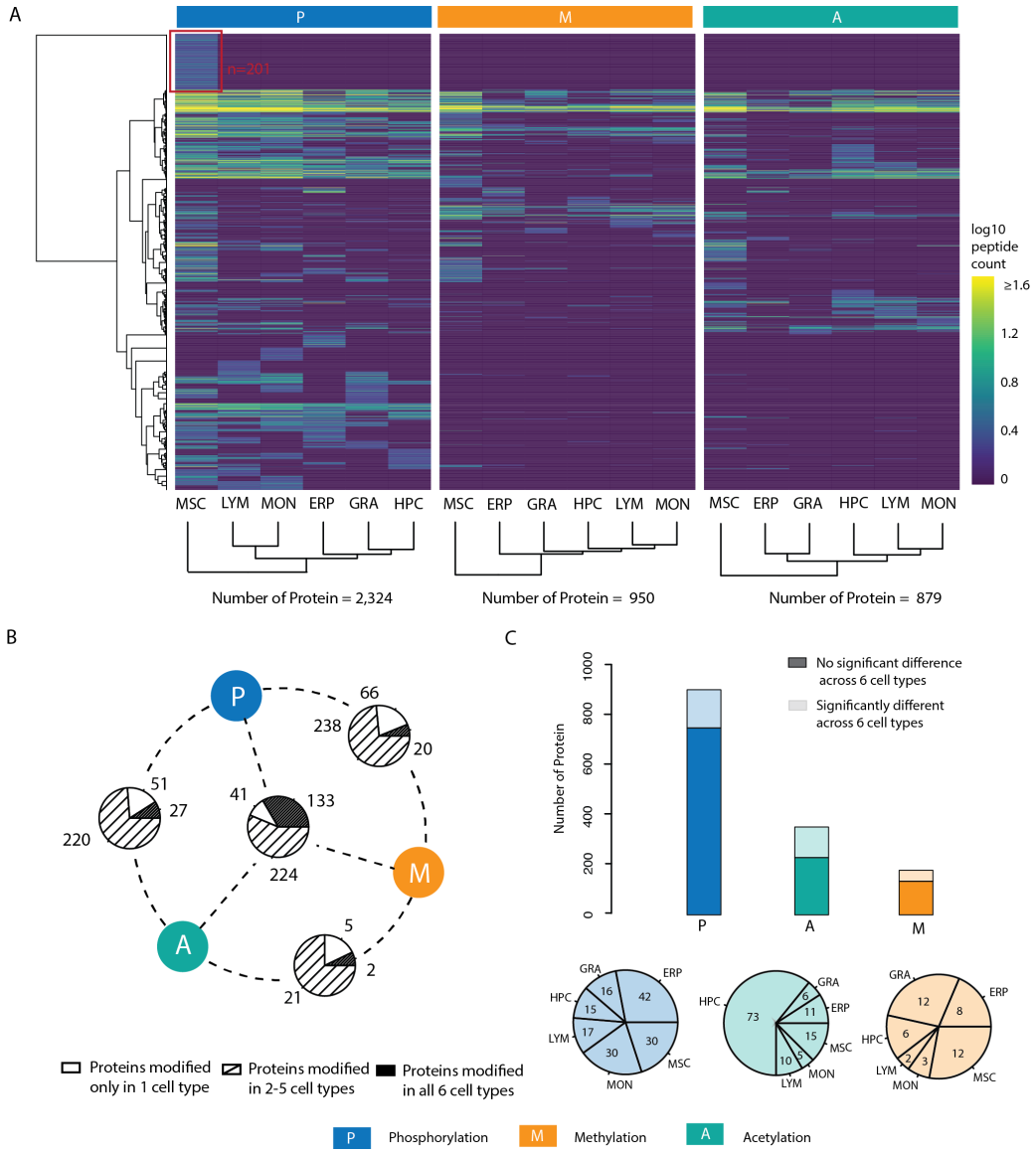


Figure 19: Distributions of phosphorylated, methylated and acetylated proteins across bone marrow cells. A) phosphorylated, methylated and acetylated proteins across 6 types of bone marrow cells were clustered based on similarity of modification profiles. Frequently modified proteins were depicted in light yellow and proteins have no modification(s) were depicted in dark blue. B) phosphorylated, methylated and acetylated proteins across 6 cell types were classified according to the cell types. C) In total 884 phosphorylated proteins, 345 acetylated proteins and 175 methylated proteins were tested for differential modification across cell types. 238, 153 and 56 proteins showed to significantly different levels of phosphorylation, acetylation and methylation among the six bone marrow derived cell types (FDR < 0.05). The cell type with the highest propensity for differentially modified proteins varied according to modification, with ERP being highest for phosphorylation, HPC for acetylation, MSC and GRA for methylation.

showed closest relationship for all three modifications. On the contrary, the modification profile of MSCs was most distant from the profiles of the five haematopoietic cell types. Overall the patterns of phosphorylation, methylation and acetylation were distinct across six cell types.

Different PTMs often modify on the same protein to co-regulate its function, therefore we further analyzed modified proteins for PTM co-occurrence. In total 1,048 proteins were identified with combinations of phosphorylation, acetylation and methylation(Figure 19B). The proportion of proteins identified with all three modification co-occurring and in all six cell types was 28% (n=133). This proportion was much higher than the percentage of proteins identified with other combinations of co-occurring modifications in all six cell types, which were in between 6 to 9%. The 133 proteins with co-occurring phosphorylation, acetylation and methylation were enriched for signalling pathways including platelet degranulation and activation, glycolysis, RHO GTPase activity and mRNA splicing (Supplementary Figure 4).

8.2.2 Phosphorylated, methylated and acetylated proteins were differentially regulated across bone marrow cells

Phosphorylated, acetylated and methylated proteins were further analyzed for differential modification rates across cell types. For each modified protein, the number of PSM detected from its modified peptides was on average 1 to 2, whereas that for the unmodified peptide is between 10 to 20. At this low number, the estimate of modification frequency (modified peptide counts divided by unmodified peptide counts) tends to be imprecise, making it difficult to identify significant variation that are above the noise. Therefore, to overcome this problem I used all modified peptides identified for each protein to collectively estimate the modification frequency at the protein level.

To compare the modification frequency across cell types for a given protein, the proportion of its modified peptides (with respect to the unmodified peptides) was compared across all 6 cell types using an extension of Fisher's exact test. Only proteins with at least 10 unmodified peptides and 3 modified peptides for each modification were analyzed. With an FDR < 0.05, in total 27 to 44% of the proteins showed significantly different levels of phosphorylation, acetylation and methylation across the six bone marrow cell types (Figure 19C). These differential regulations were often driven by very strong modification signals in one (or two) cell types(s) compared to the others. For phosphorylation, ERPs had the highest modification propensity, followed by MONs and MSCs. For methylation, GRAs and ERPs had the most differentially regulated proteins, whereas the majority of the differentially acetylated proteins (61%) can be attributed to highly acetylated proteins in HPCs.

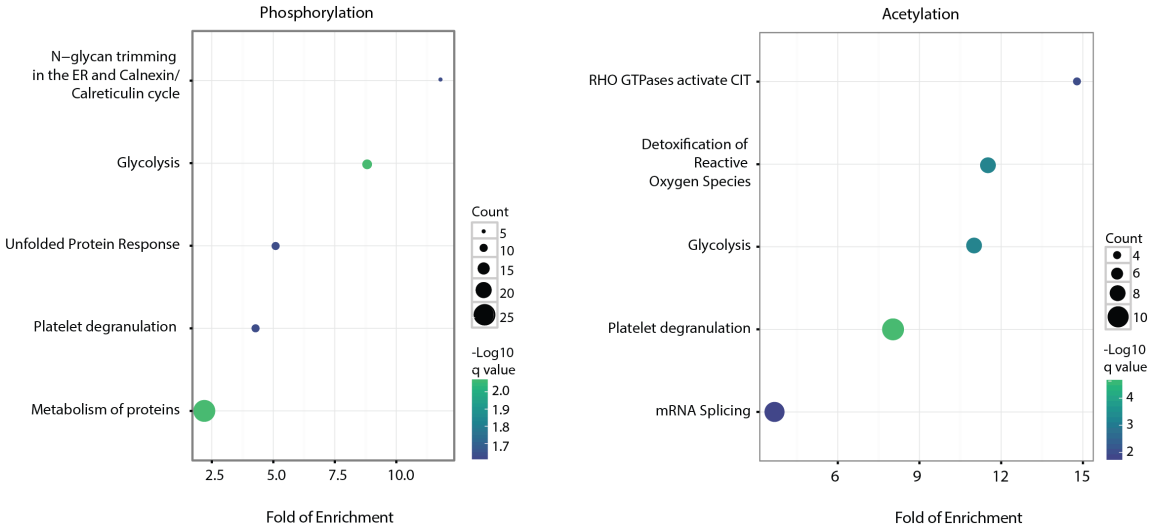


Figure 20: Mapping differentially phosphorylated and acetylated proteins with signalling pathway revealed cell-type specific regulations. Proteins differentially phosphorylated (n=238) and methylated (n=153) were mapped to Reactome database. Pathways that were significantly enriched for each modification were shown (FDR < 0.05). The size of the dots indicated the number of proteins that fall into each pathway, and colour of the dots represented the statistical significance of the enrichment. Pathways at the same level but under the same parental pathways were collapsed and only the ones with lowest p values were shown here.

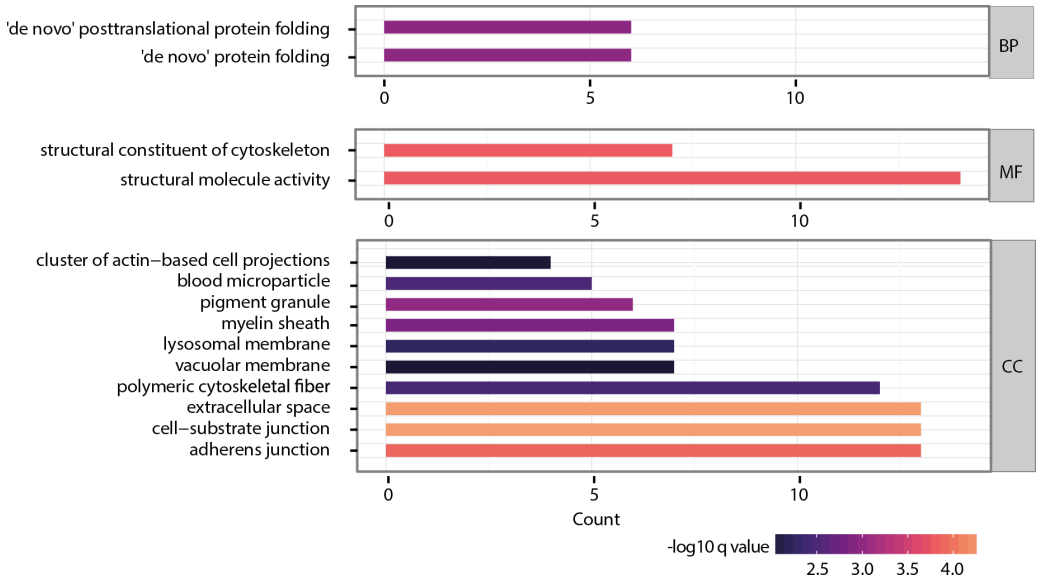


Figure 21: Over-represented biological processes, molecular functions and cellular components of differentially methylated proteins. 56 differentially methylated proteins were tested for over-representation in Gene Ontology association for biological process(BP), molecular function(MF) and cellular components (CC) (FDR < 0.05). Similar GO terms at the same level but under the same term were collapsed and only the ones with highest count were shown here.

In order to reveal the functions of those differentially modified proteins, I tested for significantly enriched pathways for the differentially phosphorylated and acetylated proteins (Figure 20). The enriched pathways indicated that the differential PTM is not restricted to a few proteins, but rather the varying regulation of phosphorylation and acetylation is spread over different junctures of pathways. It is reasonable to consider that these pathways may be relevant for the function of individual cell types.

Proteins differentially methylated were not enriched in any particular pathways. Gene ontology analysis showed that these proteins were involved in 'de novo protein folding' and 'de novo post-translational protein folding' (Figure 21). The majority of the differentially methylated proteins were enriched for structural components of cells, mostly located on the cell membrane, extracellular components and cell junction.

8.2.3 Site-specific regulation on phosphorylated, methylated and acetylated bone marrow proteins

For the same modification, proteins can also display significant differences in terms of the modification sites while the overall modification frequencies remain unchanged across cell types. Therefore, I investigated the heterogeneity of modification sites for phosphorylated, acetylated and methylated proteins across six cell types. To overcome the problem with low number of PSM per modified peptide as mentioned, a Chi-square based test was applied to detect significant variations in modification frequency for each site in a given protein. In total 87 proteins (FDR < 0.05) showed significant variability for modification sites, of these 70 were identified with phosphorylation differences, 5 for methylation and 36 for acetylation (Figure 22A).

In addition to differences on single PTM, 25 proteins showed positional differences for more than one modification(Figure 22A). Figure 22B showed EEF1A1 as an example to illustrate the bias at each modification site. T261p was the most predominant phosphorylation site, however the secondary phosphorylation site of EEF1A1 varied across cell types. Acetylation of EEF1A1 displayed heterogeneous patterns as the dominant modification sites varied across cell types. The most frequently acetylated sites were C411 in MSCs, K255 in LYM, K439 in HPC and K386 in ERP. For MON and GRA the acetylation sites were identified to be evenly distributed across the entire protein.

8.2.4 Heteronuclear binding proteins are differentially modified RNA-binding proteins in human bone marrow

To understand functions of the differentially modified proteins, I further analyzed the composition of proteins in the common pathways enriched for differentially phosphorylation and acetylation (Figure 20). Interestingly, proteins enriched in the mRNA splicing pathway

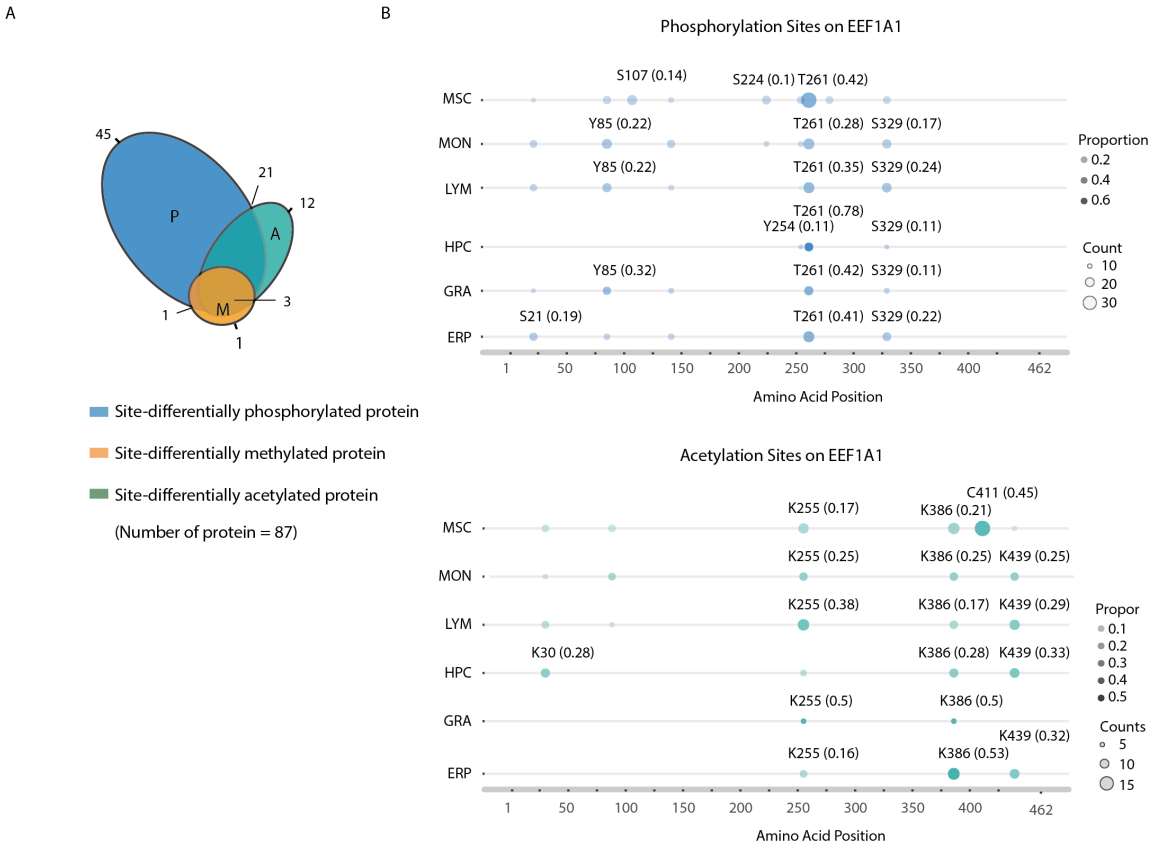


Figure 22: Preferences of modification site varied across cell types in human bone marrow. A) in total 70 proteins in the human bone marrow showed significant variations for phosphorylation sites, 5 proteins for methylation sites and 36 proteins for acetylation sites. B) dot plots depicted ratio of spectral counts of the phosphorylated and acetylated peptides to the unmodified peptides observed across six cell types for sites along the length of EEF1A1. T261p was found as the dominant phosphorylation site in all cell types whereas secondary phosphorylation sites showed great variability across cell types. Acetylation of EEF1A1 displayed heterogeneous patterns as the dominant modification sites varied across cell types. The most frequently acetylated sites were C411 in MSCs, K255 in LYM, K439 in HPC and K386 in ERP. For MON and GRA the modifications were identified to be evenly distributed across the entire protein.

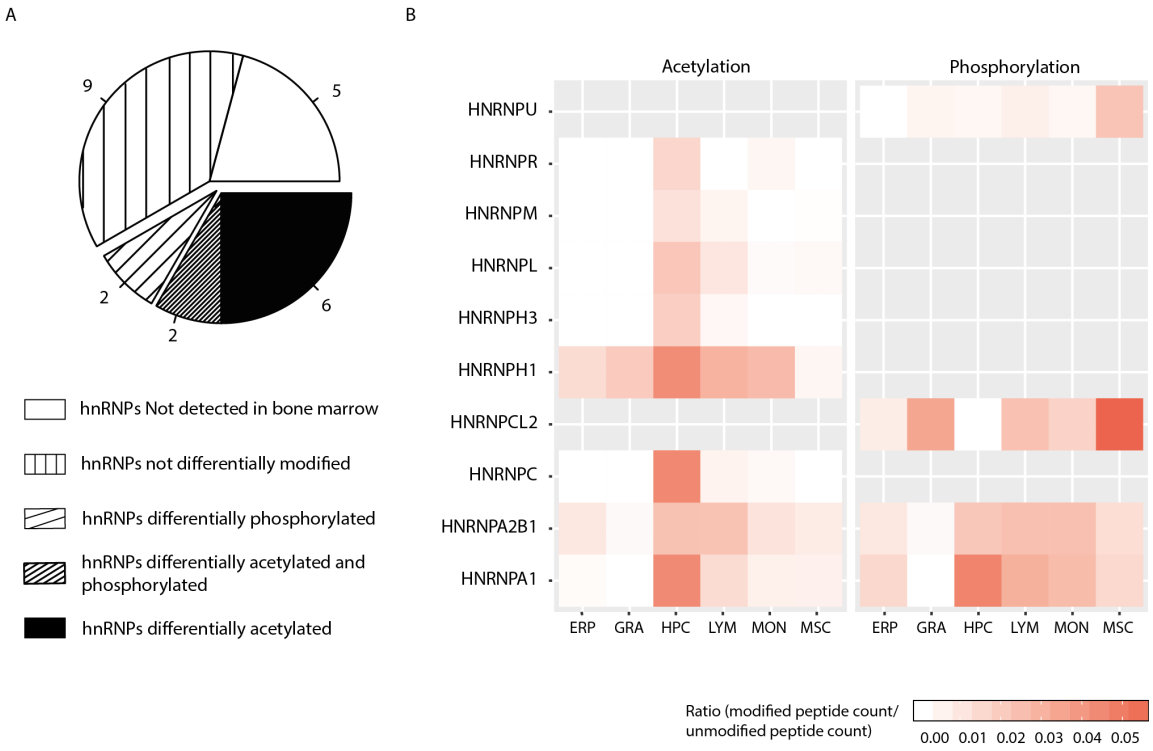


Figure 23: Heteronuclear binding proteins were preferentially acetylated in haematopoietic progenitor cells. A) over half of the hnRNP proteins identified in the human bone marrow proteome were differentially modified. Acetylation regulate the 8 of the 10 differentially modified hnRNP proteins. B) percentage of modified peptide counts to the unmodified peptides counts for each differentially modified hnRNP in six cell types were shown to illustrate the modification preference across six cell types. Acetylation regulated and co-regulated the majority of the differentially modified hnRNP proteins. Proteins that did not have the corresponding modified form detected were shown in white.

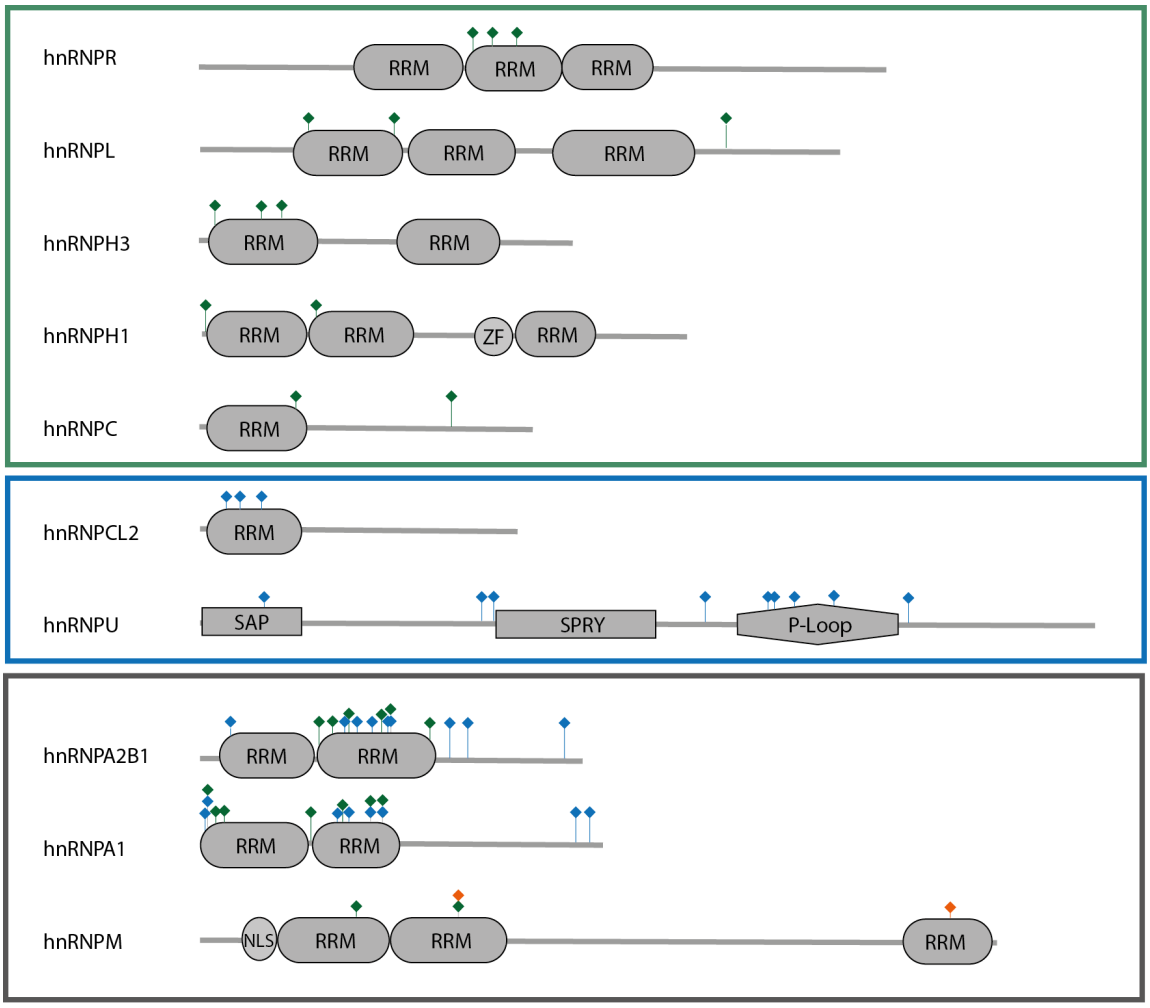


Figure 24: Differentially regulated phosphorylation, acetylation and methylation sites were concentrated in RNA-recognition motif domains on heteronuclear binding proteins. The differentially phosphorylated, methylated and acetylated hnRNP proteins and the modification sites found for each hnRNP were shown. Together with other two modification, acetylation co-regulated the majority of the differentially modified proteins. For proteins with RNA-recognition motif (RRM) domain, majority of the acetylation sites were located within the domain. For hnRNP U, which does not contain RRM domain, the modifications were mostly on its P-loop domain. In addition, competitions between acetylation and phosphorylation, acetylation and methylation were also observed within the RRM domains of hnRNP A2B1, hnRNP A1 and hnRNP M.

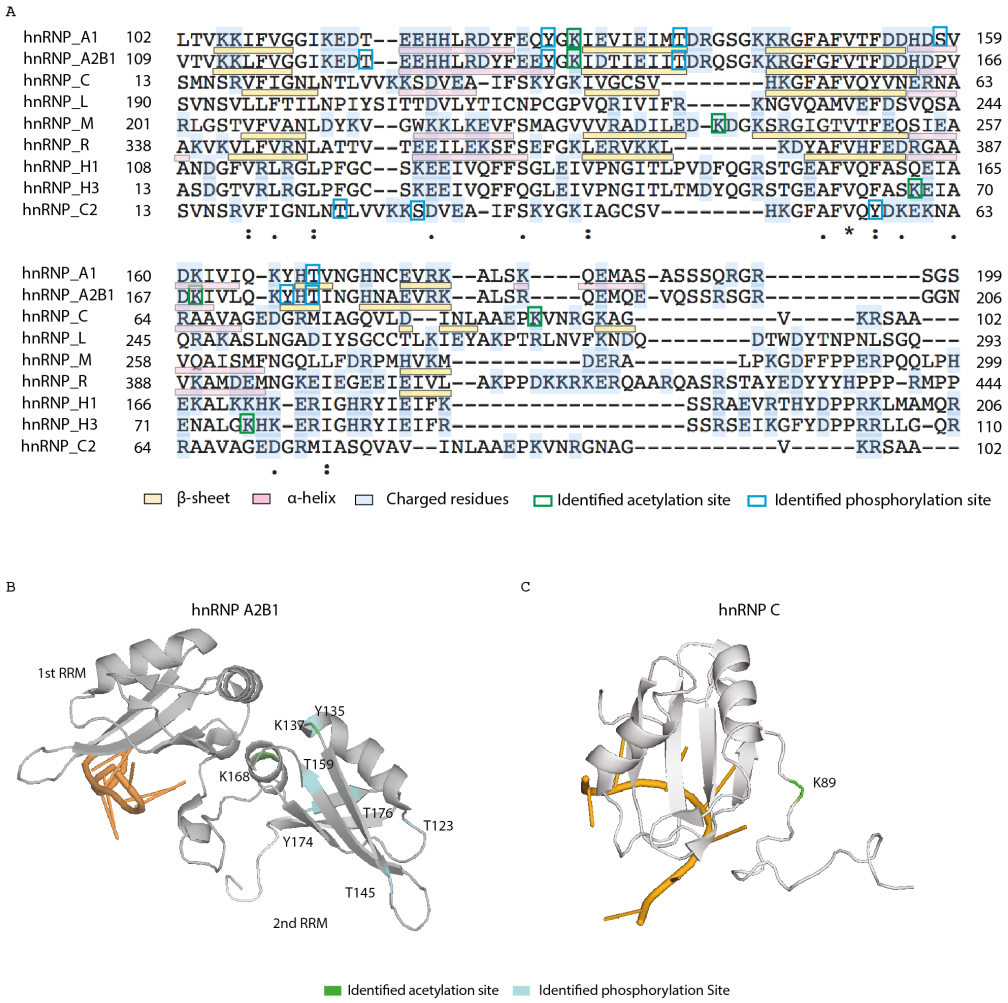


Figure 25: Modification sites on differentially regulated heteronuclear binding proteins are are in selected regions on RRM. A) sequence alignment of differentially modified hnRNP proteins with RRM domains revealed that the modification sites were mostly located in regions in between β-sheets, which form RNA-binding surface on RRM domain. The identified modification sites shared similar chemical properties across subset of hnRNP proteins. Modifications identified within the selected regions were highlighted in colour according to the modification, which were: Y128p, K130ac, T138p, S158p and T169p for hnRNP A1; T123,Y135p, K137ac, T145p, K168ac, Y174p and T176p for hnRNP A2B1; K67ac and K76ac for hnRNP H3; K239p and K239ac for hnRNP M; K89ac for hnRNP C; T25p, S38p and Y67p for hnRNP CL2. B) X-ray crystal structure of hnRNP A2B1(residue 15-193) in complex with 10-mer RNA (PDB: 5HO4). Acetylation sites K137 and K168 were highlighted in green. Phosphorylation sites T123 , Y135, T145, Y174 and T176 were highlighted in cyan. C) solution structure of hnRNP C (residue 2-106) in complex with 5-mer RNA (PDB: 2mxy). The identified acetylation site K89 was highlighted in green.

were all members from the heteronuclear binding protein (hnRNP) family. This further prompted me to examine this group of proteins in detail.

In total, there are 24 human hnRNP proteins annotated in UniprotKB database, 19 of which were detected in the human bone marrow (Figure 23A). Among the 19 hnRNP proteins detected, 10 of them were found to be differentially phosphorylated or acetylated. The differentially acetylated hnRNP proteins constituted the majority of the differentially modified ones, 42% of all hnRNP proteins detected. The 5 hnRNP proteins not detected in this proteomics dataset could be explained by their low mRNA level in the human bone marrow, as shown in the study by Uhlen et al. [Uhlén et al., 2015].

The 10 differentially modified hnRNP proteins displayed distinct patterns across cell types (Figure 23B). Acetylation of hnRNP proteins were predominantly observed in HPCs. Furthermore, hnRNP A2B1 and hnRNP A1 also contained modification sites that showed positional differences across different cell types (Supplementary Figure 8).

I verified the modification sites on hnRNP proteins through cross-checking with modification database PhosphoSitePlus. The majority of the modified residues detected on hnRNP proteins have been identified in previous studies. Phosphorylation sites, such as Y135, T145, Y174 and T176 on hnRNP A2B1, Y128, T138, S158 and T169 on hnRNP A1, have been identified in other proteomics studies [Tsai et al., 2015, Pinto et al., 2015, Levin et al., 2016]. Acetylation site such as K168 on hnRNP A2B1 has also been found in previous studies [Wu et al., 2015]. Interestingly, the majority of the detected acetylation sites, such as K67 and K76 on hnRNP H3, K137 on hnRNP A2B1, K89 on hnRNP C, were identified ubiquitination sites [Kim et al., 2011, Mertins et al., 2013]. The detected methylation site K692 on hnRNP M has also been identified as a ubiquitination site [Kim et al., 2011].

To understand the effect of modifications, I further examined the locations of those modification on hnRNP proteins. The modification sites identified for all differentially modified hnRNPs were predominantly located within the RRM domains of the hnRNP proteins (Figure 24). Phosphorylation of hnRNP U, which atypically does not contain a RRM domain, was on its SAP domain, p-loop domain as well as on regions in between the three domains. For the hnRNP proteins differentially modified by combination of PTMs, the detected acetylation sites were often located in close proximity with phosphorylation or methylation within the RRM domains.

Why is there a preference for modification on RRM domains? The RRM domains of hnRNP proteins are highly conserved across species (Supplementary Figure 5 - 7). Sequence alignment of the differentially modified hnRNP proteins with RRM domain showed that the identified modification sites were mostly clustered in selected regions within the RRM

Table 2: Resolved structures of the differentially modified heteronuclear binding proteins

Name	PDB	Type	Residue	Domain
hnRNP R	2dk2	NMR	333-416	3rd RRM
hnRNP L	3to8	X-ray diffraction 1.82Å	380-589	C-terminal RRM
hnRNP H1	2lxu	NMR	7-111	N-terminal RRM
hnRNP C	2mxy, 2mz1	NMR	2-106	N-terminal RRM with 5-mer RNA
hnRNP U	1zrj	NMR	1-37	SAP
hnRNP A2B1	5ho4	X-ray diffraction 1.85Å	15-193	N-terminal RRM with 10-mer RNA
hnRNP A1	1l3k	X-ray diffraction 1.1Å	1-196	two N-terminal RRM
hnRNP M	2do0, 2dgv	NMR	196-296, 652-730	2nd and 3rd RRM

domain. Aligned regions where modifications were clustered revealed that modification sites appeared to share similar properties across subsets of hnRNP proteins in human (Figure 25A). For hnRNP A2B1 and hnRNP A1, multiple residues aligned on the same positions were identified to have the same modification, such as acetylation site T110ac on hnRNP A2B1 and T102ac on hnRNP A1, as well as co-occurring pattern Y135p/ K137ac on hnRNP A2B1 and Y128p/K130ac on hnRNP A1. Taken together, these suggest that RRM domains have evolutionary conserved features and PTMs on this regions could be an effective mechanism for regulating the function of hnRNP proteins.

To understand the potential regulatory mechanisms of these modifications on RRM, the identified modification sites were further mapped to the crystal structures of the hnRNP proteins. Among the 10 differentially regulated hnRNP proteins, 8 proteins have their partial structures resolved (Table 2). Among those, hnRNP A2B1 and hnRNP C have structures in complex with RNA and are illustrated as examples (Figure 25 C-D). hnRNP A2B1 (residue 15-193, first two RRM domain) was crystallised in complex with a 10-mer RNA (PDB: 5HO4), whereas the structure of hnRNP C (residue 2-106) was resolved in complex with a 5-mer RNA by NMR spectroscopy (PDB: 2mxy). The identified modification residues on hnRNP A2B1 and hnRNP C (highlighted in colours) were all located in regions away from the RNA binding groove, mostly in between the β-sheets as well as sporadically on the α-helices. Crystal structures of other hnRNP proteins were also in agreement with the observation that the majority of the identified modification sites were away from the RNA binding surfaces (Supplementary Figure 9). These modifications sites were unlikely to directly interfere but could play an auxiliary role in RNA binding.

8.2.5 Glucose metabolism is differentially regulated through phosphorylation and acetylation across different cell types in bone marrow

Glycolysis pathway was enriched for both differentially phosphorylated and acetylated proteins, highlighting its importance as a central energy metabolism pathway across different bone marrow cells (Figure 20). I further examined all proteins associated with glycolytic pathways for their modification frequencies in order to have a comprehensive view

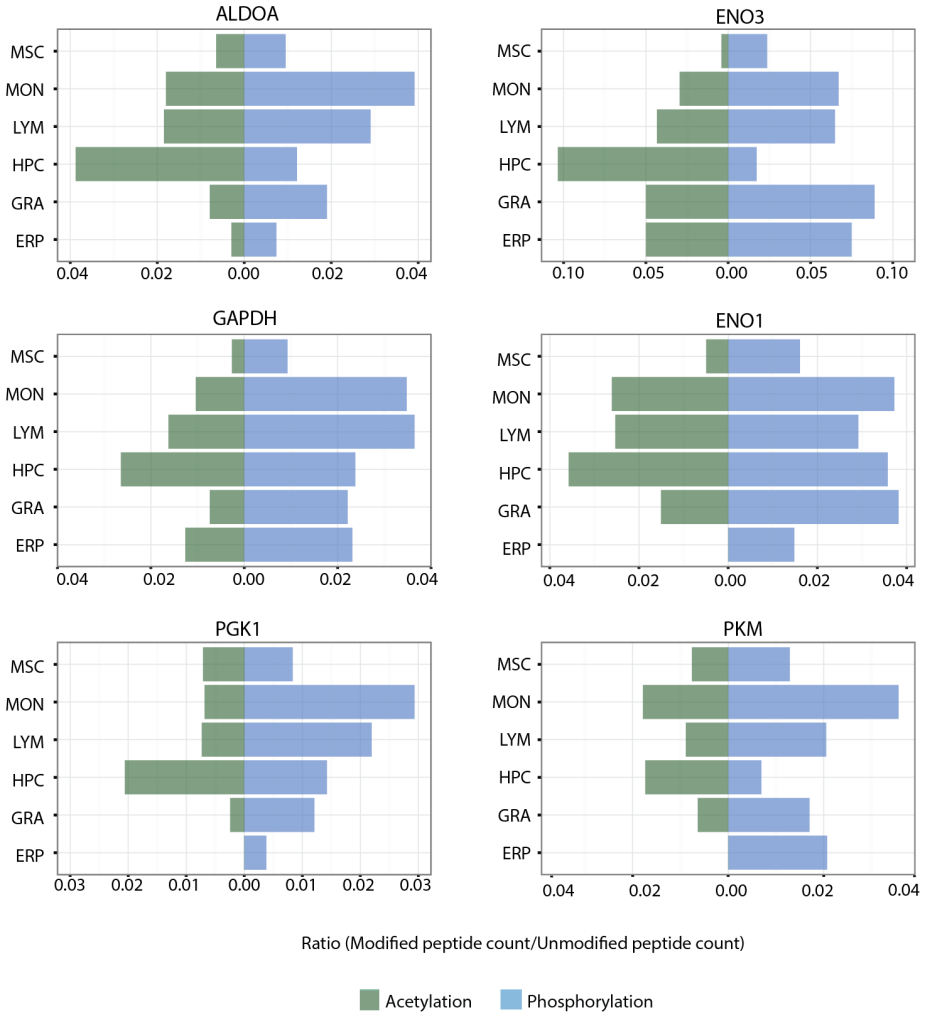


Figure 26: Differentially modified proteins in glycolysis pathways displayed preferences for acetylation and phosphorylation across cell types. α-enolase (ENO1), β-enolase (ENO3), phosphoglycerate kinase 1(PGK1), fructose-bisphosphate aldolase A (ALDOA), pyruvate kinase (PKM) and glyceraldehyde-3-phosphate dehydrogenase (GAPDH) showed significant variability for frequencies of both phosphorylation and acetylation in bone marrow cells. For a given protein, the modification frequency was shown by the ratio of modified peptide counts versus its unmodified peptide counts, with acetylation coloured in green and phosphorylation in blue. HPCs displayed a preference towards acetylation for all proteins. The frequencies of acetylation of those proteins in HPCs were also the highest among all six cell types. For the other 5 cell types, modification of the 6 differentially modified proteins showed preferences towards phosphorylation.

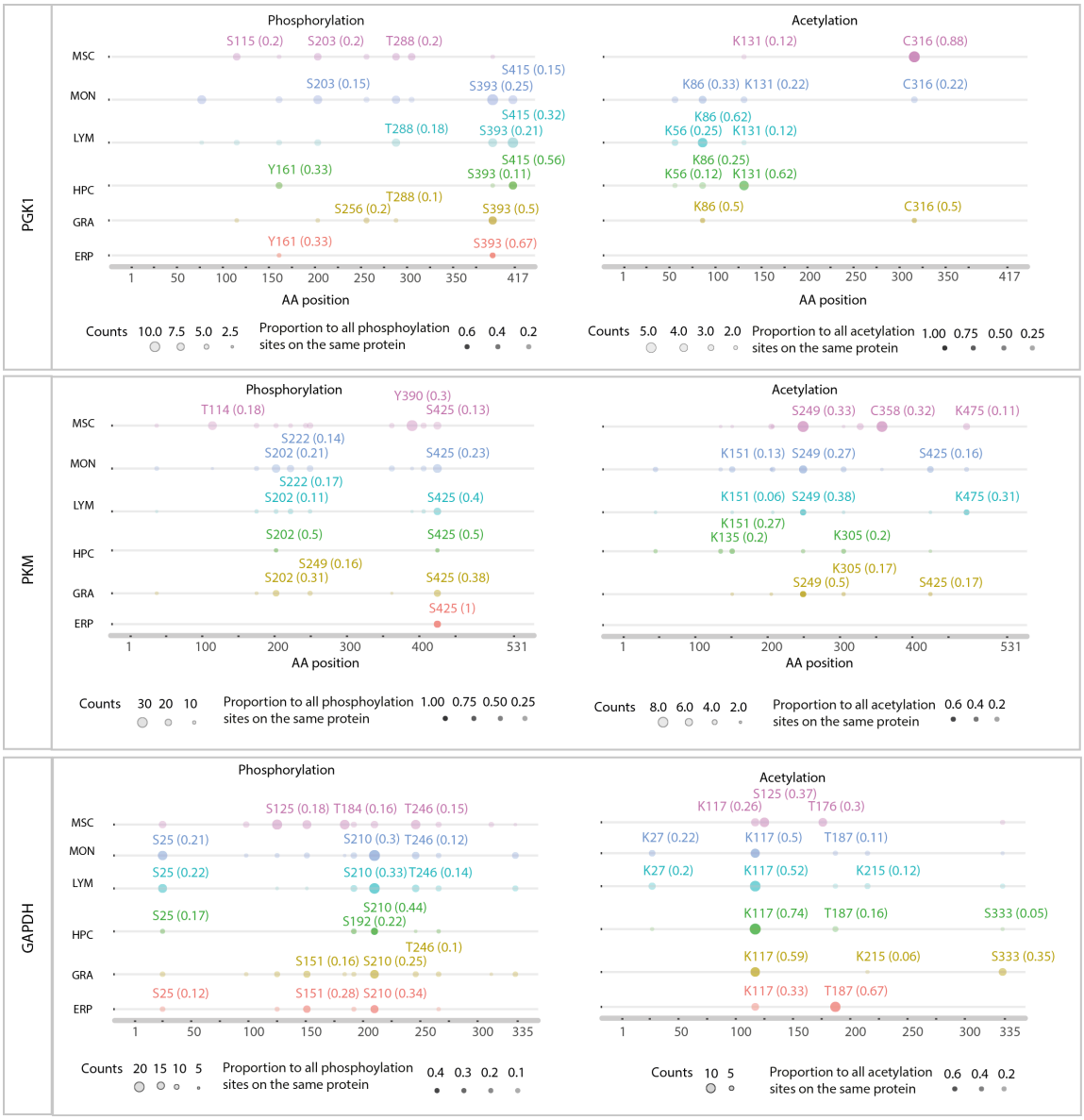


Figure 27: Phosphorylation and acetylation on glycolytic proteins displayed positional differences across cell types. Three proteins, namely PGK1, PKM, GAPDH, displayed cell-type dependent preferences for both phosphorylation and acetylation sites. For the three proteins, MSCs displayed unique modification pattern different from ones from cells from haematopoietic lineage. Dominant modification sites, such as Y390p and C358ac in PKM, S125p, T184p, S125ac, T176ac in GAPDH were only detected in high frequencies in MSCs. For modification sites identified in haematopoietic lineage cells, the phosphorylation and acetylation sites also showed high variability across cell types.

of the entire pathway. The results showed that the majority of the modified proteins in the glycolytic pathway have similar modification frequencies across cell types (Supplementary Figure 11). The differentially phosphorylated and acetylated proteins were involved only in the fructose metabolism pathway.

In total 7 proteins annotated in the glycolytic pathways showed variability for their phosphorylation levels, including α-enolase (ENO1), β-enolase (ENO3), phosphoglycerate kinase 1(PGK1), fructose-bisphosphate aldolase A (ALDOA), pyruvate kinase (PKM), glyceraldehyde-3-phosphate dehydrogenase (GAPDH) and phosphoglucomutase-1 (PGM1). Except PGM1, all differentially phosphorylated glycolytic proteins were also differentially acetylated.

In order to find the potential PTM co-regulation in the glycolytic pathway, I further compared the frequencies of phosphorylation and acetylation for the 6 proteins both differentially phosphorylated and acetylated (Figure 26). HPCs displayed preferred acetylation for all 6 proteins. The frequencies of acetylation of those proteins in HPCs were also the highest among all six cell types, whereas the frequency of phosphorylation for those proteins in HPCs were lower. For the other 5 cell types, the 6 differentially modified proteins showed tendencies to be phosphorylated as opposed to be acetylated.

The majority of the phosphorylation and acetylation sites on those proteins have been detected in previous studies, but none of them directly affect the enzymatic functions of those proteins. Most of the frequently detected acetyl-lysines and phosphoryation sites on PGK1, PKM2 and GAPDH have all been detected in previous proteomics studies [Mertins et al., 2013, Kim et al., 2011, Kettenbach et al., 2011, Tsai et al., 2015]. A few of the modified residues have been validated in biochemical studies and these include S203p on PGK1, K305ac, S202p and Y390p on PKM2, K117ac and T246p on GAPDH [Lv et al., 2011, Ventura et al., 2010, Li et al., 2016, Park et al., 2016b, Yogalingam et al., 2013, Qvit et al., 2016, Hitosugi et al., 2009].

To further understand the regulation on glycolytic pathway through PTMs, I analyzed the modification sites identified for each differentially phosphorylated and acetylated proteins. Among the three proteins that have both phosphorylation and acetylation sites differentially regulated, MSCs displayed unique modification pattern different from ones from the haematopoietic lineage cells (Figure 27). Dominant modification sites, such as Y390p in PKM, S125p and T184p in GAPDH were only detected and identified with high frequencies in MSCs. For modification sites identified from cells of haematopoietic origin, the phosphorylation and acetylation sites also showed cell-type preferences. For example, S202 phosphorylation on PKM2 was consistently observed in cells from haematopoietic

lineage but proportionally higher in the HPC population. Similarly, K117 acetylation on GAPDH was also observed in all cell types but the percentage is highest in HPCs.

In conclusion, proteins in the glycolytic pathway show different tendencies for PTMs and we can localize these differences to a few sites that can have potential functional implications.

8.3 Discussion

8.3.1 Profiles of phosphorylation, acetylation and methylation of the bone marrow-derived cells reflect functions of each cell types

Profiles of protein phosphorylation, methylation and acetylation were heterogeneous across all cell types. Comparison of the profiles across cell types revealed that the major difference lay along the line of haematopoietic versus non-haematopoietic origin. The PTM profiles of the haematopoietic lineage cells did not show a clear separation between the myeloid (GRA, MON, ERP) and the lymphoid (LYM) populations. Some of the differences might be attenuated by the heterogeneity of the cell populations analyzed in this study. All cell populations from the haematopoietic lineage were isolated through positive selection of a cocktail of surface markers, and will include cells at various stages of development.

One hallmark feature I observed is the preferential acetylation in the HPC population. Previous studies have revealed that metabolic profiles of pluripotent stem cells were different from their terminally differentiated progenitors [Yanes et al., 2010, Panopoulos et al., 2012]. The metabolism of pluripotent stem cells shifts from oxidative phosphorylation to aerobic glycolysis, feature known as Warburg effect characteristic of cancer cells [Panopoulos et al., 2012,Folmes et al., 2011]. Acetate is accumulated during reprogramming of mouse induced pluripotent stem cells, and is shown to delay early differentiation in both human and mouse embryonic stem cells [Folmes et al., 2011]. In addition, metabolic and transcriptional analysis in embryonic stem cells showed that pluripotent cells produce acetyl-CoA, the source material of acetylation, through glycolysis but the function is rapidly lost upon differentiation [Moussaieff et al., 2015]. Together these evidences suggested that the shift of metabolism to glycolysis in stem cells may contribute to the observed elevated acetylation level in HPCs.

Despite both being multiple-potent stem cell populations within the human bone marrow, MSCs do not show preferential acetylation of proteins. In fact, its modification patterns were distinct from HPCs. The bone marrow HPCs is among the best studied adult stem cell population in human and is defined by their potentials to regenerate and produce all blood and immune cells throughout entire lifespan, whereas the bone marrow-derived

non-haematopoietic MSCs are characterized as being able to replicate as undifferentiated cells and also have the tri-lineage differentiation capacity [Morrison et al., 1995, Mackay et al., 1999]. It has been discussed that some of the inherent stem cell properties of the HPCs may not be transferable to other types of adult stem cells including MSCs [Lindner et al., 2010]. The differences in the properties of HPCs and MSCs as well as their functional differences most likely explain the large differences observed for each modifications in this dataset.

8.3.2 Acetylation of hnRNP proteins regulates alternative splicing in haematopoietic progenitor cells

A large fraction of proteins from the hnRNP family was discovered to be differentially modified in the human bone marrow. The hnRNP proteins are among the most abundant nuclear proteins ubiquitous expressed at different levels in most tissues [Chaudhury et al., 2010]. They form complexes with heterogeneous nuclear RNA (hnRNAs), and are constantly being added or removed from the transcript during RNA splicing. They also facilitate the export of mRNA to cytoplasm and translation machinery. Acetylation of hnRNP proteins has been reported in hnRNP A1, and more recently in hnRNP F, hnRNP C1/C2 and hnRNP L [Kim et al., 2006,Koumbadinga et al., 2015,Kim et al., 2015]. Deacetylation of hnRNP C1/C2 and hnRNP L by SIRT1 has been demonstrated both in-vitro and in-vivo through mass spectrometry and immunoblotting [Kim et al., 2015]. Administration of deacetylase inhibitor trichostatin A (TSA) in hnRNP F resulted in changes of splicing, further suggesting the function of acetylation in regulating the alternative splicing of hnRNP proteins [Koumbadinga et al., 2015]. Given the extensive acetylation events observed on hnRNP proteins in primitive HPCs, it is tempting to postulate that genes targeted by these acetylated hnRNP proteins are transcribed differently upon haematopoiesis. In fact, target genes of the hnRNP A1 and hnRNP A2B1 showed increased exon skipping (on average 20%) in the differentiated CD4+ T cells compared to undifferentiated HSC (analyzed dataset from [Casero et al., 2015, Huelga et al., 2012], Supplementary Figure 10). The increased exon skipping events hinted that regulation of RNA splicing by specific hnRNP proteins during differentiation might be linked to its acetylation profile.

Besides regulating splicing activity through acetylation, modifications can also indirectly modulate the function of hnRNPs. Modifications such as phosphorylation, sumoylation, ubiquitination, and methylation are all known to affect the stability and cellular localization of the hnRNP proteins [Glisovic et al., 2008, Chaudhury et al., 2010, Gao G, 2017]. Acetylation of hnRNP F on the mutually exclusive acetylation and ubiquitination sites K87, K98 and K224 are shown to be critical for the enhanced protein stability and alternative splicing induced by the deacetylase inhibitor TSA [Koumbadinga et al., 2015]. Degradation of hnRNP A1, hnRNP I and hnRNP L can also be prevented by TSA, further suggesting a conserved mechanism for regulating the stability and function of hnRNP proteins through

protein PTMs. As the results showed that a considerable amount of acetylation sites and some methylation sites identified on hnRNP proteins were previously identified as ubiquitination and SUMOylation sites, it is plausible that during haematopoiesis the dynamics of protein abundance within the repertoire of hnRNP proteins are also regulated by different PTM codes.

As the most common among the three RNA-binding motifs in hnRNP proteins, RRM domain also showed differential modification in the human bone marrow. The typical RRM contains four anti-parallel β-strands and two α-helices, arranged in a $\beta\alpha\beta\beta\alpha\beta$ fold [Dreyfuss et al., 1988]. The β-sheet surface of the RRM constitute an exposed surface for direct interaction with RNA [Görlach et al., 1992]. Specificity of RNA binding of the hnRNP proteins is determined by loop-regions connecting the β-strands as well as the terminal regions of the RRM [Görlach et al., 1992]. A single RRM typically can interact with 2-6 nucleotides, which is sufficient for RNA binding. Multiple copies of RRM, as in the case of several hnRNPs, can provide enhanced specificity and affinity for RNA binding, thus allowing recognition of larger and complex RNA sequences [Maris et al., 2005]. Since the identified modification residues in bone marrow were mostly mapped outside the β-sheets in the loop-regions, these modifications are unlikely to interfere with the RNA binding. However the specificity of the hnRNP proteins to their target RNA as well as the stability and function of the hnRNP proteins may be altered upon modification. Hence, additional in-vitro and in-vivo functional study on these identified modification sites may provide further insights on the key residues and modifications that govern the activity of hnRNP proteins in the human bone marrow.

Given that most of the modifications were identified in the conserved RRM domain present in the most of the hnRNP proteins, it is very likely that the single PTM as well as co-occurred PTMs observed in hnRNP A2B1, hnRNP A1 and hnRNP M may be conserved across the entire family. One could suggest that there are potentially more PTMs on these sites that are not identified in this study and could be essential for the function of hnRNP proteins in bone marrow cells.

8.3.3 Glycolytic proteins are differentially directed for sub-cellular localization through phosphorylation and acetylation

Enzymes in the glycolysis pathways were differently phosphorylated across cell types in human bone marrow. The most frequently detected modification sites do not directly impact on the catalytic activity of those enzymes, but in-vitro and in-vivo experiments indicated that these PTMs may direct the cellular location of these proteins. S202 phosphorylation on PKM2 was consistently observed in cells from haematopoietic lineage. AKT-directed S202p is essential for the nuclear translocation of PKM2, which is required for STAT5A activation and IGF-induced cell growth [Park et al., 2016b]. Constitutive

activation of STAT5A has been shown to promote human haematopoietic stem cell self-renewal, which is in agreement with the observed high frequency of S202p in HPCs population [Schuringa et al., 2004]. Phosphorylation of S203 on PGK1 by ERK is required for its mitochondrial translocation, which inhibit the mitochondrial pyruvate metabolism and enhances glycolysis [Li et al., 2016]. For GAPDH, PKC-directed T246 phosphorylation decreases GAPDH tetramerization, reduces GAPDH glycolytic activity and inhibits GAPDH-directed mitophagy in-vitro and ex-vivo [Yogalingam et al., 2013,Qvit et al., 2016]. High proportion of S203p on PGK1 were observed in MSC and MON, and at a comparable frequency of T246p on GAPDH in MSC, MON, LYM and GRA. The low frequency of phosphorylation on these two sites in HPCs and ERPs implied that the glycolysis rate in the two cell types may be comparatively higher than in the other four cell types.

The differentially regulated acetylation sites also affect the localization of glycolytic enzymes in various ways. Acetylation directly regulates the abundance of PKM2. Enhanced acetylation of K305 on PKM2 has been shown to prevent the tetramerization of active enzymatic form and reduces the synthesis of pyruvate and acetyl-CoA [Park et al., 2016a]. K305 acetylation on PKM2 enhances its interaction with HSP70, which in turn increased its uptake by lysosome, leading to degradation [Xiong and Guan, 2012]. Mutation of PKM2 on K305 further results in accumulation of several glycolytic intermediates and stimulation of cell proliferation and growth [Lv et al., 2011]. Acetylation of K305 on PKM2 was observed in all bone marrow cells, and proportionally higher in HPCs and GRAs, suggesting that the turnover of PKM2 in bone marrow through acetylation is stronger in the two cell types. In addition, acetylation directly regulate the subcellular localization of metabolic enzyme GAPDH, interfering with its enzymatic and non-enzymatic functions. Cytoplasmic GAPDH is a key enzyme in glycolysis, whereas nuclear GAPDH have many other functions including transcriptional regulation, DNA repair, and telomere maintenance [Zheng et al., 2003]. Acetylation of K117 on GAPDH by PCAF is required for nuclear translocation of GAPDH [Ventura et al., 2010]. Hence the high proportion of K117 acetylation on GAPDH in HPCs presumable plays an active role in maintaining the genomic integrity of the stem cell population in the human bone marrow.

Taken together, the observed cell type specific acetylation and phosphorylation differences can be linked to the localization of glycolytic proteins in the bone marrow cells. This may indirectly regulate the catalytic and non-catalytic activities of those metabolic proteins.

8.4 Conclusion and perspective

Proteins from the hnRNP family were preferentially acetylated in the primitive HPC populations, which can potentially affect the pattern of alternative splicing in these bone

marrow stem cells. In addition, proteins in the glycolysis pathways are differentially phosphorylated and acetylated, modulating their subcellular localization. The compartmentalization may differentially regulate turn-over rate, catalytic and non-catalytic activity of those metabolic proteins across cell types in the human bone marrow.

Further in-vitro and in-vivo experiments could be performed to verify the identified modification sites. Of particular interest, inhibition of phosphorylation or acetylation either with inhibitors or site-directed mutagenesis could reveal the regulatory function of the identified modification sites on the glycolytic enzymes. Similarly, applying mutagenesis approach on the hnRNP proteins will enable us to understand whether the identified modification sites regulate binding specificity of the RNA transcripts, activity of alternative splicing and stability of hnRNP proteins. Alternatively, the hnRNP proteins could also be examined using acetylation inhibitors followed by transcriptomics analysis to understand down-stream effects on different haematopoietic cell populations.

9 Dynamics of Protein Phosphorylation, Acetylation and Methylation As Readouts of Biological Ageing in Human Bone Marrow

9.1 Introduction

Several studies have described age related alterations in either changes in protein abundance or levels of PTMs such as acetylation and phosphorylation [Johnson et al., 2013, Walther et al., 2015, Pal and Tyler, 2016]. In this study, we are able to consider both the unmodified proteins and the modified proteins as a whole and examine their variations with respect to age across six different bone marrow cell types. Through the combination of standard and mass-tolerant database search method, unmodified and modified proteins were identified. Proteins regulated by phosphorylation, acetylation and methylation were analyzed for age related changes. We seek to define primarily, how the profiles of PTMs change with age across different cell types, and further to understand if any biological processes are particularly changed and might serve as potential biomarkers for ageing. The results presented in this chapter show the age-associated dynamics of PTMs across different types of bone marrow cells.

9.2 Result

9.2.1 Protein post-translational modifications in human bone marrow largely remain stable during ageing

Bone marrow cells collected from 59 donors with age from 20 to 60 were analyzed using quantitative shotgun proteomics. For every protein identified, I was able to quantitatively measure one to two spectrum for the modified peptides in each donor using TMT isobaric tags. This low number of quantified spectrum per modified peptides implies estimates of modification levels at the peptide level would be more sensitive to measurement noise. Therefore, I implemented a more robust estimate of the level of modification at the protein level. For each modified protein, I averaged the TMT intensity across all modified peptides and termed this **PTM score**. This PTM score reflects the likelihood for a protein to be modified. When comparing two PTM scores for the same protein between two samples, a higher PTM score indicates that the protein in one sample is more frequently modified. However to apply statistical methods such as linear model or any other tests assuming a normal distribution, one might not have sufficient power to infer the exact modified peptide that effects this difference.

With the PTM scoring method, 695 phosphorylated proteins, 258 methylated proteins and 328 acetylated proteins were quantified across six bone marrow cell types (Figure 28A). For each modification, the proportion of proteins quantified to all modified proteins

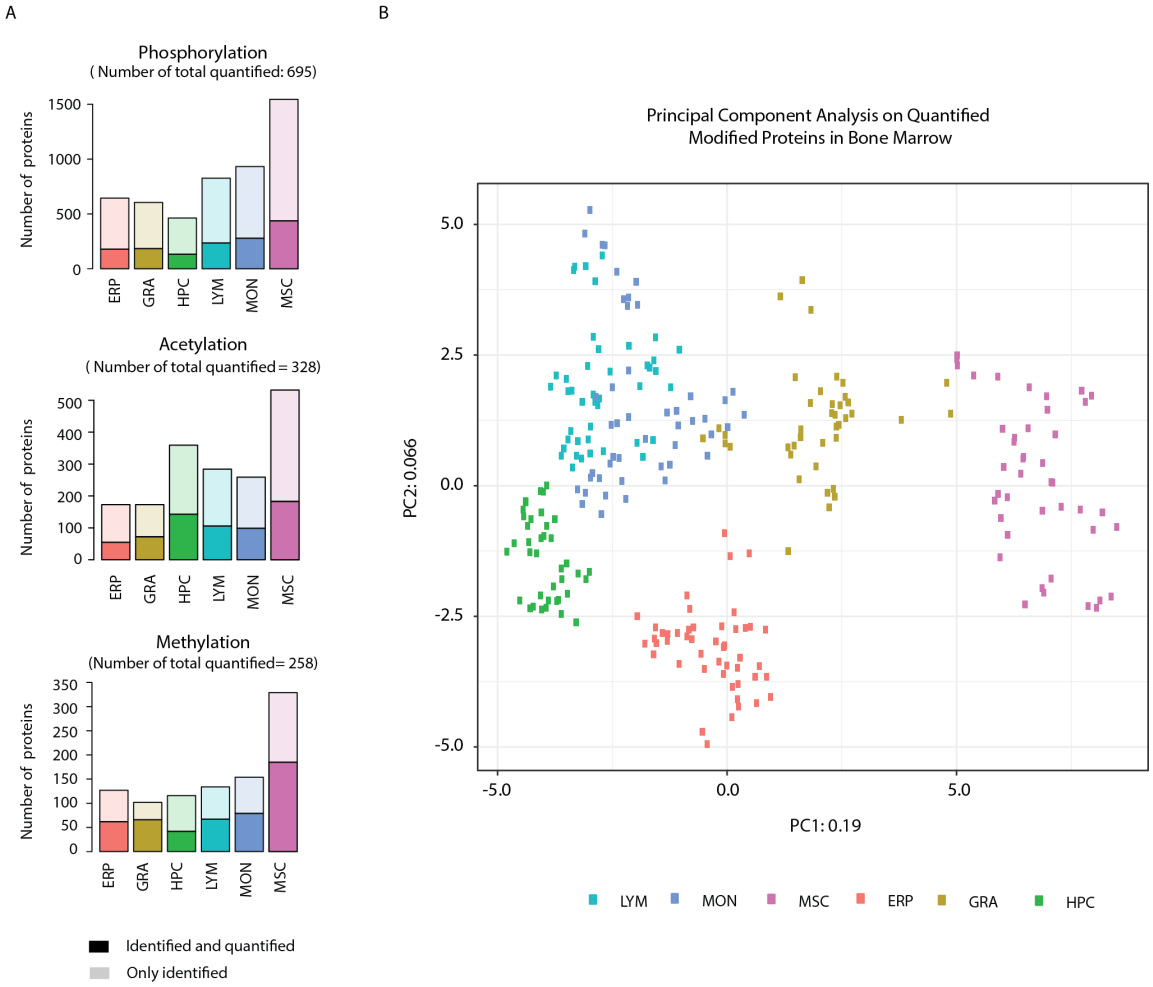


Figure 28: Modification levels of the modified bone marrow proteins showed cell-type specific differences. A) for the proteins identified to be phosphorylated, methylated and acetylated, the percentages of proteins quantified based on PTM score were at comparable levels across six cell types. The numbers of phosphorylated proteins that can be quantified for ERPs, GRAs, HPCs, LYMs, MONs and MSCs were 178, 184, 133, 235, 278 and 437, representing 28 to 30% of the identified phosphorylated proteins. The numbers of acetylated proteins that can be quantified for the six cell types (same order as above) were 55, 72, 143, 106, 99 and 183, representing 32 to 42% of the identified. The numbers of methylated proteins that can be quantified for the six cell types (same order as above) were 62, 66, 42, 67, 79 and 185, representing 36 to 65% of the identified. B) principle component analysis on the PTM scores of the modified proteins (including phosphorylated, acetylated and methylated forms) showed that a large part of the variations across samples can be explained by cell types.

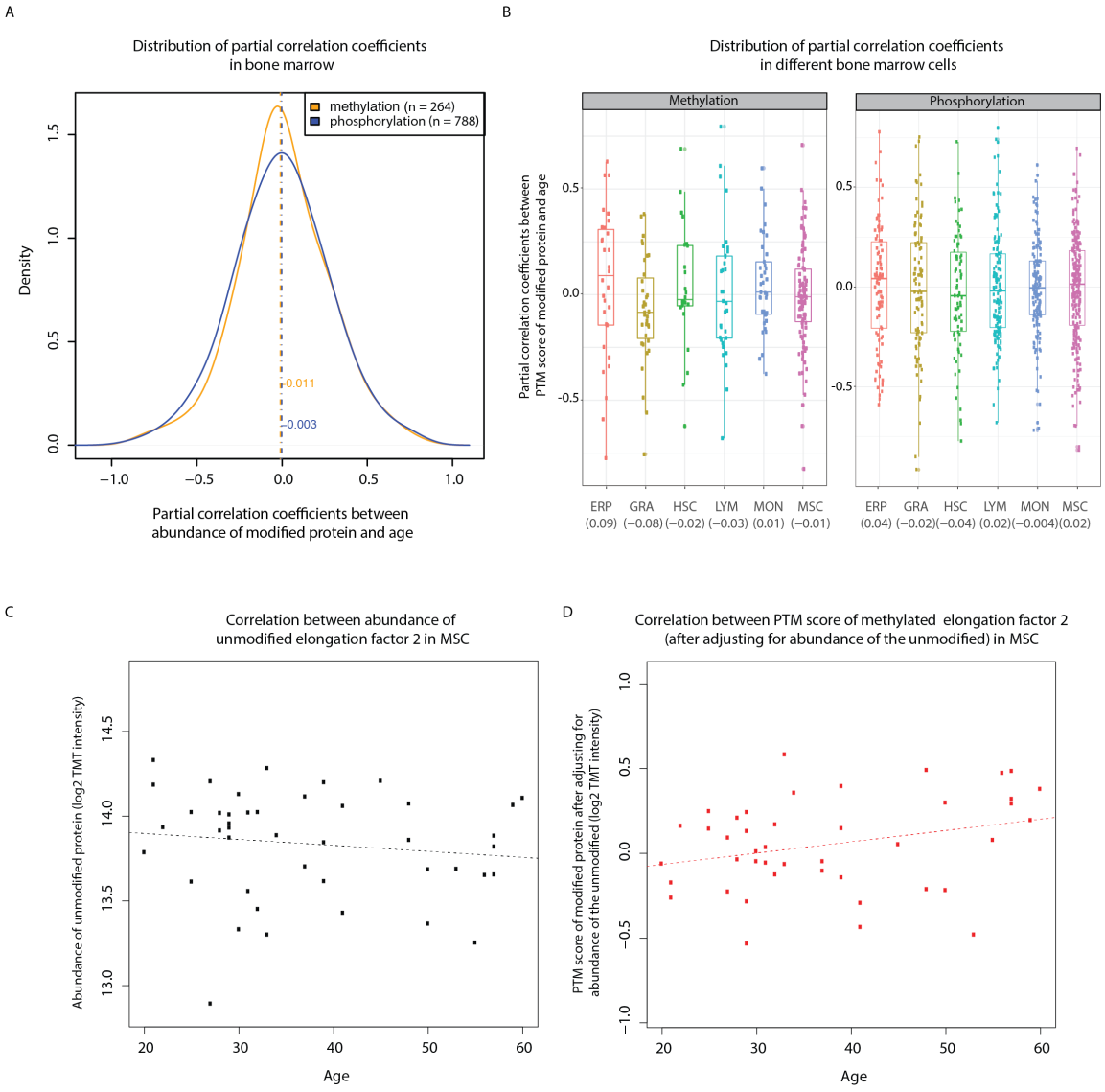


Figure 29: Modification levels for phosphorylated and methylated proteins in the human bone marrow remained largely stable during ageing. A) the partial correlation coefficients of phosphorylated proteins and methylated proteins in the bone marrow were centred around 0, with median values of -0.003 and -0.011 for phosphorylation and methylation, respectively. B) the overall distributions of partial correlation coefficients of phosphorylated proteins and methylated proteins showed similar distribution across six bone marrow derived cell types. The median of partial correlation coefficients were shown with box plots and annotated below the plot. C) in MSCs the abundance of the unmodified elongation factor 2 had a slight negative correlation with increasing age, whereas D) the PTM score of the methylated elongation factor 2 (corrected for the abundance of the unmodified) were positively correlated with age. The partial correlation coefficient was 0.31.

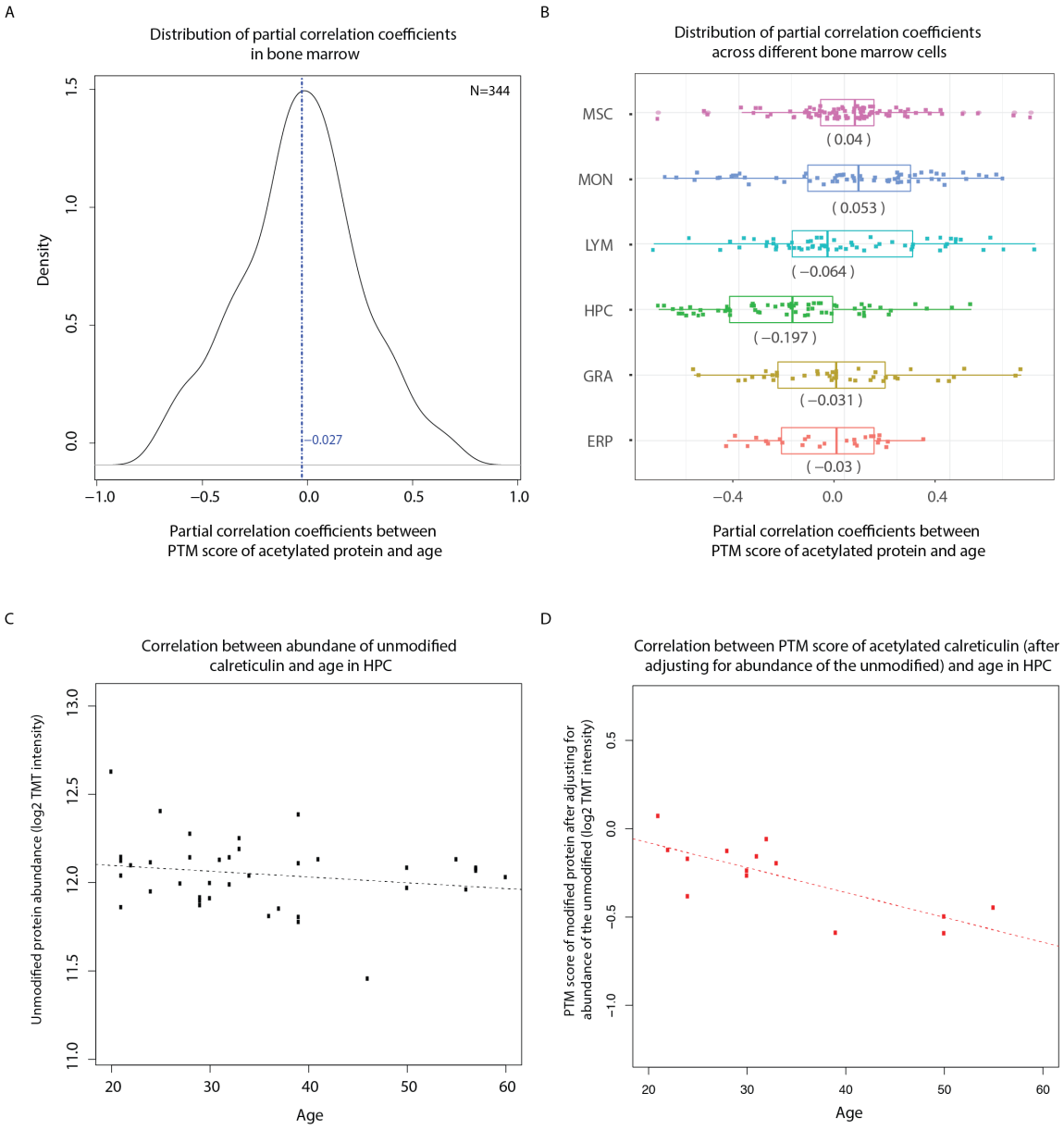


Figure 30: Modification level for the acetylated proteins in HPCs were biased towards negative correlation with age. A) the partial correlation coefficients of acetylated proteins in the bone marrow were centred around 0, with a median of -0.027. B) the distributions of partial correlation coefficients of acetylated proteins across different bone marrow cells showed that the distribution in HPCs was towards a negative correlation with age than in the other bone marrow cells. The median of partial correlation coefficients of the six cell types were shown with box plots and annotated below. C) in HPCs the abundance of the unmodified calreticulin showed slightly negative to no correlation with increasing age, whereas D) the PTM score of the acetylated calreticulin (corrected for the abundance of the unmodified) were negative correlated with age in HPCs. The partial correlation coefficient was -0.70.

detected were comparable across six cell types. The modified proteins that can be quantified were 28 to 30% of identified phosphorylated proteins, 32 to 42% of the acetylated proteins, and 36 to 65% of the methylated proteins.

After quantifying the modified proteins with PTM scores, the first question I asked is what are the factors that best describe the observed variations in the PTM scores across all samples. Principal component analysis (PCA) on PTM scores of modified bone marrow proteins revealed that cell type difference can explain a large part of the variation between different samples (Figure 28B). Modified proteins in the haematopoietic and non-haematopoietic lineages were well separated on the first component. These indicates that there are many proteins detected more in their modified forms in MSCs as compared to other cell types. In addition, cells from the haematopoietic lineage were clustered into individual cell types, though cells from the lymphoid and myeloid lineage were not distinctively separated. This indicates that the modified proteins that can be quantified with PTM scoring method was reliable and is representative for the proteome of each cell type. When we further examining the other PC components, there were 2 components showed significant correlation with age (unadjusted pvalue < 0.01) (Supplementary Table 1). This suggested that there were variations associated with age across all individuals and cell types, indicating that the quantification of modified proteins can be further analyzed for age associated variations.

To find out the association between the level of protein modification and the biological age of individual, one fact that needs to be accounted for is that for a given protein, its PTM score is directly associated with the abundance of its unmodified form. The abundance of the unmodified protein may also change during ageing. Therefore I used partial correlation analysis to address this question. In essence, with partial correlation analysis, associations between the PTM score to biological age and between abundance of unmodified proteins to PTM score were first examined. Then the degree of association between PTM score and biological age was measured while controlling for potential effects of the unmodified proteins.

For phosphorylation and methylation, we did not observe heavy tailed distributions or shifts in the median from zero, for the partial correlation coefficients across all proteins (Figure 29A). This implied the bone marrow proteins showed no strong consensus in terms of increased or decreased modification with respect to age. The distributions of partial correlation coefficients between the PTM score and age stratified by cell types further revealed that distributions of partial correlation coefficients were similar across the six bone marrow cell types (Figure 29B). A small proportion of proteins do have modifications changes that are related to age. As exemplified in Figure 29C, elongation factor 2 (eEF2) had a positive partial correlation coefficient of 0.31 in MSCs. The abundance of unmodified

eEF2 was slightly negative correlated with increasing age in MSC populations, whereas on the contrary the PTM score of the methylated eEF2 (after correcting for the abundance of the unmodified) in MSCs showed a clear positive correlation with age (Figure 29D).

Acetylation showed a similar profile of its partial correlation with age as phosphorylation and methylation. The acetylated fraction of the bone marrow proteome displayed no strong correlation with age (Figure 30A). However, when examining the distribution across different cell types, the results showed a bias towards negative association in the HPC populations (Figure 30B). This indicated that the acetylation of a large number of proteins decreased with age in HPCs. Calreticulin was among one of the proteins with negative association between PTM score and age. The partial correlation coefficient of calreticulin was -0.70 in HPCs. As illustrated in Figure 30C and Figure 30D, the abundance of unmodified calreticulin showed slight negative to no correlation with age in HPCs, whereas the PTM score of the acetylated calreticulin (after correcting for the abundance of the unmodified) displayed a clear negative correlation with age.

In conclusion, combining partial correlation analysis with PTM score revealed cell-type specific trends in terms of PTM changes during ageing. Proteins with high partial correlation coefficients could be potential candidates for further analysis on age-association.

9.2.2 Modified proteins in haematopoietic progenitor cells forms correlated clusters during ageing

After assessing the association between changes of protein phosphorylation, acetylation and methylation and age with partial correlation analysis, we asked the question whether these changes are coordinated during ageing. In gene regulatory network analysis, genes that show a coordinated expression pattern across numerous samples are grouped into co-expression clusters based on correlation measures [van Dam et al., 2017]. A similar strategy for grouping modified proteins whose abundance levels are highly similar across samples was applied. We termed this group of modified proteins as correlated clusters.

To determine the correlated clusters of proteins, the pairwise correlation coefficients between the modified proteins were calculated based on their PTM scores. This pairwise correlation is calculated across all cell types and samples. Phosphorylated, acetylated and methylated forms of the same protein were considered separately. The modified proteins were then clustered according to their pairwise correlation coefficient to find the correlated clusters (Figure 31A and Supplementary Figure 12). After defining groups of these coordinated modified proteins independently of age, I further asked whether any of these defined groups of modified proteins show a concerted change during ageing.

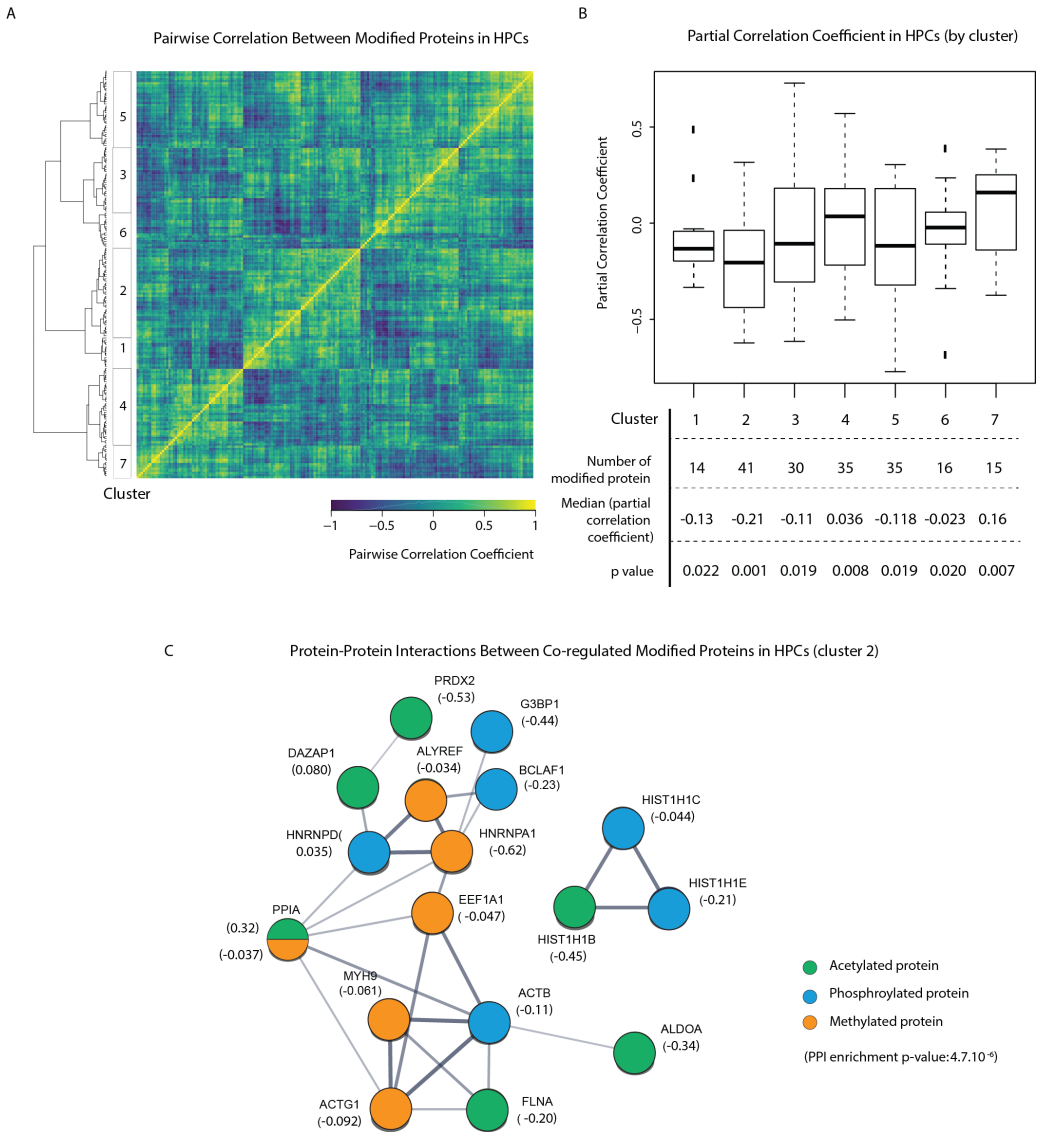


Figure 31: Modified proteins in haematopoietic progenitor cells co-evolved during ageing. A) pairwise correlation between modified proteins in HPCs were calculated based on their PTM scores. The modified proteins were then clustered according to pairwise correlation coefficient using Ward's method, and clustered into 7 groups. B) distribution of the partial correlation coefficient between PTM scores of modified proteins and age within each cluster in HPCs were shown. With ANOVA test, proteins belong to group 2, 4 and 7 were found to have significantly different distributions of partial correlation coefficient from the others after multiple testing adjustment with Bonferroni correction. C) interaction network of proteins belonged to the second group of co-evolved proteins in HPCs. Protein-protein interaction network was built using STRING. For each protein, its partial correlation coefficient between PTM scores and age were annotated in bracket.

Table 3: Modification sites differentially regulated during ageing

Protein	Site	Modification	Partial Correlation Coefficient	p. adj
14-3-3 protein beta/alpha	T32	Acetylation	-0.42	0.042
Methionine–tRNA ligase	Y410	Phosphorylation	0.30	0.079
Clathrin	Y169	Phosphorylation	-0.55	0.079
SEC22b	S137	Phosphorylation	-0.56	0.079

The degree of similarity between modified proteins showed various patterns across different bone marrow cell types. The distribution of partial correlation coefficients between cluster showed differences in the HPC population (Supplementary Table 2). The distribution of partial correlation coefficients between the PTM score and age for modified proteins within each cluster in HPCs were shown in Figure 31B. Overall proteins belonging to cluster 1, 2, 3, 5 and 6 were mostly negatively associated with increasing age (with median partial correlation coefficient from -0.21 to -0.023) , whereas proteins in group 4 and 7 were mostly positively associated during ageing (median partial correlation coefficient 0.036 and 0.16). Proteins belonging to 2,4 and 7 clusters were found to have significantly different distributions of their partial correlation coefficient from proteins in the other groups (adjusted p value < 0.05).

We further asked whether there are any known interactions in the clusters that showed overall association with age. Network analysis was performed on proteins belonging cluster 2, 4 and 7. Proteins belonging to cluster 2 and 4 had significantly more interactions than the number of interactions expected at the genome level (Figure 31C and Supplementary Figure 13). Proteins in cluster 4 were mostly mediated by phosphorylation, whereas all three modifications were found in modified proteins in cluster 2. There were 17 modified proteins in cluster 2 that formed interaction networks, including RNA binding proteins such as hnRNP A1, hnRNP D and elongation factor 1-α1 (EEF1A1), structural component of the cell such as myosin (MYH9), actin (ACTG1, ACTB) and filamin-A (FLNA), and variants of histone H1 (HIST1H1C, HIST1H1B and HIST1H1E). Modifications on these proteins were mostly negatively correlated with age, further suggesting that these modified proteins may participate in processes that are co-operatively suppressed in the process of ageing.

9.2.3 Differential regulation of site-specific modifications in bone marrow during ageing

We attempted to elucidate whether there are any site-specific modifications that changes with age using partial correlation method. As mentioned above, the estimate of the relationship between age and modified peptide abundance comes with large errors due to the low number of quantified spectrum per modified peptides. To circumvent this and to increase statistical power, I combined measurements across cell types to enable a more

accurate estimate of the relationship between age and modified peptide abundance. We have stronger evidence of a relationship between abundance of modified peptide and age if an age-related effect is consistently observed across more than one cell type. To analyse the degree of association between modification frequency at a modification site (the abundance of the specific peptide with modification) and biological age, the potential effect of the unmodified was controlled using partial correlation analysis.

In total 547 modified peptides were analysed, including 303 phosphorylated peptides, 98 methylated and 146 acetylated. The distribution of the partial correlation coefficients of those modified suggested that the majority of the modification sites were not significantly altered during ageing. After adjusting for multiple testing using Benjamini-Hochberg, four peptides were found to be differentially modified during ageing (Table 3). However, no functional information is known about modifications on these sites.

9.3 Discussion

9.3.1 Abundance of modified proteins in bone marrow derived cells shows cell-type dependencies and correlations with age

We observed a bias towards negative correlation between frequency of acetylation and age in HPCs. This may be attributed to the altered pluripotency in the aged stem cell populations. Bone marrow haematopoietic stem cells are known to have high susceptibility to age-related changes such as skewed differentiation potential [Morrison et al., 1995, Wahlestedt et al., 2017]. Perturbation of global acetylation can affect stemness of cells. For example, studies on embryonic stem cell showed that the differentiation of stem cell can be delayed through using inhibitors upstream of acetyl-CoA [Moussaieff et al., 2015]. In addition, over-expression of deacetylase SIRT2 in human pluripotent stem cells (hPSCs) also affect stem cell functions including differentiation capacity [Cha et al., 2017]. Therefore, the decreased amount of acetylated proteins in HPCs was in good agreement with previous observations.

Although levels of modification for most modified bone marrow proteins were relatively stable during ageing, there were individual proteins observed to have strong correlations with age across different cell types. For example, the abundance of methylated eEF2 was positively correlated with increasing age in MSC population. Function of eEF2 is positively regulated by longevity- promoting rapamycin through inhibition of eukaryotic elongation factor 2 kinase (eEF2K) [Redpath et al., 1996, Proud, 2015]. Hence, one might hypothesize that the increased eEF2 methylation in MSC upon ageing is associated with its diminishing activity. There are few studies on the mechanism of eEF2 methylation, but one possible mediator of this methylation - eEF2 activity relationship would be FGF signalling [Jung et al., 2011]. The fibroblast growth factors (FGFs) are important regulators

for the development of bone from MSCs and methylation of eEF2 was shown to be induced by basic fibroblast growth factor (bFGF) through MAPK signalling pathway in mouse embryo fibroblast NIH3T3 cells [Jung et al., 2011, Coutu et al., 2011]. The exact functional consequence of eEF2 methylation needs to be clarified with further experiments.

9.3.2 Changes of modification levels on histone H1 variants may modulate chromatin structure in bone marrow haematopoietic progenitor cells during ageing

We were able to infer clusters of modified proteins in HPCs that are highly coordinated in terms of modification frequency, which might be a result of common regulatory pathways during ageing. Three variants of histone H1, namely H1.2 (HIST1H1C), H1.4 (HIST1H1E) and H1.5 (HIST1H1B) were correlated through phosphorylation and acetylation. There are 7 human histone H1 isoforms ubiquitously expressed in somatic cells. H1.1, H1.2, H1.3, H1.4 and H1.5 are expressed during replication, and H1.x and H1.0 are expressed throughout the cell cycle [Izzo and Schneider, 2016]. H1 histones are linker proteins that binds to the entry and exit sites of DNA on the surface of nucleosomal core particles. They are dynamic components of chromatin and are required for the stability of chromatin structure [Hergeth and Schneider, 2015]. Like other histone proteins, H1 histones were found to carry different modifications. Phosphorylation and acetylation of histone H1 variants have been observed in regions where directly involved in DNA binding and are known to affect DNA condensation capacity of H1 proteins [Talasz et al., 2009, Wiśniewski et al., 2007, Roque et al., 2008, Hergeth and Schneider, 2015].

We observed negative correlations between levels of phosphorylation of histone H1.2 and H1.4 and age in HPCs. One possible explanation would be that during ageing, the phosphorylation level on histone H1 in HPCs, especially H1.4 decreases, reducing the compaction of chromatin and contributing to immunosenescence. Recent studies on mRNA expression and mouse model have suggested that individual H1 variants play significant roles in embryonic stem (ES) cell differentiation [Terme et al., 2011, Zhang et al., 2012]. In pluripotent human NTERA-2/D1 (NT2) cells and mouse ES cells the levels of H1.4S187 phosphorylation and H1.2/H1.5S173 phosphorylation decrease globally during differentiation [Liao and Mizzen, 2017]. Furthermore, in human peripheral blood lymphocytes, it has been observed that the phosphorylation of H1.4 and H1.5 is significantly decreased with increasing age, contributing to an increase in senescence-associated heterochromatin formation [Happel et al., 2008]. All these evidences substantiate a potential link between histone H1 phosphorylation and reduced differentiation potential in bone marrow HPCs during ageing.

In addition to phosphorylation, the decreased level of acetylation on histone H1.5 may also impact the chromatin compaction, potentially affecting differentiation and prolifera-

tion potential of the HPCs during ageing. In differentiated human ES cells, histone H1.5 is shown to bind preferentially to genes encoding membrane and membrane-related proteins [Li et al., 2012]. Binding of H1.5 is associated with gene repression and is required for chromatin compaction, whereas depletion of H1.5 resulted in increased chromatin accessibility, deregulation of gene expression, and decreased cell growth. There are few studies on histone H1.5 acetylation, but functions of acetylation have been analyzed on the other histone H1 variant H1.4. Acetylation of H1.4 is known to activate transcription by directly reducing H1 affinity to chromatin and by recruiting subunit of the transcription factor [Kamieniarz et al., 2012, Terme et al., 2014]. One could hypothesise that the decreased acetylation of histone H1.5 may follow a similar mechanism, disrupting the maintenance of condensed chromatin and selectively affecting transcription of a subset of genes, but this remains to be elucidated in future studies.

9.4 Conclusion and perspective

At the level of the entire bone marrow proteome, the levels of phosphorylation, acetylation and methylation on the bone marrow proteins remained relatively stable during ageing. A small proportion of modified proteins displayed age-associated changes on modification frequencies, but no specific pathways enriched for these specific modified proteins were observed. Of note, the HPC populations showed a bias towards negative association between acetylation level and age, which may correlate with its altered pluripotency. We also observed that the coordinated modification patterns of histone H1 variants. The decreased levels of phosphorylation and acetylation may have an impact on the chromatin structure, contributing to immunosenescent features in the elderly population .

As the results of this study recovered only a part of the entire modified bone marrow proteome, the number of modification sites that allows detection of significant associations with age is rather limited. Nonetheless it is reasonable to believe that there are potentially more proteins differentially modified at specific modification sites upon ageing, which alter the functions of the proteins and ultimately contribute to the age-associated senescence of the immune system. Further studies with targeted approaches, focus on particular cell types (e.g HPC) and increased number of samples can increase the power greatly, and yield more significant results to help to uncover the correlation between protein modification and ageing in the human bone marrow.

10 Conclusion and Perspectives

The human bone marrow proteome was represented by six main sub-populations of cells and quantitatively measured by shotgun proteomics. The combination of standard and mass-tolerant database search methods effectively identified abundant PTMs in the bone marrow and quantified bone marrow proteins with those abundant modifications.

The bone marrow proteins were extensively modified. The modification profiles of phosphorylation, acetylation and methylation were cell-type dependent, with distinct separation between cells from the haematopoietic and non-haematopoietic lineages. Proteins differentially modified across cell types regulate a diverse spectrum of processes inside the bone marrow, including glucose metabolism and alternative splicing.

On top of cell type differences in PTMs, this study allowed us to examine changes of the human bone marrow proteome across timespan of 40 years. Modification level for the majority of the modified proteins in bone marrow remained stable during ageing. In the HPC populations a negative association between acetylation level and age was observed, which may be linked to diminished differentiation potential of the stem cells during ageing.

Taken together, these results suggest that post-translational modification patterns can inform us of the differences in pathway regulation across cell types, and also have the potential to serve as either biological markers of the ageing processor indications of different pathway activities during ageing.

As the modified proteins identified in this study only represented a partial image of the entire modification landscape of the human bone marrow proteome, one can anticipate that further studies with more elaborate study designs can deepen our insight on the importance of PTMs in bone marrow during ageing. Targeted approaches will allow us to survey more broadly the dynamics of protein modifications across cell types. Apart from further complementary experiments to confirm the functions of these modification during ageing, one can also envision the application and development of computational strategies to model and predict the degree of ageing in an individual, based on integrating post translational profiles with other genetic and phenotypic information, for example, gene expression and epigenetic changes.

11 Material and Method

11.1 Sample collection

Cells were collected from donated bone marrow (donor age between 20 and 69) at Heidelberg University Hospital under the approval by the ethics committee. Bone marrow aspirates from the iliac crest of healthy human subjects were washed with phosphate-buffered saline (Sigma-Aldrich) and transferred to Ficoll-Paque (15 mL, Biochrom). Mononuclear cells (MNCs) were obtained through centrifugation at 800g for 30 min. For cell-type-specific internal standards, MNCs were collected from umbillical cord blood using the same procedure. Cells from the five cell types from haematopoietic lineage,namely monocyte progenitor cells (MONs, CD34-, CD14+, CD45+), granulocyte progenitor cells (GRAs, CD34-, CD14-, CD45+, high Ssc-a), lymphocyte progenitor cells (LYMs, CD34-, CD14-, CD45+, low Ssc-a), erythroid progenitor cells (ERPs, CD34-, CD14-, CD45-) and haematopoietic progenitor cells (HPCs, CD34+) were isolated by fluorescence-activated cell sorting (FACS) using haematopoietic lineage marker antibodies and side-scattered light index accordingly. The bone marrow mesenchymal stromal cells were obtained through in vitro culture of MNCs. The MNCs were seeded in fibronectincoated culture flasks (75 cm^2) with a density of ≈100 cells/mL in Verfaillie medium till adherent colonies were formed. MSCs were then scraped from the culture flask. All cells were stored at -20°C before lysis.

11.2 Cell lysis

The quantity of cells used for each donor in the analysis is cell-type dependent due to the differences in the protein concentrations across different cell types and sample availability. For each donor, 1.10^5 of HPCs, 1.10^6 of ERPs, 5.10^5 of LYMs, MONs, GRAs and MSCs were used. Frozen cell pallets were lysed in HEPES buffer (200mM HEPES, PH 8) containing 1% protease inhibitor cocktail (Sigma-Aldrich) and 0.5% RapiGest SF (Waters). Cell lysate were incubated at 90°C for 5 min, followed by 20 min sonication and 5 min centrifugation with 16,000g to precipiate DNA and cell debris. Proteins were collected from supernatant layer and incubated with 2mM dithiothreitol (Sigma-Aldrich) at 23°C for 30 min with 300 rpm. The denatured proteins were then reacted with 5mM iodoacetamide (Sigma-Aldrich) in dark for 30 min, and then quenched by 5 min UV-exposure with natural light. The reduced and alkylated protein mixtures were firstly digested by Lys-C enzyme (1:1000 (w/w), 0.01μg/μl Lys-C in HEPES buffer, Wako) for 3 hours at 37°C, followed by digestion with trypsin (1: 500 (w/w), 0.01μg/μl, in 1% (v/v) trypsin resuspension buffer) for 16 hours at 37°C.

11.3 Peptide labelling and separation

Digested peptides were labeled with TMT-6plex isobaric mass tags (Thermo) according to the manufacturer's instructions. Cell-type specific internal standards were labelled with

TMT channel 126, whereas the samples from donors were randomized for their age and labelled with TMT channel 127 to TMT channel 131. 20μl of TMT (0.02mg/μl in acetonitrile) was directly added to the digested peptide mixture and reacted at 23°C for 1 hour with 300 rpm. The reaction was terminated by incubation with 5% (v/v) hydroxylamine at 23°C for 15 min with 300 rpm. 10% (v/v) trifluoroacetic acid was added to the labelled mixture to condition the buffer to PH below 3. The mixture was incubated for 45 min at 37°C with 300 rpm to cleave the acid-labile structure on RapiGest, and centrifuged for 20 min at 4°C with 16,000 g to pallet the insoluble part of RapiGest. TMT-labelled peptide was collected from supernatant and desalted on SepPak C18 column. SepPak column was first equilibrated with 1ml elution buffer (0.6% (v/v) acetic acid in 80% (v/v) acetonitrile), and 1ml washing buffer (0.6% (v/v) acetic acid). Peptide mixture was conditioned with additional 500 μl washing buffer and loaded to the equilibrated SepPak column, washed with 1ml washing buffer and eluted with 500μl elution buffer. Eulate from each TMT channel were concentrated in speed vacuum concentrator to 50 μl and adjusted to PH >7 using 2 M sodium hydroxide (Sigma-Aldrich).Subsequently the eluate were labelled again using same amount of TMT of the same channel, using aforementioned procedure. The double-labelled samples were concentrated to 50 μl, washed using SepPak column following the same procedures and eluated with 500 μl elution buffer.

To determine the labelling efficiency, 25 μl eluate from each TMT channel were combined, concentrated in vacuum concentrator, protonated with 1% formic acid and analyzed by Q Exactive mass spectrometer (Thermo Scientific). Ratio among TMT channels were obtained by analyzing data using Proteome Discoverer (version 1.4.1.14, Thermo Fisher Scientific). If the results shows that any of the samples have over 2 fold deviation from the mean of the ratio across all six TMT channel, a second check were performed prepared based on the value obtained from the first mixture. After the mixing ratios were adjusted within 2 fold deviation from the mean of the ratio across all six TMT channel, samples were concentrate to 50 μl for peptide separation.

11.4 Reversed-phase chromatography separation of the TMT-labeled peptides

The concentrated peptide mixture was fractionated using reversed-phase liquid chromatography (Agilent 1260 infinity HPLC system) with equipped with XBridge C18 reversed-phase column (3,5 μm 1 x 100 mm, Waters). Fractions were concentrated and desalted using Oasis HLB μElution Plate. Plate was equilibrated with 200 μl elution buffer, followed by flushing with 200 μl washing buffer. Peptide mixture was conditioned with additional 100 μl washing buffer and loaded onto plate. Fractions were combined to 18 samples as shown in Supplementary Figure 14. Peptides were washed with 200 μl washing buffer and eluated with 50 μl elution buffer. All fractions were concentrated to less than 10 μl and protonated with 1% formic acid for LC-MS/MS analysis.

11.5 Nano reversed-phase chromatography tandem mass spectrometry

The fractions were analyzed on LTQ Orbitrap Velos Prof mass spectrometer (Thermo) equipped with nanoAquity UPLC flow column (Waters) and Proxeon nanospray source (Proxeon Biosystems) connected. Each fraction was separated on C18 column. The HPLC eluate was directly electrosprayed into the mass spectrometer. Full scan spectra, from 300 m/z to 1,700 m/z at resolution of 30,000 (profile mode) were acquired in the Orbitrap mass analyzer. Ions with 10 highest intensities from the full scan MS were selected for fragmentation in the ion trap. Only multiply charged (2+, 3+) precursor ions were selected for MS/MS. The dynamic exclusion list was restricted to 500 entries with maximum retention period of 30s and relative mass window of 10 ppm.

11.6 Database search

To check the TMT labelling efficiency, MS/MS spectra were searched against the Uniprot human database using Proteome Discoverer (version 1.4.1.14, Thermo) and MASCOT search engine (Matrix Science). The following search parameters were used: peptide mass tolerance at +/- 20 ppm, MS/MS tolerance at 0.1 Da, trypsin enzyme with up to 2 missed cleavages, fixed modification of TMT 6-plex (lysine), TMT 6-plex (N-term), variable modification of methionine oxidation and cysteine carbamidomethylation, and a reverse decoy database pattern. All reported data were based on 95% confidence levels for protein and peptide identification as determined by the false discovery rate (FDR) of no more than 1%; All identified proteins were filtered for with at least one unique peptide.

For identification of unmodified proteins and proteins with modifications, ProteoWizard was used to convert mass spectrometry data from Thermo raw file format to mzML format. All MS/MS spectra assignments were performed using Comet software (version "2016.01 rev. 2") against Uniprot human database. Decoy database was constructed by reversing sequences from Uniprot human database. The database search parameters were as follows: trypsin was selected as the enzyme; a maximum of two missed cleavages were allowed; precursor ion tolerance (MS1 search window) was set to 15 ppm; fragment ion bin tolerance was 0.02; cysteine carbamidomethylation (+57.021464 Da), methionine oxidation (+15.9949 Da), lysine and peptide N-terminal TMT-6plex tagging (+229.1629 Da) were selected as dynamic modifications. When performing search for proteins with modifications, the modifications of interests were added to the dynamic modification search parameter; for each file 20 CPU were used for spectra assignment, which took approximately 20 to 40 min. For identification of potential modifications, search parameters were the same as for identification of unmodified proteins with a few differences: maximally five missed cleavages were allowed; precursor ion tolerances were set to 100 Da, 200 Da, 300 Da, 400 Da and 500 Da; for each file 40 CPU were used for spectra assignment. Depending on the precursor ion tolerance, the database search for one file may take up to 8 -12 hours.

A protein was accepted as being identified when the following criteria were all fulfilled: 1) for single spectrum to sequence assignment, the assignment was the best match by Comet. 2) at least 3 unique peptides were detected for the same protein. 3) passed 1% false discovery rate at protein level.

11.7 Protein and peptide quantification

Intensity values of reporter ions were extracted from raw data using Bioconductor R package "mzR". The reference molecular weight of TMT tags used were 126.2193 Da for TMT126, 127.2127 Da for TMT127, 128.2046 Da for TMT128, 129.1981 Da for TMT129, 130.1900 Da for TMT130 and 131.1834 Da for TMT131. When obtaining intensity values from raw data, the mass value between two TMT channels were used to separate signal from different TMT channels. Intensity values of one TMT channel was accepted only when the values are above the median intensity value of all spectra obtained within the same scan. The intensity of one TMT channel of one scan were the sum of all intensity values of spectra that passed the median cutoff.

For quantification of proteins, all 18 fractions from one pool were consolidated. Only proteins with more than 2 unique peptides were used for quantification. The abundance of each protein is estimated by taking the mean intensities of all its PSMs. Modified protein abundance were estimated separately from unmodified proteins, for example, abundance of modified phosphorylated protein A was estimated using only phosphorylated peptides found in protein A. After obtaining estimates of abundance for all proteins in each sample, the protein abundance for each sample was normalized by taking the log2 value, followed by scale function in R (center around its mean and scaled by standard deviation).

11.8 Mass shift peak picking

For identification of mass shift peaks, spectra which has no peptide matching during identification of unmodified peptides and has a matching in searches for potential modifications were kept, whereas any conflicting spectra, i.e. spectra identified in both searches for unmodified peptides and modified peptides were discarded. For each spectra, mass shift was calculated by subtracting the experimentally acquired mass value by its predicted theoretical mass value. Gaussian kernel function was applied to calculate the density of the mass shifts. Mass shifts were partitioned into discrete bins, with a bandwidth of 0.0025. Second derivatives of density of each bin was calculated. Local maxima of the second derivatives were called out as mass shift peaks. We estimated the background signal as the mean of the density of those non-empty bins, and used 1.5 of the mean as the cutoff to select the mass shift peaks that above background. Peak picking was performed for all open search results from 100Da to 500Da.

11.9 Deduction of PTM from mass shift

Protein PTMs were deduced from comparing the molecular weight of known protein PTMs in Unimod database with the mass of the mass shift peaks. Mass shift peaks identified from all mass-tolerant searches were checked for whether the mass shift values of the identified mass shift peaks match to the molecular weight of any known protein PTMs would induce on peptide. The mono-isotopic masses of modifications annotated in protein modification database Unimod were used as reference. Peptides with mass shift values within the range of 0.01 Da from molecular weight of any known modifications in Unimod database were collected and considered as candidates. Furthermore, for the peptides under the mass shift peaks that have a match to a modification to Unimod database, the frequencies amino acids were calculated. A Fisher's exact test was applied to to determine whether there is any enrichment of animo acids ($p < 0.05$). A modification was considered to be present in the dataset only if the enriched amino acids correspond to the ones required for the matched Unimod modification. Modifications inferred from all mass tolerant search results from all six cell types were combined.

11.10 Cross-checking modified peptides in PhosphoSitePlus Database

Peptides identified with phosphorylation from the standard search including phosphorylation as dynamic search parameter were collected and compared with phosphorylated peptides annotated in PhosphoSitePlus database (version 03-21-2017). For the reference peptides, phosphorylated peptides in PhosphoSitePlus, both the peptides annotated to have been identified in high-throughput and low-throughput studies were chosen. A phosphopeptide was considered as verified if the same modification site has been identified in previous studies. Similarly, acetylated peptides and methylated peptides identified from the standard searches were compared against the acetylated peptides and methylated peptides annotated in PhosphoSitePlus database.

11.11 Gene Ontology enrichment

Bioconductor R package "biomaRt" was used for mapping protein identity to the corresponding Gene Ontology indeitity. The Gene Ontology (GO) data from homo sapiens was used for enrichment analysis. For enriched GO terms, a Benjamini-Hochberg (0.01%) was applied to address the multiple testing issue.

11.12 Enrichment of protein domains for modified peptides

Unmodified, phosphorylated, methylated and acetylated peptides were mapped to InterPro database to retrieve information on the corresponding protein domains. Occurrence of each peptide in each cell type was calculated. A Fisher's exact test followed by Benjamini-Hochberg (0.01%) was applied to the ratios of occurrence for each protein domain in the

modified to the unmodified to determine the enrichment of protein domains in the modified proteins.

11.13 Frequencies of amino acids flanking modification site

We extracted the sequences flanking the modification site, using 10 amino acids in both directions. For each of the 20 amino acids, we tested at each position (-10 to +10 of the modification site), whether the frequency of the particular amino acid at the particular position is different from its frequency across all other positions. A two -sided Fisher test is used to determine the significance of the enrichment. We converted the p values from the fisher test using a $-log10$ transformation. The transformed $log10$ p values were assigned a positive value if the amino acid is enriched, and a negative value if the amino acid is depleted.

11.14 Peptide sequence and phosphopeptide motif enrichment

Phosphorylated peptides were extracted from the data and visualized for phosphopeptide motif using pLogo [O'shea et al., 2013]. Peptide motif enrichment analysis was performed using MEME [Bailey et al., 2009]. The Uniprot human database was used to calculate background amino acid distribution. We restricted to search motif with minimum length of 6 amino acids to maximum length of 10 amino acids.

11.15 Identification of PTM hotspot

A protein PTM hotspot was defined as a 15 amino acid sequence on protein with more than 3 detected modification sites. Phosphorylated, acetylated and methylated peptides were counted for calculating PTM hotspot. All modified peptides that were within 3 amino acids of another peptide were collapsed into one group. The most common peptide within a group was used to represent the group. Proteins identified with PTM hotspots were clustered according to the types of modifications within the hotspot using binary function.

11.16 Distribution of phosphorylation frequency around other PTMs

For each protein, I calculated for each phosphorylation site, the nearest distance to each type of PTM that reside on the same protein (for example methylation). Then I can classified each phosphorylation site based on this distance, for example, whether it is 10 amino acids, 20 amino acids or other amino acids acid distances (in steps of 10 amino acids) away from its closest PTM of interest. I also calculated the total number of phosphorylated peptides harbouring each phosphorylation site. Using the classification and the phosphorylated number as a proxy for phosphorylation frequency, I estimated at each distance, the proportion of phosphorylation that were detected at this distance. Confidence interval for this proportion were estimated using the binomial proportion interval.

11.17 Reactome pathway enrichment

Bioconductor R package "ReactomePA" was used for mapping modified proteins to pathways annotated Reactome database. All proteins identified in the bone marrow were taken as background to determine the phosphorylated, acetylated and methylated proteins are significantly enriched in a pathway. The p value obtained for each pathway was corrected for multiple testing using Benjamini-Hochberg. A final FDR cutoff of 0.05 was applied.

11.18 Cell-type specific heterogeneity of modification sites

All modified peptides that were within 3 amino acids of another peptide were collapsed into one group and represented by the most common one within the group. For a given protein, the difference across all modification sites detected in six cell types were compared using Chi-square based test with simulated p value. Only proteins with at least 5 modified peptides in at least two cell types for each modification were analyzed. Benjamini-Hochberg was used to address multiple testing. A final FDR cutoff of 0.05 was applied.

11.19 Identification of differentially modified proteins in bone marrow

For a given protein, frequency of modification was estimated as the ratio of its modified peptide count to the unmodified peptide count. For each modified protein, the differences of the modification status was compared using frequency of modification across cell types using an extension of Fisher's exact test, followed by multiple testing correction with Benjamini-Hochberg (FDR < 0.05). Only proteins with at least 10 unmodified peptides and 3 modified peptides for each modification were analyzed.

11.20 Comparison of modification frequency of proteins in glycolysis pathway

Proteins involved in glycolytic pathway were extracted from ReactomePA database. Assuming that for each modification, each potential modified residue on proteins within the glycolytic pathway has the same probability of being modified, the total number of peptides for a given protein was counted (adjusting for protein abundance), and multiplied by the total number of modifiable residues per peptide (adjusting for number of modification sites). The resultant number is the total expected modification sites for the given protein. Scaling factor for this protein thus can be calculated by diving the total expected modification sites of the protein by the mean of the expected modifications sites for all proteins in the glycolytic pathway. To determine whether there is any cell-type preference for modifications (i.e. more dominant in one cell type than the others) for individual protein in the pathway, for each protein and each modification, the number of its modified forms was normalised by the total amount of its protein forms (both modified and unmodified). A binomial test was

applied with the null hypothesis that for each modification, the modification rate of the same protein was the same among all six cell types. Proteins were then filtered for with at least one modified peptides being detected in any of the 6 cell types. Because all proteins within the glycolytic pathways were tested for cell-type preference for each modification, Benjamini-Hochberg method was applied to address for multiple comparison and the final adjusted p values were obtained.

11.21 Structural visualisation and analysis

Protein structures were assembled using software MacPymol (version v1.7.4.5; Schrodinger, LLC). Crystal structures and NMR structures of the hnRNP proteins were downloaded from PDB database.

11.22 Imputation of missing values in modified proteins

Modified proteins from acetylation, phosphorylation, methylation and deamidation were used. Use of deamidation is to facilitate more accurate normalization of samples. The protein abundance values for all these proteins are normalized in each sample (as described above) and used for imputation using the R package missForest. The maximum iteration was set to 100 and the final imputation value was taken after convergence of the algorithm.

11.23 Clustering of modified proteins

The full protein abundance values for methylated, acetylated and phosphorylated proteins, containing imputed values (from above) was used to calculate pairwise spearman correlation. Hierachical clustering using the ward.D method was performed on the pairwise spearman correlation matrix, and clusters were defined by using the cutree function in R, defining the number of clusters.

11.24 Partial correlation of proteins abundance with age

Abundance of modified and unmodified proteins were corrected for pool effects using a linear mixed model that includes age as fixed effect and pool as random effect. Spearman partial correlation between age and modified protein abundance was calculated, while controlling for the influence of unmodified protein abundance.

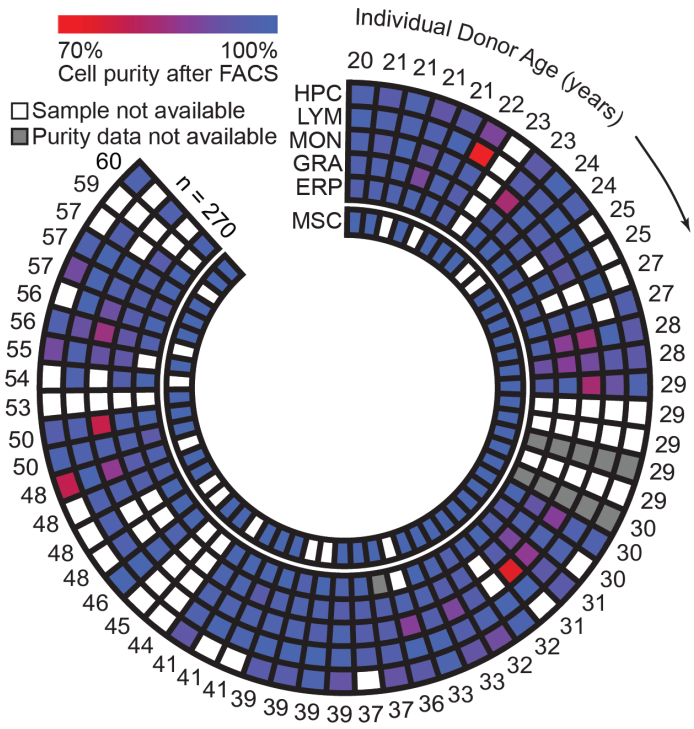
12 Supplementary

Supplementary Table 1: Top 10 principle components most correlated with age

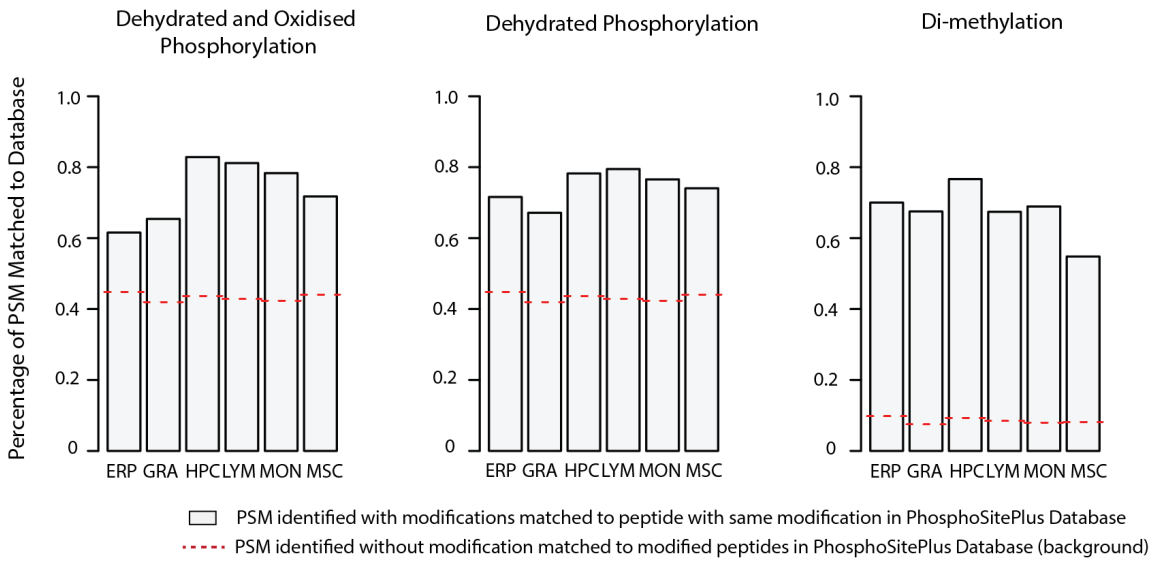
PC number	Spearman correlation coefficient	p value
PC34	-0.200	0.0012
PC120	0.177	0.0044
PC149	0.150	0.0158
PC61	-0.150	0.0160
PC92	-0.147	0.0176
PC54	0.145	0.0192
PC53	0.139	0.0249
PC167	0.138	0.0268
PC44	0.133	0.0330
PC102	-0.123	0.0488

Supplementary Table 2: Differences in partial correlation coefficients between clusters across six cell types

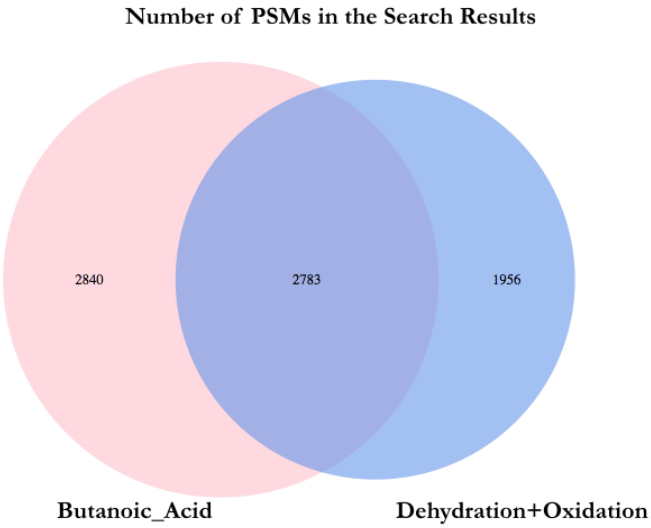
cluster	median	p	celltype
1	−0.111771200	0.618	ERP
2	−0.081393717	0.726	ERP
3	0.068258948	0.610	ERP
4	0.192844181	0.267	ERP
5	0.031916208	0.595	ERP
6	0.052353893	0.946	ERP
7	0.031910986	0.954	ERP
1	−0.061918222	0.143	GRA
2	−0.067479500	0.366	GRA
3	−0.062709653	0.763	GRA
4	−0.028762172	0.785	GRA
5	−0.004601149	0.753	GRA
6	0.067677212	0.091	GRA
7	0.219946629	0.140	GRA
1	−0.133254734	0.022	HSC
2	−0.204872322	0.001	HSC
3	−0.107067116	0.019	HSC
4	0.035558604	0.008	HSC
5	−0.117676208	0.019	HSC
6	−0.023224655	0.020	HSC
7	0.159136407	0.007	HSC
1	−0.101292302	0.374	LYM
2	−0.031576858	0.482	LYM
3	−0.154492129	0.406	LYM
4	−0.107226851	0.265	LYM
5	0.014899795	0.375	LYM
6	−0.015814440	0.475	LYM
7	0.108209754	0.036	LYM
1	−0.024852759	0.131	MON
2	0.042325546	0.273	MON
3	0.042886819	0.202	MON
4	0.100930354	0.104	MON
5	0.013495843	0.203	MON
6	−0.015656193	0.128	MON
7	−0.009411443	0.204	MON
1	0.016692131	0.567	MSC
2	−0.013761592	0.251	MSC
3	0.063817567	0.278	MSC
4	0.045833942	0.668	MSC
5	0.019426921	0.605	MSC
6	0.029746308	0.646	MSC
7	−0.014316932	0.371	MSC



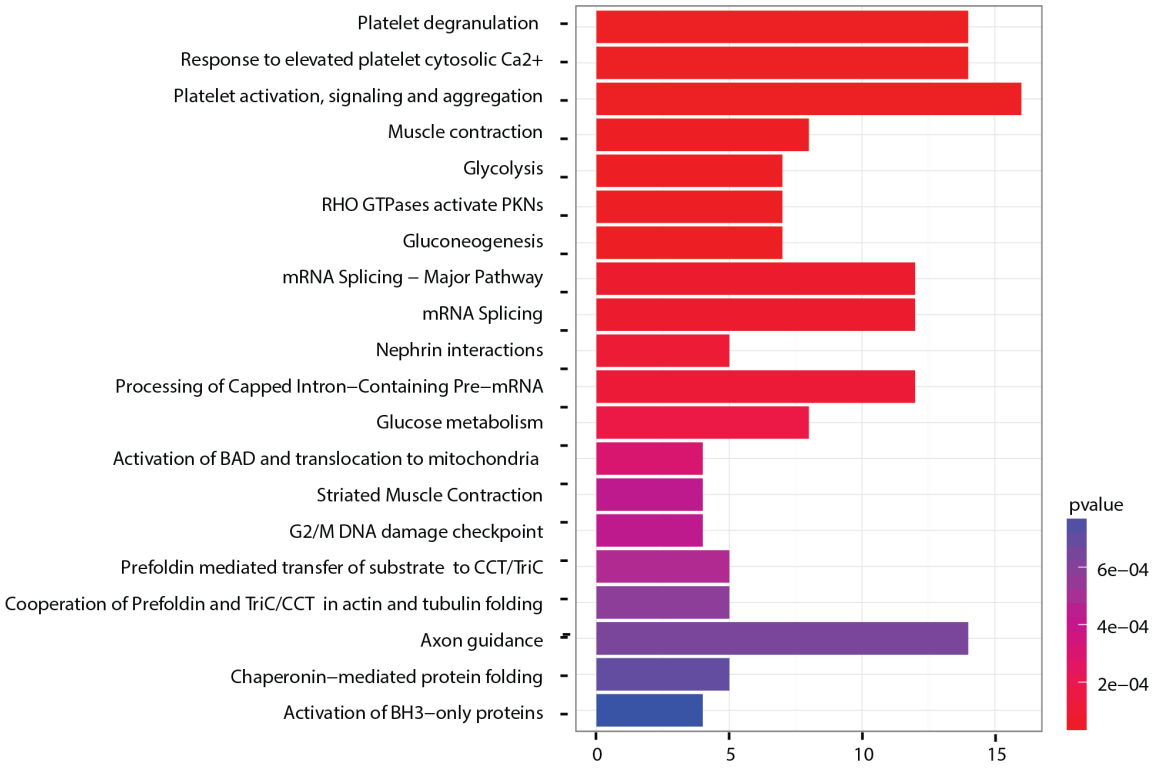
Supplementary Figure 1: Sample purify of fluorescence-activated cell sorting. Purity of all samples separated through FACS were indicated with colours from red (70%) to blue (100%). (this figure is from my collaborator Ximing Ding)



Supplementary Figure 2: Validation of modified peptide with protein modification database. Dehydrated and oixidised phophorylated peptides, dehydrated phosphorylated peptides and di-methylated peptides were compared to peptide entires annotated in curated protein phorsphorylation and methylation database. The percentages of peptides matching to a database entry were 67% to 79%, 62% to 83% and 55% to 77%, respectively. The peptide matching rates were at least 1.5 higher than the background (indicated by the red dashed lines).

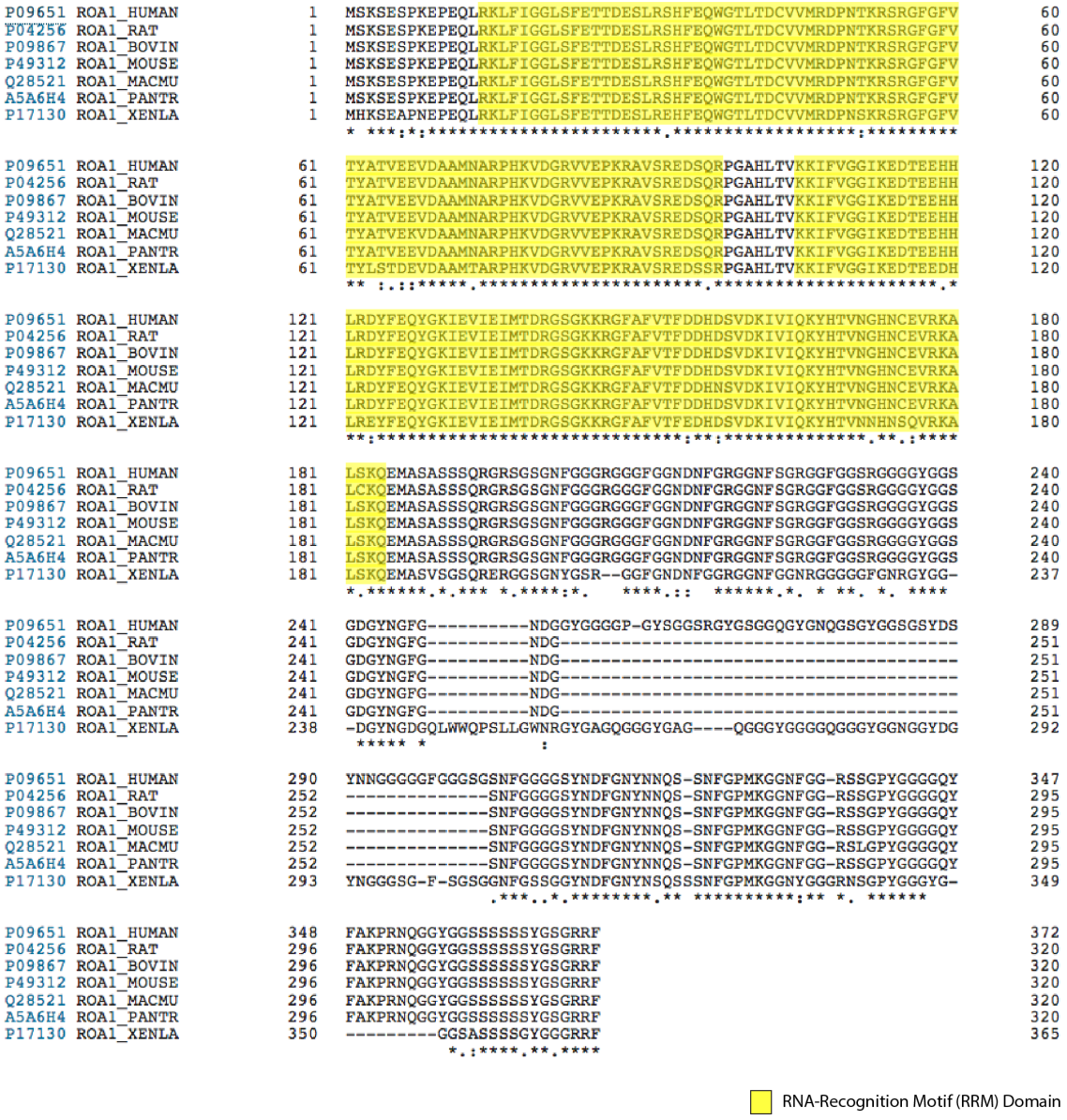


Supplementary Figure 3: Comparison between spectrum identified with 2-amino-3-oxo-butanoic acid modification and with oxidised and dehydrated phosphorylation. For data from one MS fraction, with standard database search include 2-amino-3-oxo-butanoic acid as dynamic modification parameter, one obtained 5623 PSM. For the same result file, with standard database search include dehydation and oxidation as dynamic modification parameter, one obtained 4739 PSM. 2783 spectrum had shared identifications between the two modification.

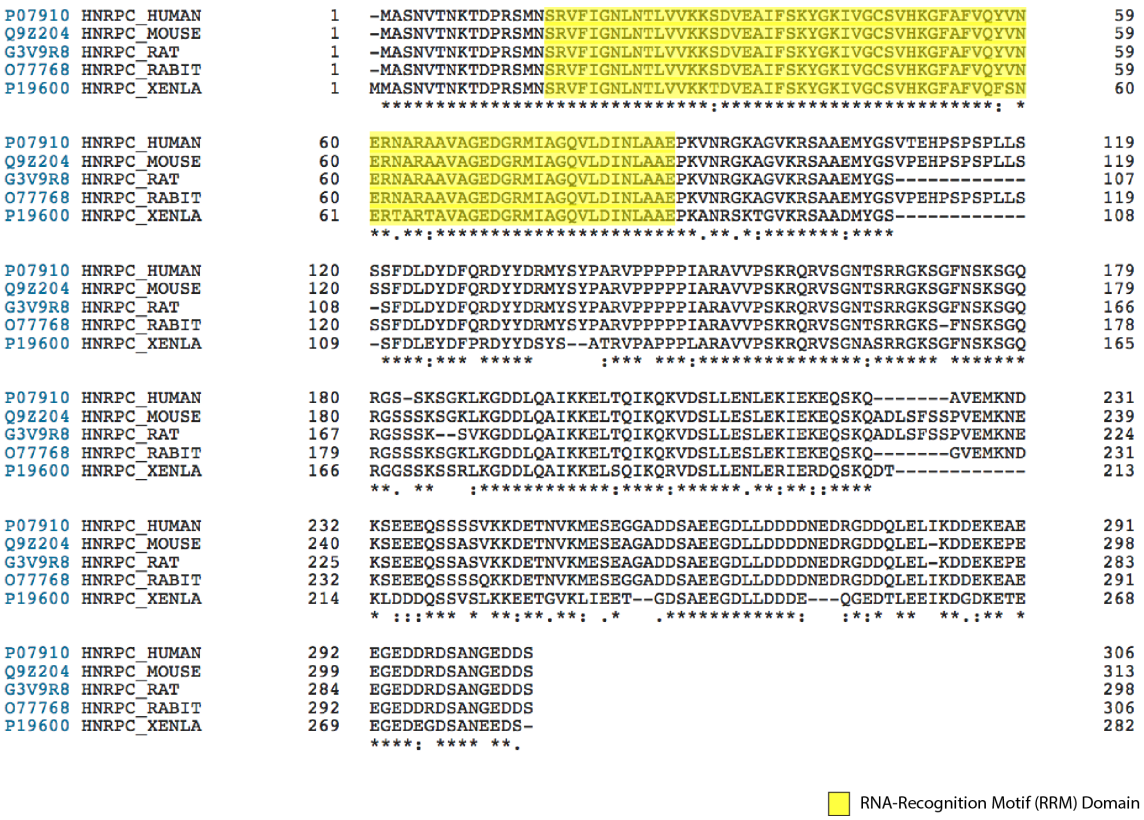


Supplementary Figure 4: Proteins co-regulated by phosphorylation, methylation and acetylation across six cell types were enriched in various pathways in human bone marrow.

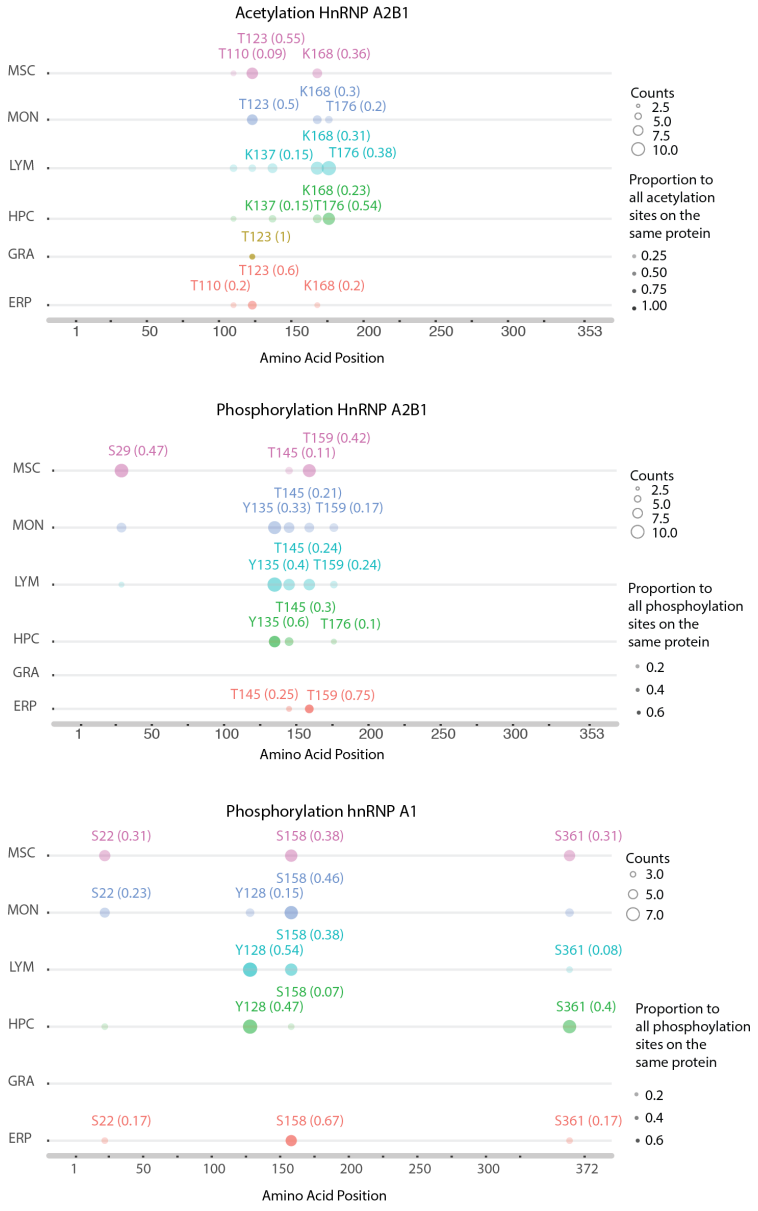
Supplementary Figure 5: Sequences of hnRNP A2B1 are highly conserved across multiple species. Species analysed include human (P22626), rat (A7VJC2), mouse (O88569), bovine (Q2HJ60), cotton-top tamarin (Q9TTV2), and sumatran orangutan (Q5RBU8).



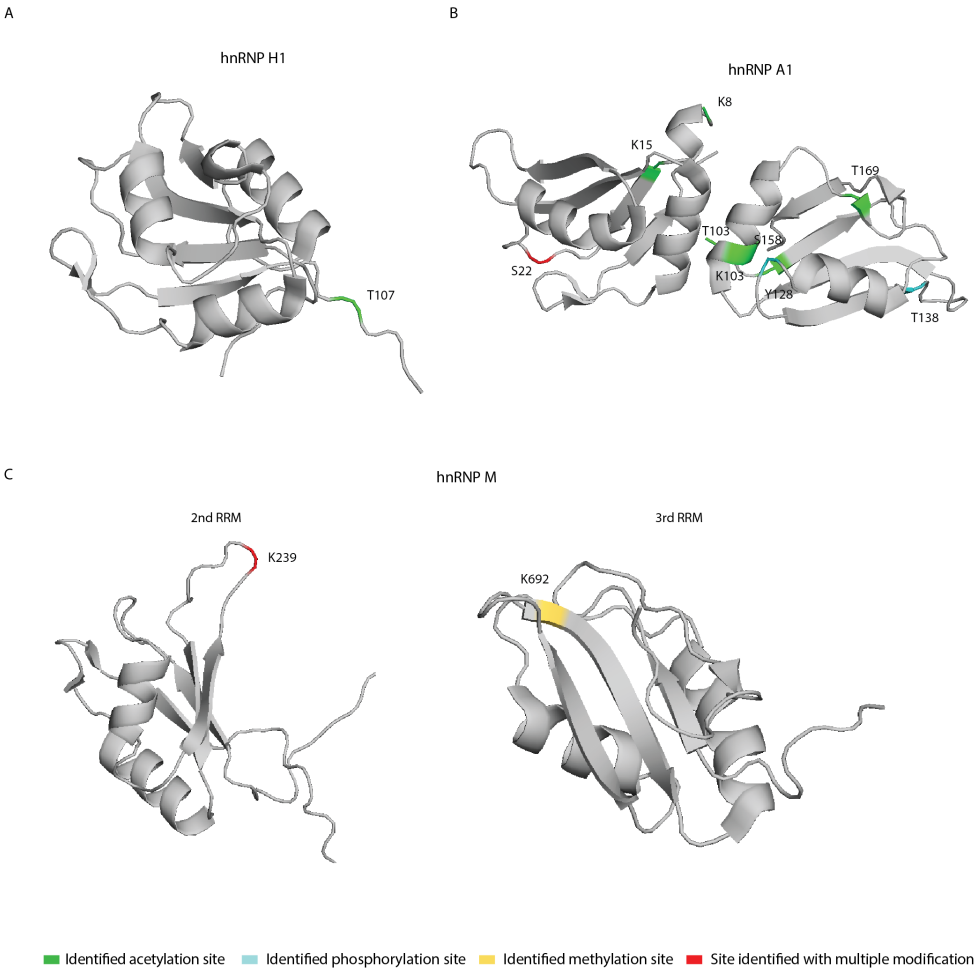
Supplementary Figure 6: RRM domain of hnRNP A1 are conserved across multiple species. Species analyzed include human (P04256), chimpanzee (A5A6H4), rhesus macaque (Q28521), mouse(P49312), rat (P04256), bovine (P09867), and african clawed frog (P17130).



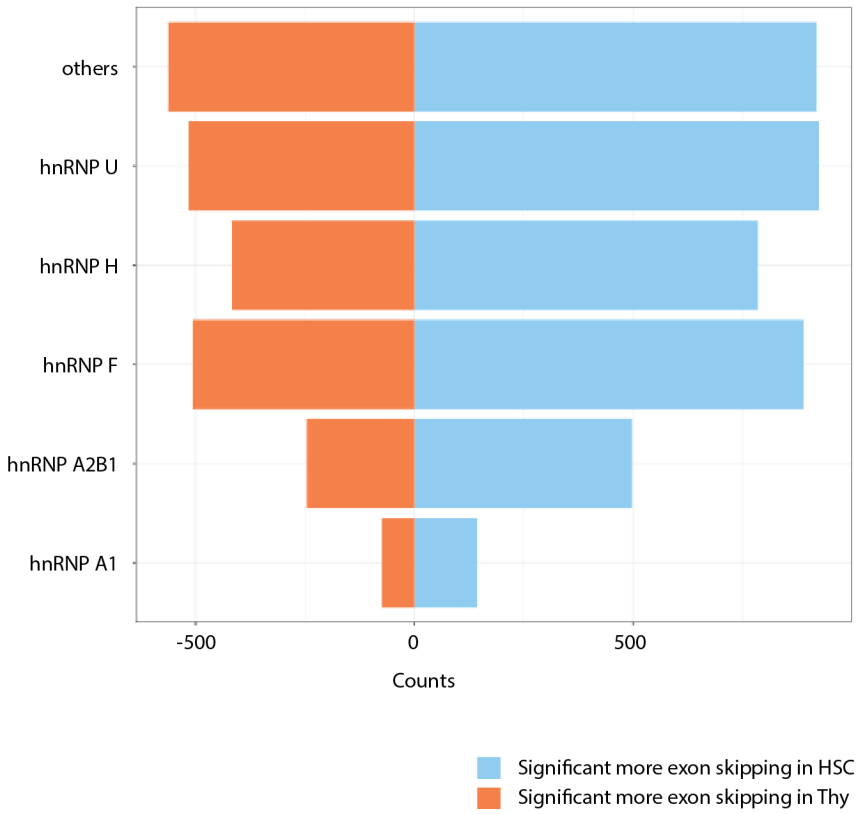
Supplementary Figure 7: RRM domain of hnRNP C are conserved across multiple species. Species analyzed include human (P07910), mouse (Q9Z204), rat (G3V9R8), chimpanzee (A5A6H4), rhesus macaque (Q28521), rabbit (O77768), and african clawed frog (P19600).



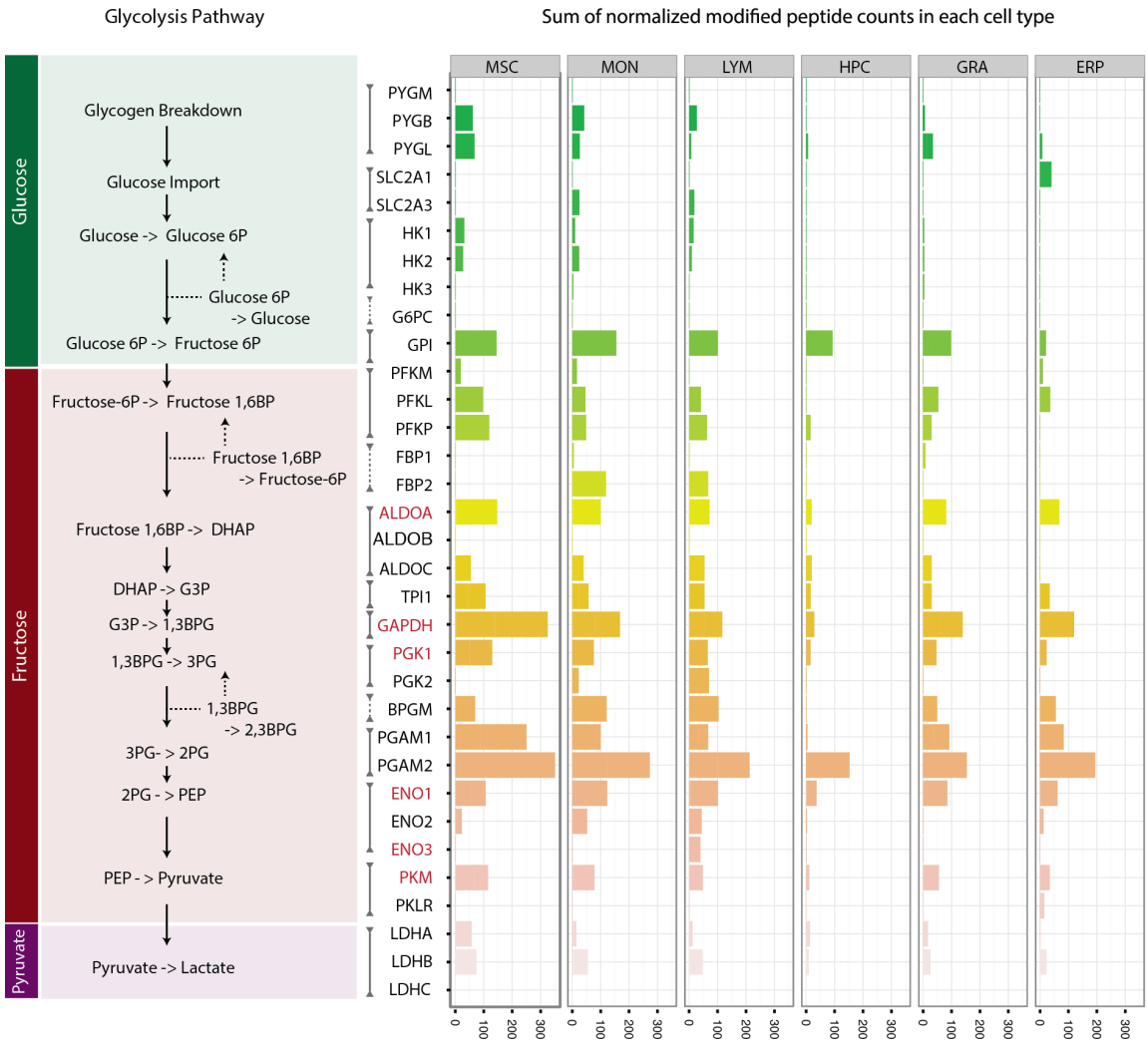
Supplementary Figure 8: hnRNP A2B1 and hnRNPA1 showed cell-type dependent preferences for phosphorylation and acetylation sites across different bone marrow cells. Dominant modification sites were different across different cell types. For acetylation site on hnRNP A2B1, T123 was the most dominantly acetylated site in MSCs, MONs, GRAs and ERPs, whereas T176 is more frequently acetylated in LYMs and HPCs. S29 was most frequently phosphorylated on hnRNP A2B1 in MSCs where as in MONs, LYMs, HPCs Y135 is the preferred phosphorylation site. For hnRNP A1, S158 was the preferred phosphorylation site in MSCs, MONs and ERPs, whereas Y1228 was most frequently detected in HPC and LYMs.



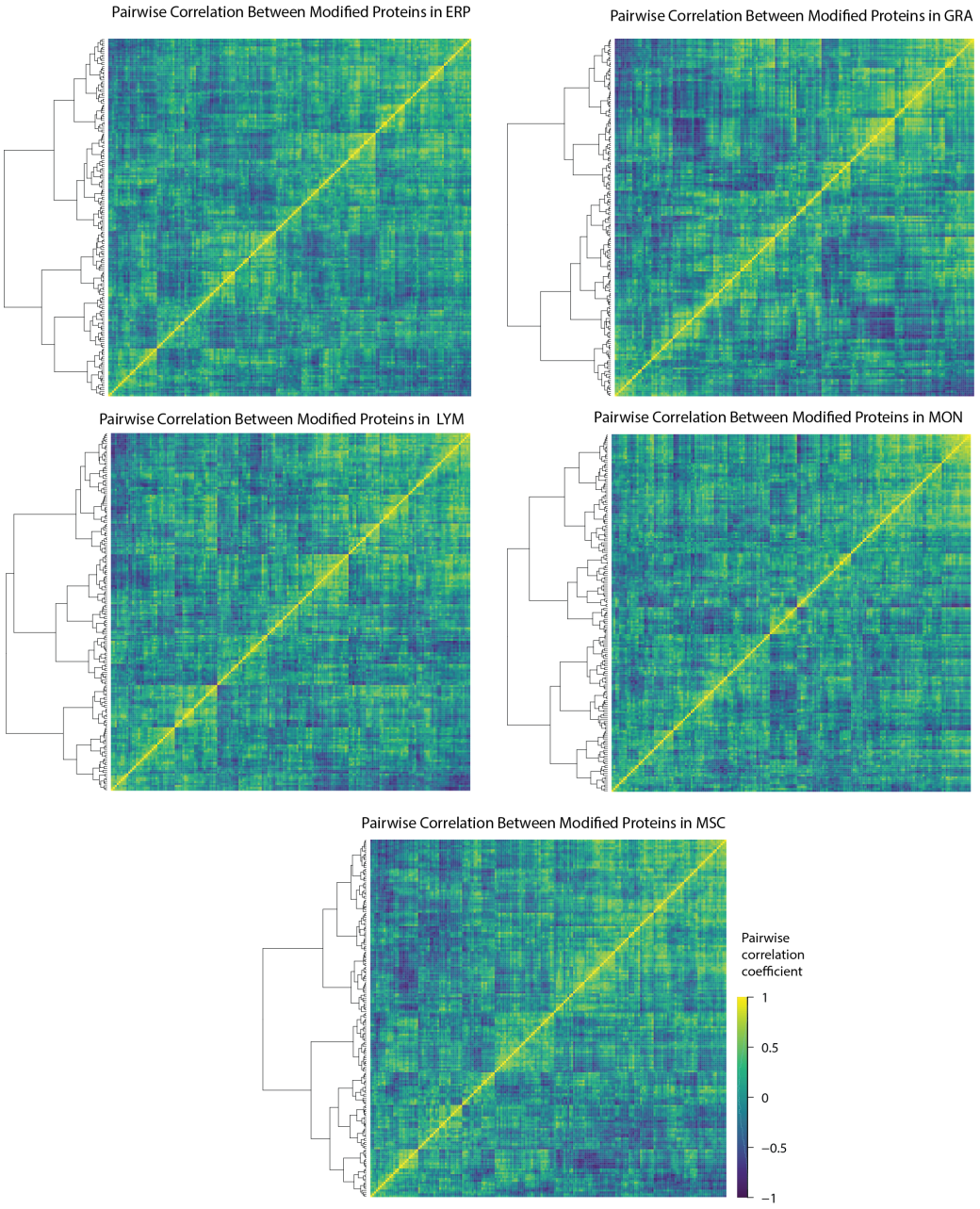
Supplementary Figure 9: Structures of RRM domains of hnRNP H1, hnRNP A1 and hnRNP M. NMR structure of N-terminal RRM domain of hnRNP H1(PDB: 2lxu), X-ray crystal structure of RRM domains of hnRNP A1(PDB:1l3k) as well as NMR structure of the second and third RRM domain of hnRNP M (PDB:2do0) with detected modification sites were shown. Acetylation site was coloured in green, phosphorylation in cyan, methylation in orange. Site with multiple modification detected was shown in red, which were S22 on hnRNP A1 for phosphorylation and acetylation, and K239 in hnRNP M for acetylation and methylation.



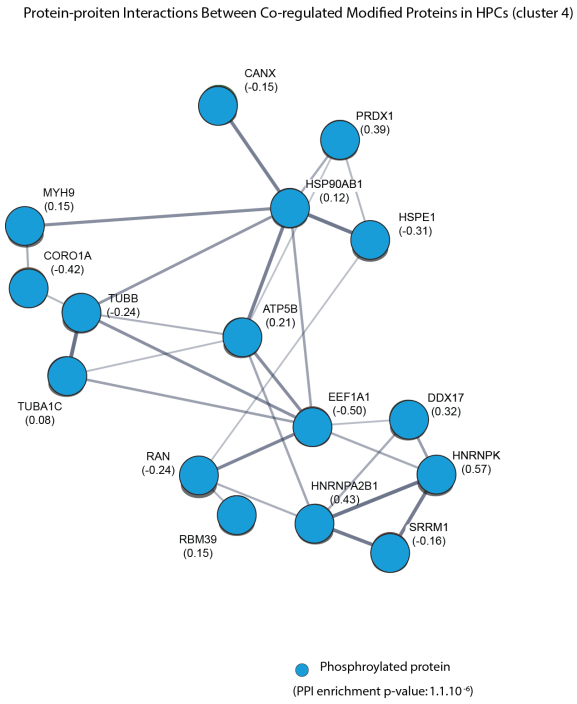
Supplementary Figure 10: Alternative splicing events of hnRNP protein-associated transcripts. Transcripts associated with hnRNP A1 and hnRNP A1B1 in human 293T cells were obtained from [Huelga et al., 2012]. RNA expressed in CD 4+ and HSC cells were obtained from [Casero et al., 2015].The targets of hnRNP A1 and hnRNP A2B1 in CD 4+ T cells showed a bias of exon skipping in HSC.



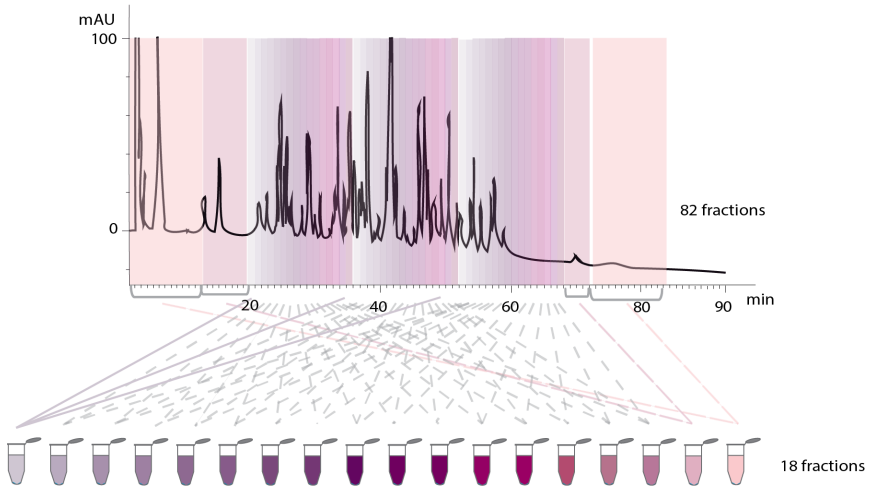
Supplementary Figure 11: Proteins regulating fructose metabolism within the glycolysis pathway were more frequently modified. For each cell type, all proteins associated with glycolytic pathway was analysed for their modification frequency. All modifications (including oxidation and deamidation) were taken into consideration. Assuming that for each modification the corresponding modification residues on each protein within the pathway have the same probability of being modified, the number of the modified peptides were normalised by the total number of peptides detected for a given protein and the number of residues that can be modified for the protein. From the normalized modified peptide counts, it can be seen that for each cell types, protein modifications were more frequent in fructose metabolism within the glycolysis pathway. The glycolytic pathway was illustrated on the left panel, and the normalized modified peptide counts were shown on the right panel, with differentially phosphorylated and acetylated proteins highlighted in red.



Supplementary Figure 12: Pairwise-correlation between modified proteins in different bone marrow cells. The correlations were calculated based on PTM scores. Modified proteins in ERPs, LYMs, MONs, GRAs and MSCs co-evolved and can be further clustered into different groups.



Supplementary Figure 13: Protein-protein interaction between modified proteins in haematopoietic progenitor cells. Proteins in cluster 4 in HPCs from clustering analysis were mostly mediated by phosphorylation. The relationship between phosphorylation level and age for these proteins were heterogeneous as 9 were positively related with age and 7 were negatively related with age.



Supplementary Figure 14: Fractionation scheme for orthogonal liquid-chromatography separation. Samples were labelled with TMT 6plex and combined. The concentrated peptide mixture was fractionated using reversed-phase liquid chromatography to 90 fractions. Fraction 1-82 were further combined according to the scheme into 18 fractions.

References

[Aebersold and Mann, 2016] Aebersold, R. and Mann, M. (2016). Mass-spectrometric exploration of proteome structure and function. *Nature*, 537(7620):347–355.

[Akue et al., 2012] Akue, A. D., Lee, J.-Y., and Jameson, S. C. (2012). Derivation and maintenance of virtual memory cd8 t cells. *The Journal of Immunology*, 188(6):2516–2523.

[Alberts, 2008] Alberts, B. (2008). *Molecular Biology of the Cell: Reference edition*. Number Bd. 1 in Molecular Biology of the Cell: Reference Edition. Garland Science.

[Aletta et al., 1998] Aletta, J. M., Cimato, T. R., and Ettinger, M. J. (1998). Protein methylation: a signal event in post-translational modification. *Trends in biochemical sciences*, 23(3):89–91.

[Andersson et al., 2014] Andersson, S., Nilsson, K., Fagerberg, L., Hallström, B. M., Sundström, C., Danielsson, A., Edlund, K., Uhlen, M., and Asplund, A. (2014). The transcriptomic and proteomic landscapes of bone marrow and secondary lymphoid tissues. *PloS one*, 9(12):e115911.

[Apic et al., 2001] Apic, G., Gough, J., and Teichmann, S. A. (2001). An insight into domain combinations. *Bioinformatics*, 17(suppl 1):S83–S89.

[Avalos et al., 2005] Avalos, J. L., Bever, K. M., and Wolberger, C. (2005). Mechanism of sirtuin inhibition by nicotinamide: altering the nad+ cosubstrate specificity of a sir2 enzyme. *Molecular cell*, 17(6):855–868.

[Bailey et al., 2009] Bailey, T. L., Boden, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., Ren, J., Li, W. W., and Noble, W. S. (2009). Meme suite: tools for motif discovery and searching. *Nucleic acids research*, page gkp335.

[Banasavadi-Siddegowda et al., 2016] Banasavadi-Siddegowda, Y., Russell, L., Frair, E., Karkhanis, V., Relation, T., Yoo, J., Zhang, J., Sif, S., Imitola, J., Baiocchi, R., et al. (2016). Prmt5–pten molecular pathway regulates senescence and self-renewal of primary glioblastoma neurosphere cells. *Oncogene*.

[Bannister and Kouzarides, 2011] Bannister, A. J. and Kouzarides, T. (2011). Regulation of chromatin by histone modifications. *Cell research*, 21(3):381–395.

[Barford et al., 1998] Barford, D., Das, A. K., and Egloff, M.-P. (1998). The structure and mechanism of protein phosphatases: insights into catalysis and regulation. *Annual review of biophysics and biomolecular structure*, 27(1):133–164.

[Beausoleil et al., 2004] Beausoleil, S. A., Jedrychowski, M., Schwartz, D., Elias, J. E., Villén, J., Li, J., Cohn, M. A., Cantley, L. C., and Gygi, S. P. (2004). Large-scale characterization of hela cell nuclear phosphoproteins. *Proceedings of the National Academy of Sciences of the United States of America*, 101(33):12130–12135.

[Beltrao et al., 2012] Beltrao, P., Albanèse, V., Kenner, L. R., Swaney, D. L., Burlingame, A., Villén, J., Lim, W. A., Fraser, J. S., Frydman, J., and Krogan, N. J. (2012). Systematic functional prioritization of protein posttranslational modifications. *Cell*, 150(2):413–425.

[Berndsen and Denu, 2008] Berndsen, C. E. and Denu, J. M. (2008). Catalysis and substrate selection by histone/protein lysine acetyltransferases. *Current opinion in structural biology*, 18(6):682–689.

[Bian et al., 2014] Bian, Y., Song, C., Cheng, K., Dong, M., Wang, F., Huang, J., Sun, D., Wang, L., Ye, M., and Zou, H. (2014). An enzyme assisted rp-rplc approach for in-depth analysis of human liver phosphoproteome. *Journal of proteomics*, 96:253–262.

[Biggar and Li, 2015] Biggar, K. K. and Li, S. S.-C. (2015). Non-histone protein methylation as a regulator of cellular signalling and function. *Nature Reviews Molecular Cell Biology*, 16(1):5–17.

[Bitterman et al., 2002] Bitterman, K. J., Anderson, R. M., Cohen, H. Y., Latorre-Esteves, M., and Sinclair, D. A. (2002). Inhibition of silencing and accelerated aging by nicotinamide, a putative negative regulator of yeast sir2 and human sirt1. *Journal of Biological Chemistry*, 277(47):45099–45107.

[Blanc and Richard, 2017] Blanc, R. S. and Richard, S. (2017). Arginine methylation: The coming of age. *Molecular Cell*, 65(1):8–24.

[Blanc et al., 2016] Blanc, R. S., Vogel, G., Chen, T., Crist, C., and Richard, S. (2016). Prmt7 preserves satellite cell regenerative capacity. *Cell reports*, 14(6):1528–1539.

[Brooks and Gu, 2003] Brooks, C. L. and Gu, W. (2003). Ubiquitination, phosphorylation and acetylation: the molecular basis for p53 regulation. *Current opinion in cell biology*, 15(2):164–171.

[Carling et al., 2011] Carling, D., Mayer, F. V., Sanders, M. J., and Gamblin, S. J. (2011). Amp-activated protein kinase: nature's energy sensor. *Nature chemical biology*, 7(8):512–518.

[Carrascal et al., 2008] Carrascal, M., Ovelleiro, D., Casas, V., Gay, M., and Abian, J. (2008). Phosphorylation analysis of primary human t lymphocytes using sequential imac and titanium oxide enrichment. *Journal of proteome research*, 7(12):5167–5176.

[Casero et al., 2015] Casero, D., Sandoval, S., Seet, C. S., Scholes, J., Zhu, Y., Ha, V. L., Luong, A., Parekh, C., and Crooks, G. M. (2015). Long non-coding rna profiling of human lymphoid progenitor cells reveals transcriptional divergence of b cell and t cell lineages. *Nature immunology*.

[Cha et al., 2017] Cha, Y., Han, M.-J., Cha, H.-J., Zoldan, J., Burkart, A., Jung, J. H., Jang, Y., Kim, C.-H., Jeong, H.-C., Kim, B.-G., et al. (2017). Metabolic control of primed human pluripotent stem cell fate and function by the mir-200c-sirt2 axis. *Nature Cell Biology*, 19(5):445–456.

[Chapman et al., 2014] Chapman, J. D., Goodlett, D. R., and Masselon, C. D. (2014). Multiplexed and data-independent tandem mass spectrometry for global proteome profiling. *Mass spectrometry reviews*, 33(6):452–470.

[Chaudhury et al., 2010] Chaudhury, A., Chander, P., and Howe, P. H. (2010). Heterogeneous nuclear ribonucleoproteins (hnrnps) in cellular processes: Focus on hnrnp e1's multifunctional regulatory roles. *Rna*, 16(8):1449–1462.

[Chen et al., 2017] Chen, M. J., Dixon, J. E., and Manning, G. (2017). Genomics and evolution of protein phosphatases. *Sci. Signal.*, 10(474):eaag1796.

[Chen et al., 2006] Chen, X., Armstrong, M. A., and Li, G. (2006). Mesenchymal stem cells in immunoregulation. *Immunology and cell biology*, 84(5):413–421.

[Chen et al., 2010] Chen, X., Tang, W., Liu, S., Yu, L., and Chen, Z. (2010). Thioredoxin-1 phosphorylated at t100 is needed for its anti-apoptotic activity in hepg2 cancer cells. *Life sciences*, 87(7):254–260.

[Chick et al., 2015] Chick, J. M., Kolippakkam, D., Nusinow, D. P., Zhai, B., Rad, R., Huttlin, E. L., and Gygi, S. P. (2015). A mass-tolerant database search identifies a large proportion of unassigned spectra in shotgun proteomics as modified peptides. *Nature biotechnology*, 33(7):743–749.

[Childs et al., 2015] Childs, B. G., Durik, M., Baker, D. J., and Van Deursen, J. M. (2015). Cellular senescence in aging and age-related disease: from mechanisms to therapy. *Nature medicine*, 21(12):1424–1435.

[Cho et al., 2008] Cho, R. H., Sieburg, H. B., and Muller-Sieburg, C. E. (2008). A new mechanism for the aging of hematopoietic stem cells: aging changes the clonal composition of the stem cell compartment but not individual stem cells. *Blood*, 111(12):5553–5561.

[Choudhary et al., 2009] Choudhary, C., Kumar, C., Gnad, F., Nielsen, M. L., Rehman, M., Walther, T. C., Olsen, J. V., and Mann, M. (2009). Lysine acetylation targets protein complexes and co-regulates major cellular functions. *Science*, 325(5942):834–840.

[Choudhary and Mann, 2010] Choudhary, C. and Mann, M. (2010). Decoding signalling networks by mass spectrometry-based proteomics. *Nature reviews Molecular cell biology*, 11(6):427–439.

[Chung et al., 2009] Chung, H. Y., Cesari, M., Anton, S., Marzetti, E., Giovannini, S., Seo, A. Y., Carter, C., Yu, B. P., and Leeuwenburgh, C. (2009). Molecular inflammation: underpinnings of aging and age-related diseases. *Ageing research reviews*, 8(1):18–30.

[Comb et al., 1966] Comb, D. G., Sarkar, N., and Pinzino, C. J. (1966). The methylation of lysine residues in protein. *Journal of Biological Chemistry*, 241(8):1857–1862.

[Coutu et al., 2011] Coutu, D. L., François, M., and Galipeau, J. (2011). Inhibition of cellular senescence by developmentally regulated fgf receptors in mesenchymal stem cells. *Blood*, 117(25):6801–6812.

[Debler et al., 2016] Debler, E. W., Jain, K., Warmack, R. A., Feng, Y., Clarke, S. G., Blobel, G., and Stavropoulos, P. (2016). A glutamate/aspartate switch controls product specificity in a protein arginine methyltransferase. *Proceedings of the National Academy of Sciences*, 113(8):2068–2073.

[Devine et al., 2003] Devine, S. M., Cobbs, C., Jennings, M., Bartholomew, A., and Hoffman, R. (2003). Mesenchymal stem cells distribute to a wide range of tissues following systemic infusion into nonhuman primates. *Blood*, 101(8):2999–3001.

[Dorshkind et al., 2009] Dorshkind, K., Montecino-Rodriguez, E., and Signer, R. A. (2009). The ageing immune system: is it ever too old to become young again? *Nature Reviews Immunology*, 9(1):57–62.

[Drazic et al., 2016] Drazic, A., Myklebust, L. M., Ree, R., and Arnesen, T. (2016). The world of protein acetylation. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics*, 1864(10):1372–1401.

[Dreyfuss et al., 1988] Dreyfuss, G., Philipson, L., and Mattaj, I. W. (1988). Ribonucleoprotein particles in cellular processes. *The Journal of cell biology*, 106(5):1419–1425.

[Duan and Walther, 2015] Duan, G. and Walther, D. (2015). The roles of post-translational modifications in the context of protein interaction networks. *PLoS Comput Biol*, 11(2):e1004049.

[Dyson and Wright, 2005] Dyson, H. J. and Wright, P. E. (2005). Intrinsically unstructured proteins and their functions. *Nature reviews Molecular cell biology*, 6(3):197–208.

[Eisenberg et al., 2009] Eisenberg, T., Knauer, H., Schauer, A., Büttner, S., Ruckenstuhl, C., Carmona-Gutierrez, D., Ring, J., Schroeder, S., Magnes, C., Antonacci, L., et al. (2009). Induction of autophagy by spermidine promotes longevity. *Nature cell biology*, 11(11):1305–1314.

[Eisenberg et al., 2014] Eisenberg, T., Schroeder, S., Andryushkova, A., Pendl, T., Küttner, V., Bhukel, A., Mariño, G., Pietrocola, F., Harger, A., Zimmermann, A., et al. (2014). Nucleocytosolic depletion of the energy metabolite acetyl-coenzyme a stimulates autophagy and prolongs lifespan. *Cell metabolism*, 19(3):431–444.

[Ermak and Davies, 2002] Ermak, G. and Davies, K. J. (2002). Calcium and oxidative stress: from cell signaling to cell death. *Molecular immunology*, 38(10):713–721.

[Estève et al., 2011] Estève, P.-O., Chang, Y., Samaranayake, M., Upadhyay, A. K., Horton, J. R., Feehery, G. R., Cheng, X., and Pradhan, S. (2011). A methylation and phosphorylation switch between an adjacent lysine and serine determines human dnmt1 stability. *Nature structural & molecular biology*, 18(1):42–48.

[Fabrizio et al., 2001] Fabrizio, P., Pozza, F., Pletcher, S. D., Gendron, C. M., and Longo, V. D. (2001). Regulation of longevity and stress resistance by sch9 in yeast. *Science*, 292(5515):288–290.

[Fang et al., 2014] Fang, L., Zhang, L., Wei, W., Jin, X., Wang, P., Tong, Y., Li, J., Du, J. X., and Wong, J. (2014). A methylation-phosphorylation switch determines sox2 stability and function in esc maintenance or differentiation. *Molecular cell*, 55(4):537–551.

[Feng et al., 2002] Feng, Q., Wang, H., Ng, H. H., Erdjument-Bromage, H., Tempst, P., Struhl, K., and Zhang, Y. (2002). Methylation of h3-lysine 79 is mediated by a new family of hmtases without a set domain. *Current Biology*, 12(12):1052–1058.

[Finkel et al., 2009] Finkel, T., Deng, C.-X., and Mostoslavsky, R. (2009). Recent progress in the biology and physiology of sirtuins. *Nature*, 460(7255):587–591.

[Folmes et al., 2011] Folmes, C. D., Nelson, T. J., Martinez-Fernandez, A., Arrell, D. K., Lindor, J. Z., Dzeja, P. P., Ikeda, Y., Perez-Terzic, C., and Terzic, A. (2011). Somatic oxidative bioenergetics transitions into pluripotency-dependent glycolysis to facilitate nuclear reprogramming. *Cell metabolism*, 14(2):264–271.

[Fontana et al., 2010] Fontana, L., Partridge, L., and Longo, V. D. (2010). Extending healthy life span?from yeast to humans. *science*, 328(5976):321–326.

[Forsgard et al., 2010] Forsgard, N., Salehpour, M., and Possnert, G. (2010). Accelerator mass spectrometry in the attomolar concentration range for 14 c-labeled biologically active compounds in complex matrixes. *Journal of Analytical Atomic Spectrometry*, 25(1):74–78.

[Franceschi et al., 2007] Franceschi, C., Capri, M., Monti, D., Giunta, S., Olivieri, F., Sevini, F., Panourgia, M. P., Invidia, L., Celani, L., Scurti, M., et al. (2007). Inflammaging and anti-inflammaging: a systemic perspective on aging and longevity emerged from studies in humans. *Mechanisms of ageing and development*, 128(1):92–105.

[Füllgrabe et al., 2013] Füllgrabe, J., Lynch-Day, M. A., Heldring, N., Li, W., Struijk, R. B., Ma, Q., Hermanson, O., Rosenfeld, M. G., Klionsky, D. J., and Joseph, B. (2013). The histone h4 lysine 16 acetyltransferase hmof regulates the outcome of autophagy. *Nature*, 500(7463):468–471.

[Fülöp et al., 2013] Fülöp, T., Larbi, A., and Pawelec, G. (2013). Humant cell aging and the impact of persistent viral infections.

[Gao et al., 2016] Gao, Y., Zhao, Y., Zhang, J., Lu, Y., Liu, X., Geng, P., Huang, B., Zhang, Y., and Lu, J. (2016). The dual function of prmt1 in modulating epithelial-mesenchymal transition and cellular senescence in breast cancer cells through regulation of zeb1. *Scientific reports*, 6.

[Gao G, 2017] Gao G, Dhar S, B. M. (2017). Prmt5 regulates ires-dependent translation via methylation of hnrnp a1. *Nucleic Acids Research*, 45(8):4359–4369.

[Geiger et al., 2013] Geiger, H., De Haan, G., and Florian, M. C. (2013). The ageing haematopoietic stem cell compartment. *Nature Reviews Immunology*, 13(5):376–389.

[Gelino and Hansen, 2012] Gelino, S. and Hansen, M. (2012). Autophagy-an emerging anti-aging mechanism. *Journal of clinical & experimental pathology*.

[Ghannam et al., 2010] Ghannam, S., Bouffi, C., Djouad, F., Jorgensen, C., and Noël, D. (2010). Immunosuppression by mesenchymal stem cells: mechanisms and clinical applications. *Stem cell research & therapy*, 1(1):2.

[Glisovic et al., 2008] Glisovic, T., Bachorik, J. L., Yong, J., and Dreyfuss, G. (2008). Rna-binding proteins and post-transcriptional gene regulation. *FEBS letters*, 582(14):1977–1986.

[Görlach et al., 1992] Görlach, M., Wittekind, M., Beckman, R., Mueller, L., and Dreyfuss, G. (1992). Interaction of the rna-binding domain of the hnrnp c proteins with rna. *The EMBO journal*, 11(9):3289.

[Goronzy and Weyand, 2013] Goronzy, J. J. and Weyand, C. M. (2013). Understanding immunosenescence to improve responses to vaccines. *Nature immunology*, 14(5):428–436.

[Greer et al., 2011] Greer, E. L., Maures, T. J., Ucar, D., Hauswirth, A. G., Mancini, E., Lim, J. P., Benayoun, B. A., Shi, Y., and Brunet, A. (2011). Transgenerational epigenetic inheritance of longevity in caenorhabditis elegans. *Nature*, 479(7373):365–371.

[Gwinn et al., 2008] Gwinn, D. M., Shackelford, D. B., Egan, D. F., Mihaylova, M. M., Mery, A., Vasquez, D. S., Turk, B. E., and Shaw, R. J. (2008). Ampk phosphorylation of raptor mediates a metabolic checkpoint. *Molecular cell*, 30(2):214–226.

[Gygi et al., 1999] Gygi, S. P., Rist, B., Gerber, S. A., Turecek, F., Gelb, M. H., and Aebersold, R. (1999). Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nature biotechnology*, 17(10):994–999.

[Haberland et al., 2009] Haberland, M., Montgomery, R. L., and Olson, E. N. (2009). The many roles of histone deacetylases in development and physiology: implications for disease and therapy. *Nature Reviews Genetics*, 10(1):32–42.

[Hait et al., 2009] Hait, N. C., Allegood, J., Maceyka, M., Strub, G. M., Harikumar, K. B., Singh, S. K., Luo, C., Marmorstein, R., Kordula, T., Milstien, S., et al. (2009). Regulation of histone acetylation in the nucleus by sphingosine-1-phosphate. *Science*, 325(5945):1254–1257.

[Hamamoto et al., 2004] Hamamoto, R., Furukawa, Y., Morita, M., Iimura, Y., Silva, F. P., Li, M., Yagyu, R., and Nakamura, Y. (2004). Smyd3 encodes a histone methyltransferase involved in the proliferation of cancer cells. *Nature cell biology*, 6(8):731–740.

[Hamamoto et al., 2015] Hamamoto, R., Saloura, V., and Nakamura, Y. (2015). Critical roles of non-histone protein lysine methylation in human tumorigenesis. *Nature Reviews Cancer*, 15(2):110–124.

[Hanschmann et al., 2013] Hanschmann, E.-M., Godoy, J. R., Berndt, C., Hudemann, C., and Lillig, C. H. (2013). Thioredoxins, glutaredoxins, and peroxiredoxins?molecular mechanisms and health significance: from cofactors to antioxidants to redox signaling. *Antioxidants & redox signaling*, 19(13):1539–1605.

[Happel et al., 2008] Happel, N., Doenecke, D., Sekeri-Pataryas, K. E., and Sourlingas, T. G. (2008). H1 histone subtype constitution and phosphorylation state of the ageing cell system of human peripheral blood lymphocytes. *Experimental gerontology*, 43(3):184–199.

[Harrison et al., 2009] Harrison, D. E., Strong, R., Sharp, Z. D., Nelson, J. F., Astle, C. M., Flurkey, K., Nadon, N. L., Wilkinson, J. E., Frenkel, K., Carter, C. S., et al. (2009). Rapamycin fed late in life extends lifespan in genetically heterogeneous mice. *nature*, 460(7253):392–395.

[Hergeth and Schneider, 2015] Hergeth, S. P. and Schneider, R. (2015). The h1 linker histones: multifunctional proteins beyond the nucleosomal core particle. *EMBO reports*, page e201540749.

[Hietakangas et al., 2006] Hietakangas, V., Anckar, J., Blomster, H. A., Fujimoto, M., Palvimo, J. J., Nakai, A., and Sistonen, L. (2006). Pdsm, a motif for phosphorylation-dependent sumo modification. *Proceedings of the National Academy of Sciences of the United States of America*, 103(1):45–50.

[Hitosugi et al., 2009] Hitosugi, T., Kang, S., Vander Heiden, M. G., Chung, T.-W., Elf, S., Lythgoe, K., Dong, S., Lonial, S., Wang, X., Chen, G. Z., et al. (2009). Tyrosine phosphorylation inhibits pkm2 to promote the warburg effect and tumor growth. *Science signaling*, 2(97):ra73.

[Hornbeck et al., 2015] Hornbeck, P. V., Zhang, B., Murray, B., Kornhauser, J. M., Latham, V., and Skrzypek, E. (2015). Phosphositeplus, 2014: mutations, ptms and recalibrations. *Nucleic acids research*, 43(D1):D512–D520.

[Hotamisligil, 2017] Hotamisligil, G. S. (2017). Inflammation, metaflammation and immunometabolic disorders. *Nature*, 542(7640):177–185.

[Hsu et al., 2003] Hsu, J.-L., Huang, S.-Y., Chow, N.-H., and Chen, S.-H. (2003). Stable-isotope dimethyl labeling for quantitative proteomics. *Analytical chemistry*, 75(24):6843–6852.

[Huang et al., 2012] Huang, H., Liu, N., Guo, H., Liao, S., Li, X., Yang, C., Liu, S., Song, W., Liu, C., Guan, L., et al. (2012). L-carnitine is an endogenous hdac inhibitor selectively inhibiting cancer cell growth in vivo and in vitro. *PloS one*, 7(11):e49062.

[Huelga et al., 2012] Huelga, S. C., Vu, A. Q., Arnold, J. D., Liang, T. Y., Liu, P. P., Yan, B. Y., Donohue, J. P., Shiue, L., Hoon, S., Brenner, S., et al. (2012). Integrative genome-wide analysis reveals cooperative regulation of alternative splicing by hnrnp proteins. *Cell reports*, 1(2):167–178.

[Hunter, 2007] Hunter, T. (2007). The age of crosstalk: phosphorylation, ubiquitination, and beyond. *Molecular cell*, 28(5):730–738.

[Imai et al., 2000] Imai, S.-I., Armstrong, C. M., Kaeberlein, M., and Guarente, L. (2000). Transcriptional silencing and longevity protein sir2 is an nad-dependent histone deacetylase. *Nature*, 403(6771):795–800.

[Inoki et al., 2003] Inoki, K., Zhu, T., and Guan, K.-L. (2003). Tsc2 mediates cellular energy response to control cell growth and survival. *Cell*, 115(5):577–590.

[Izzo and Schneider, 2016] Izzo, A. and Schneider, R. (2016). The role of linker histone h1 modifications in the regulation of gene expression and chromatin dynamics. *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms*, 1859(3):486–495.

[Jain et al., 2016] Jain, K., Warmack, R. A., Debler, E. W., Hadjikyriacou, A., Stavropoulos, P., and Clarke, S. G. (2016). Protein arginine methyltransferase product specificity is mediated by distinct active-site architectures. *Journal of Biological Chemistry*, 291(35):18299–18308.

[Johnson, 2009] Johnson, L. N. (2009). The regulation of protein phosphorylation. *Biochemical Society Transactions*, 37(4):627–641.

[Johnson et al., 2013] Johnson, S. C., Rabinovitch, P. S., and Kaeberlein, M. (2013). mtor is a key modulator of ageing and age-related disease. *Nature*, 493(7432):338–345.

[Jung et al., 2011] Jung, G. A., Shin, B. S., Jang, Y. S., Sohn, J. B., Woo, S. R., Kim, J. E., Choi, G., Lee, K.-M., Min, B. H., Lee, K.-H., et al. (2011). Methylation of eukaryotic elongation factor 2 induced by basic fibroblast growth factor via mitogen-activated protein kinase. *Experimental & molecular medicine*, 43(10):550–560.

[Kamieniarz et al., 2012] Kamieniarz, K., Izzo, A., Dundr, M., Tropberger, P., Ozretić, L., Kirfel, J., Scheer, E., Tropel, P., Wiśniewski, J. R., Tora, L., et al. (2012). A dual role of linker histone h1. 4 lys 34 acetylation in transcriptional activation. *Genes & development*, 26(8):797–802.

[Kapahi et al., 2004] Kapahi, P., Zid, B. M., Harper, T., Koslover, D., Sapin, V., and Benzer, S. (2004). Regulation of lifespan in drosophila by modulation of genes in the tor signaling pathway. *Current Biology*, 14(10):885–890.

[Keller et al., 2002] Keller, A., Nesvizhskii, A. I., Kolker, E., and Aebersold, R. (2002). Empirical statistical model to estimate the accuracy of peptide identifications made by ms/ms and database search. *Analytical chemistry*, 74(20):5383–5392.

[Kettenbach et al., 2011] Kettenbach, A. N., Schweppe, D. K., Faherty, B. K., Pechenick, D., Pletnev, A. A., and Gerber, S. A. (2011). Quantitative phosphoproteomics identifies substrates and functional modules of aurora and polo-like kinase activities in mitotic cells. *Science signaling*, 4(179).

[Kim et al., 2006] Kim, S. C., Sprung, R., Chen, Y., Xu, Y., Ball, H., Pei, J., Cheng, T., Kho, Y., Xiao, H., Xiao, L., et al. (2006). Substrate and functional diversity of lysine acetylation revealed by a proteomics survey. *Molecular cell*, 23(4):607–618.

[Kim et al., 2015] Kim, S.-Y., Sim, C. K., Tang, H., Han, W., Zhang, K., and Xu, F. (2015). Acetylome analysis identifies sirt1 targets in mrna-processing and chromatin-remodeling in mouse liver. *PloS one*, 10(10):e0140619.

[Kim et al., 2011] Kim, W., Bennett, E. J., Huttlin, E. L., Guo, A., Li, J., Possemato, A., Sowa, M. E., Rad, R., Rush, J., Comb, M. J., et al. (2011). Systematic and quantitative assessment of the ubiquitin-modified proteome. *Molecular cell*, 44(2):325–340.

[Koepp et al., 2001] Koepp, D. M., Schaefer, L. K., Ye, X., Keyomarsi, K., Chu, C., Harper, J. W., and Elledge, S. J. (2001). Phosphorylation-dependent ubiquitination of cyclin e by the scffbw7 ubiquitin ligase. *Science*, 294(5540):173–177.

[Koumbadinga et al., 2015] Koumbadinga, G. A., Mahmood, N., Lei, L., Kan, Y., Cao, W., Lobo, V. G., Yao, X., Zhang, S., and Xie, J. (2015). Increased stability of heterogeneous ribonucleoproteins by a deacetylase inhibitor. *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms*, 1849(8):1095–1103.

[Kreegipuu et al., 1998] Kreegipuu, A., Blom, N., Brunak, S., and Järv, J. (1998). Statistical analysis of protein kinase specificity determinants. *FEBS letters*, 430(1-2):45–50.

[Larsen et al., 2016] Larsen, S. C., Sylvestersen, K. B., Mund, A., Lyon, D., Mullari, M., Madsen, M. V., Daniel, J. A., Jensen, L. J., and Nielsen, M. L. (2016). Proteome-wide analysis of arginine monomethylation reveals widespread occurrence in human cells. *Sci. Signal.*, 9(443):rs9–rs9.

[Lee and Workman, 2007] Lee, K. K. and Workman, J. L. (2007). Histone acetyltransferase complexes: one size doesn't fit all. *Nature reviews Molecular cell biology*, 8(4):284–295.

[Lehninger et al., 2005] Lehninger, A., Nelson, D., and Cox, M. (2005). *Lehninger Principles of Biochemistry*. W. H. Freeman.

[Levin et al., 2016] Levin, R. S., Hertz, N. T., Burlingame, A. L., Shokat, K. M., and Mukherjee, S. (2016). Innate immunity kinase tak1 phosphorylates rab1 on a hotspot for posttranslational modifications by host and pathogen. *Proceedings of the National Academy of Sciences*, 113(33):E4776–E4783.

[Li et al., 2012] Li, J.-Y., Patterson, M., Mikkola, H. K., Lowry, W. E., and Kurdistani, S. K. (2012). Dynamic distribution of linker histone h1. 5 in cellular differentiation. *PLoS genetics*, 8(8):e1002879.

[Li et al., 2016] Li, X., Jiang, Y., Meisenhelder, J., Yang, W., Hawke, D. H., Zheng, Y., Xia, Y., Aldape, K., He, J., Hunter, T., et al. (2016). Mitochondria-translocated pgk1 functions as a protein kinase to coordinate glycolysis and the tca cycle in tumorigenesis. *Molecular cell*, 61(5):705–719.

[Liao and Mizzen, 2017] Liao, R. and Mizzen, C. A. (2017). Site-specific regulation of histone h1 phosphorylation in pluripotent cell differentiation. *Epigenetics & chromatin*, 10(1):29.

[Lindner et al., 2010] Lindner, U., Kramer, J., Rohwedel, J., and Schlenke, P. (2010). Mesenchymal stem or stromal cells: toward a better understanding of their biology? *Transfusion Medicine and Hemotherapy*, 37(2):75–83.

[Liu et al., 2015] Liu, F., Cheng, G., Hamard, P.-J., Greenblatt, S., Wang, L., Man, N., Perna, F., Xu, H., Tadi, M., Luciani, L., et al. (2015). Arginine methyltransferase prmt5 is essential for sustaining normal adult hematopoiesis. *The Journal of clinical investigation*, 125(9):3532–3544.

[López-Otín et al., 2013] López-Otín, C., Blasco, M. A., Partridge, L., Serrano, M., and Kroemer, G. (2013). The hallmarks of aging. *Cell*, 153(6):1194–1217.

[López-Otín et al., 2016] López-Otín, C., Galluzzi, L., Freije, J. M., Madeo, F., and Kroemer, G. (2016). Metabolic control of longevity. *Cell*, 166(4):802–821.

[Lv et al., 2011] Lv, L., Li, D., Zhao, D., Lin, R., Chu, Y., Zhang, H., Zha, Z., Liu, Y., Li, Z., Xu, Y., et al. (2011). Acetylation targets the m2 isoform of pyruvate kinase for degradation through chaperone-mediated autophagy and promotes tumor growth. *Molecular cell*, 42(6):719–730.

[Mackay et al., 1999] Mackay, A., Beck, S., Jaiswal, R., Douglas, R., Mosca, J., Moorman, M., Simonetti, D., Craig, S., and Marshak, D. (1999). Multilineage potential of adult human mesenchymal stem cells. *Science*, 284(5411):143–147.

[Mair et al., 2011] Mair, W., Morantte, I., Rodrigues, A. P., Manning, G., Montminy, M., Shaw, R. J., and Dillin, A. (2011). Lifespan extension induced by ampk and calcineurin is mediated by crtc-1 and creb. *Nature*, 470(7334):404–408.

[Manning et al., 2002a] Manning, G., Plowman, G. D., Hunter, T., and Sudarsanam, S. (2002a). Evolution of protein kinase signaling from yeast to man. *Trends in biochemical sciences*, 27(10):514–520.

[Manning et al., 2002b] Manning, G., Whyte, D. B., Martinez, R., Hunter, T., and Sudarsanam, S. (2002b). The protein kinase complement of the human genome. *Science*, 298(5600):1912–1934.

[Mariño et al., 2014] Mariño, G., Pietrocola, F., Eisenberg, T., Kong, Y., Malik, S. A., Andryushkova, A., Schroeder, S., Pendl, T., Harger, A., Niso-Santano, M., et al. (2014). Regulation of autophagy by cytosolic acetyl-coenzyme a. *Molecular cell*, 53(5):710–725.

[Maris et al., 2005] Maris, C., Dominguez, C., and Allain, F. H.-T. (2005). The rna recognition motif, a plastic rna-binding platform to regulate post-transcriptional gene expression. *The FEBS journal*, 272(9):2118–2131.

[Matsuzaki et al., 2005] Matsuzaki, H., Daitoku, H., Hatta, M., Aoyama, H., Yoshimochi, K., and Fukamizu, A. (2005). Acetylation of foxo1 alters its dna-binding ability and sensitivity to phosphorylation. *Proceedings of the National Academy of Sciences of the United States of America*, 102(32):11278–11283.

[Mazumder et al., 2014] Mazumder, M., Padhan, N., Bhattacharya, A., and Gourinath, S. (2014). Prediction and analysis of canonical ef hand loop and qualitative estimation of ca 2+ binding affinity. *PloS one*, 9(4):e96202.

[Mertins et al., 2013] Mertins, P., Qiao, J. W., Patel, J., Udeshi, N. D., Clauser, K. R., Mani, D., Burgess, M. W., Gillette, M. A., Jaffe, J. D., and Carr, S. A. (2013). Integrated proteomic analysis of post-translational modifications by serial enrichment. *Nature methods*, 10(7):634–637.

[Mihaylova and Shaw, 2011] Mihaylova, M. M. and Shaw, R. J. (2011). The ampk signalling pathway coordinates cell growth, autophagy and metabolism. *Nature cell biology*, 13(9):1016–1023.

[Millard et al., 2013] Millard, C. J., Watson, P. J., Celardo, I., Gordiyenko, Y., Cowley, S. M., Robinson, C. V., Fairall, L., and Schwabe, J. W. (2013). Class i hdacs share a common mechanism of regulation by inositol phosphates. *Molecular cell*, 51(1):57–67.

[Morrison and Scadden, 2014] Morrison, S. J. and Scadden, D. T. (2014). The bone marrow niche for haematopoietic stem cells. *Nature*, 505(7483):327–334.

[Morrison et al., 1995] Morrison, S. J., Uchida, N., and Weissman, I. L. (1995). The biology of hematopoietic stem cells. *Annual review of cell and developmental biology*, 11(1):35–71.

[Morselli et al., 2010] Morselli, E., Maiuri, M., Markaki, M., Megalou, E., Pasparaki, A., Palikaras, K., Criollo, A., Galluzzi, L., Malik, S., Vitale, I., et al. (2010). Caloric restriction and resveratrol promote longevity through the sirtuin-1-dependent induction of autophagy. *Cell death & disease*, 1(1):e10.

[Moussaieff et al., 2015] Moussaieff, A., Rouleau, M., Kitsberg, D., Cohen, M., Levy, G., Barasch, D., Nemirovski, A., Shen-Orr, S., Laevsky, I., Amit, M., et al. (2015). Glycolysis-mediated changes in acetyl-coa and histone acetylation control the early differentiation of embryonic stem cells. *Cell metabolism*, 21(3):392–402.

[Neault et al., 2012] Neault, M., Mallette, F. A., Vogel, G., Michaud-Levesque, J., and Richard, S. (2012). Ablation of prmt6 reveals a role as a negative transcriptional regulator of the p53 tumor suppressor. *Nucleic acids research*, 40(19):9513–9521.

[Ng and Tang, 2013] Ng, F. and Tang, B. L. (2013). Sirtuins' modulation of autophagy. *Journal of cellular physiology*, 228(12):2262–2270.

[Nguyen et al., 2013] Nguyen, L. K., Kolch, W., and Kholodenko, B. N. (2013). When ubiquitination meets phosphorylation: a systems biology perspective of egfr/mapk signalling. *Cell Communication and Signaling*, 11(1):52.

[Nielsen et al., 2006] Nielsen, M. L., Savitski, M. M., and Zubarev, R. A. (2006). Extent of modifications in human proteome samples and their effect on dynamic range of analysis in shotgun proteomics. *Molecular & Cellular Proteomics*, 5(12):2384–2391.

[Oda et al., 2001] Oda, Y., Nagasu, T., and Chait, B. T. (2001). Enrichment analysis of phosphorylated proteins as a tool for probing the phosphoproteome. *Nature biotechnology*, 19(4):379–382.

[Oh et al., 2014] Oh, J., Lee, Y. D., and Wagers, A. J. (2014). Stem cell aging: mechanisms, regulators and therapeutic opportunities. *Nature medicine*, 20(8):870–880.

[Oishi and Manabe, 2016] Oishi, Y. and Manabe, I. (2016). Macrophages in age-related chronic inflammatory diseases. *npj Aging and Mechanisms of Disease*, 2:16018.

[Olsen et al., 2010] Olsen, J. V., Vermeulen, M., Santamaria, A., Kumar, C., Miller, M. L., Jensen, L. J., Gnad, F., Cox, J., Jensen, T. S., Nigg, E. A., Brunak, S., and Mann, M. (2010). Quantitative phosphoproteomics reveals widespread full phosphorylation site occupancy during mitosis. *Science Signaling*, 3(104):ra3–ra3.

[Ong et al., 2002] Ong, S.-E., Blagoev, B., Kratchmarova, I., Kristensen, D. B., Steen, H., Pandey, A., and Mann, M. (2002). Stable isotope labeling by amino acids in cell culture, silac, as a simple and accurate approach to expression proteomics. *Molecular & cellular proteomics*, 1(5):376–386.

[Ong et al., 2004] Ong, S.-E., Mittler, G., and Mann, M. (2004). Identifying and quantifying in vivo methylation sites by heavy methyl silac. *Nature methods*, 1(2):119–126.

[O'shea et al., 2013] O'shea, J. P., Chou, M. F., Quader, S. A., Ryan, J. K., Church, G. M., and Schwartz, D. (2013). plogo: a probabilistic approach to visualizing sequence motifs. *Nature methods*, 10(12):1211–1212.

[Ou et al., 2005] Ou, Y.-H., Chung, P.-H., Sun, T.-P., and Shieh, S.-Y. (2005). p53 c-terminal phosphorylation by chk1 and chk2 participates in the regulation of dna-damage-induced c-terminal acetylation. *Molecular biology of the cell*, 16(4):1684–1695.

[P De Miguel et al., 2012] P De Miguel, M., Fuentes-Julian, S., Blazquez-Martinez, A., Y Pascual, C., A Aller, M., Arias, J., and Arnalich-Montiel, F. (2012). Immunosuppressive properties of mesenchymal stem cells: advances and applications. *Current molecular medicine*, 12(5):574–591.

[Pal and Tyler, 2016] Pal, S. and Tyler, J. K. (2016). Epigenetics and aging. *Science advances*, 2(7):e1600584.

[Pang et al., 2013] Pang, L., Tian, H., Chang, N., Yi, J., Xue, L., Jiang, B., Gorospe, M., Zhang, X., and Wang, W. (2013). Loss of carm1 is linked to reduced hur function in replicative senescence. *BMC molecular biology*, 14(1):15.

[Panopoulos et al., 2012] Panopoulos, A. D., Yanes, O., Ruiz, S., Kida, Y. S., Diep, D., Tautenhahn, R., Herrerías, A., Batchelder, E. M., Plongthongkum, N., Lutz, M., et al. (2012). The metabolome of induced pluripotent stem cells reveals metabolic changes occurring in somatic cell reprogramming. *Cell research*, 22(1):168–177.

[Park et al., 2016a] Park, S.-H., Ozden, O., Liu, G., Song, H. Y., Zhu, Y., Yan, Y., Zou, X., Kang, H.-J., Jiang, H., Principe, D. R., et al. (2016a). Sirt2-mediated deacetylation and tetramerization of pyruvate kinase directs glycolysis and tumor growth. *Cancer research*, 76(13):3802–3812.

[Park et al., 2016b] Park, Y. S., Kim, D. J., Koo, H., Jang, S. H., You, Y.-M., Cho, J. H., Yang, S.-J., Yu, E. S., Jung, Y., Lee, D. C., et al. (2016b). Akt-induced pkm2 phosphorylation signals for igf-1-stimulated cancer cell growth. *Oncotarget*, 7(30):48155.

[Paz et al., 2014] Paz, J. C., Park, S., Phillips, N., Matsumura, S., Tsai, W.-W., Kasper, L., Brindle, P. K., Zhang, G., Zhou, M.-M., Wright, P. E., et al. (2014). Combinatorial regulation of a signal-dependent activator by phosphorylation and acetylation. *Proceedings of the National Academy of Sciences*, 111(48):17116–17121.

[Peng et al., 2003] Peng, J., Schwartz, D., Elias, J. E., Thoreen, C. C., Cheng, D., Marsischky, G., Roelofs, J., Finley, D., and Gygi, S. P. (2003). A proteomics approach to understanding protein ubiquitination. *Nature biotechnology*, 21(8):921–926.

[Picotti and Aebersold, 2012] Picotti, P. and Aebersold, R. (2012). Selected reaction monitoring-based proteomics: workflows, potential, pitfalls and future directions. *Nature methods*, 9(6):555–566.

[Pinkse et al., 2004] Pinkse, M. W., Uitto, P. M., Hilhorst, M. J., Ooms, B., and Heck, A. J. (2004). Selective isolation at the femtomole level of phosphopeptides from proteolytic digests using 2d-nanolc-esi-ms/ms and titanium oxide precolumns. *Analytical chemistry*, 76(14):3935–3943.

[Pinto et al., 2015] Pinto, S. M., Nirujogi, R. S., Rojas, P. L., Patil, A. H., Manda, S. S., Subbannayya, Y., Roa, J. C., Chatterjee, A., Prasad, T., and Pandey, A. (2015). Quantitative phosphoproteomic analysis of il-33-mediated signaling. *Proteomics*, 15(2-3):532–544.

[Proud, 2015] Proud, C. G. (2015). Regulation and roles of elongation factor 2 kinase.

[Qvit et al., 2016] Qvit, N., Joshi, A. U., Cunningham, A. D., Ferreira, J. C., and Mochly-Rosen, D. (2016). Glyceraldehyde-3-phosphate dehydrogenase (gapdh) protein-protein interaction inhibitor reveals a non-catalytic role for gapdh oligomerization in cell death. *Journal of Biological Chemistry*, 291(26):13608–13621.

[Raftery and Geczy, 1998] Raftery, M. J. and Geczy, C. L. (1998). Identification of posttranslational modifications and cdna sequencing errors in the rat s100 proteins mrp8 and 14 using electrospray ionization mass spectrometry. *Analytical biochemistry*, 258(2):285–292.

[Rathert et al., 2008] Rathert, P., Dhayalan, A., Murakami, M., Zhang, X., Tamas, R., Jurkowska, R., Komatsu, Y., Shinkai, Y., Cheng, X., and Jeltsch, A. (2008). Protein lysine methyltransferase g9a acts on non-histone targets. *Nature chemical biology*, 4(6):344–346.

[Rattan et al., 1992] Rattan, S. I., Derventzi, A., and Clark, B. F. (1992). Protein synthesis, posttranslational modifications, and aginga. *Annals of the New York Academy of Sciences*, 663(1):48–62.

[Rauniyar and Yates III, 2014] Rauniyar, N. and Yates III, J. R. (2014). Isobaric labeling-based relative quantification in shotgun proteomics. *Journal of proteome research*, 13(12):5293–5309.

[Redpath et al., 1996] Redpath, N., Foulstone, E., and Proud, C. G. (1996). Regulation of translation elongation factor-2 by insulin via a rapamycin-sensitive signalling pathway. *The EMBO journal*, 15(9):2291.

[Rikova et al., 2007] Rikova, K., Guo, A., Zeng, Q., Possemato, A., Yu, J., Haack, H., Nardone, J., Lee, K., Reeves, C., Li, Y., et al. (2007). Global survey of phosphotyrosine signaling identifies oncogenic kinases in lung cancer. *Cell*, 131(6):1190–1203.

[Roque et al., 2008] Roque, A., Ponte, I., Arrondo, J. L. R., and Suau, P. (2008). Phosphorylation of the carboxy-terminal domain of histone h1: effects on secondary structure and dna condensation. *Nucleic acids research*, 36(14):4719–4726.

[Ross et al., 2004] Ross, P. L., Huang, Y. N., Marchese, J. N., Williamson, B., Parker, K., Hattan, S., Khainovski, N., Pillai, S., Dey, S., Daniels, S., et al. (2004). Multiplexed protein quantitation in saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. *Molecular & cellular proteomics*, 3(12):1154–1169.

[Roth et al., 2001] Roth, S. Y., Denu, J. M., and Allis, C. D. (2001). Histone acetyltransferases. *Annual review of biochemistry*, 70(1):81–120.

[Rush et al., 2005] Rush, J., Moritz, A., Lee, K. A., Guo, A., Goss, V. L., Spek, E. J., Zhang, H., Zha, X.-M., Polakiewicz, R. D., and Comb, M. J. (2005). Immunoaffinity profiling of tyrosine phosphorylation in cancer cells. *Nature biotechnology*, 23(1):94–101.

[Schullery et al., 1999] Schullery, D. S., Ostrowski, J., Denisenko, O. N., Stempka, L., Shnyreva, M., Suzuki, H., Gschwendt, M., and Bomsztyk, K. (1999). Regulated interaction of protein kinase cδ with the heterogeneous nuclear ribonucleoprotein k protein. *Journal of Biological Chemistry*, 274(21):15101–15109.

[Schuringa et al., 2004] Schuringa, J. J., Chung, K. Y., Morrone, G., and Moore, M. A. (2004). Constitutive activation of stat5a promotes human hematopoietic stem cell self-renewal and erythroid differentiation. *Journal of Experimental Medicine*, 200(5):623–635.

[Sharma et al., 2014] Sharma, K., D?Souza, R. C., Tyanova, S., Schaab, C., Wiśniewski, J. R., Cox, J., and Mann, M. (2014). Ultradeep human phosphoproteome reveals a distinct regulatory nature of tyr and ser/thr-based signaling. *Cell reports*, 8(5):1583–1594.

[Shi et al., 2004] Shi, Y., Lan, F., Matson, C., Mulligan, P., Whetstine, J. R., Cole, P. A., Casero, R. A., and Shi, Y. (2004). Histone demethylation mediated by the nuclear amine oxidase homolog lsd1. *Cell*, 119(7):941–953.

[Shimazu et al., 2013] Shimazu, T., Hirschey, M. D., Newman, J., He, W., Shirakawa, K., Le Moan, N., Grueter, C. A., Lim, H., Saunders, L. R., Stevens, R. D., et al. (2013). Suppression of oxidative stress by β-hydroxybutyrate, an endogenous histone deacetylase inhibitor. *Science*, 339(6116):211–214.

[Simon et al., 2015] Simon, A. K., Hollander, G. A., and McMichael, A. (2015). Evolution of the immune system in humans from infancy to old age. In *Proc. R. Soc. B*, volume 282, page 20143085. The Royal Society.

[Skinner and Kelleher, 2015] Skinner, O. S. and Kelleher, N. L. (2015). Illuminating the dark matter of shotgun proteomics. *Nature biotechnology*, 33(7):717–719.

[Smith et al., 2017] Smith, P., Willemsen, D., Popkes, M. L., Metge, F., Gandiwa, E., Reichard, M., and Valenzano, D. R. (2017). Regulation of life span by the gut microbiota in the short-lived african turquoise killifish. *bioRxiv*, page 120980.

[Sol et al., 2012] Sol, E. M., Wagner, S. A., Weinert, B. T., Kumar, A., Kim, H.-S., Deng, C.-X., and Choudhary, C. (2012). Proteomic investigations of lysine acetylation identify diverse substrates of mitochondrial deacetylase sirt3. *PloS one*, 7(12):e50545.

[Swaney et al., 2013] Swaney, D. L., Beltrao, P., Starita, L., Guo, A., Rush, J., Fields, S., Krogan, N. J., and Villén, J. (2013). Global analysis of phosphorylation and ubiquitylation cross-talk in protein degradation. *Nature methods*, 10(7):676–682.

[Swaney et al., 2009] Swaney, D. L., Wenger, C. D., Thomson, J. A., and Coon, J. J. (2009). Human embryonic stem cell phosphoproteome revealed by electron transfer dissociation tandem mass spectrometry. *Proceedings of the National Academy of Sciences*, 106(4):995–1000.

[Talasz et al., 2009] Talasz, H., Sarg, B., and Lindner, H. H. (2009). Site-specifically phosphorylated forms of h1. 5 and h1. 2 localized at distinct regions of the nucleus are related to different processes during the cell cycle. *Chromosoma*, 118(6):693–709.

[Tang et al., 2013] Tang, Q., Koh, L. K., Jiang, D., and Schwarz, H. (2013). Cd137 ligand reverse signaling skews hematopoiesis towards myelopoiesis during aging. *Aging (Albany NY)*, 5(9):643–652.

[Terme et al., 2014] Terme, J.-M., Millán-Ariño, L., Mayor, R., Luque, N., Izquierdo-Bouldstridge, A., Bustillos, A., Sampaio, C., Canes, J., Font, I., Sima, N., et al. (2014). Dynamics and dispensability of variant-specific histone h1 lys-26/ser-27 and thr-165 post-translational modifications. *FEBS letters*, 588(14):2353–2362.

[Terme et al., 2011] Terme, J.-M., Sesé, B., Millán-Ariño, L., Mayor, R., Belmonte, J. C. I., Barrero, M. J., and Jordan, A. (2011). Histone h1 variants are differentially expressed and incorporated into chromatin during differentiation and reprogramming to pluripotency. *Journal of Biological Chemistry*, 286(41):35347–35357.

[Thangaraju et al., 2009] Thangaraju, M., Carswell, K. N., Prasad, P. D., and Ganapathy, V. (2009). Colon cancer cells maintain low levels of pyruvate to avoid cell death caused by inhibition of hdac1/hdac3. *Biochemical Journal*, 417(1):379–389.

[Thompson et al., 2003] Thompson, A., Schäfer, J., Kuhn, K., Kienle, S., Schwarz, J., Schmidt, G., Neumann, T., and Hamon, C. (2003). Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by ms/ms. *Analytical chemistry*, 75(8):1895–1904.

[Thornton et al., 1999] Thornton, J. M., Orengo, C. A., Todd, A. E., and Pearl, F. M. (1999). Protein folds, functions and evolution. *Journal of molecular biology*, 293(2):333–342.

[Tran et al., 2011] Tran, J. C., Zamdborg, L., Ahlf, D. R., Lee, J. E., Catherman, A. D., Durbin, K. R., Tipton, J. D., Vellaichamy, A., Kellie, J. F., Li, M., et al. (2011). Mapping intact protein isoforms in discovery mode using top-down proteomics. *Nature*, 480(7376):254–258.

[Tsai et al., 2015] Tsai, C.-F., Wang, Y.-T., Yen, H.-Y., Tsou, C.-C., Ku, W.-C., Lin, P.-Y., Chen, H.-Y., Nesvizhskii, A. I., Ishihama, Y., and Chen, Y.-J. (2015). Large-scale determination of absolute phosphorylation stoichiometries in human cells by motif-targeting quantitative proteomics. *Nature communications*, 6.

[Tsukada et al., 2006] Tsukada, Y.-i., Fang, J., Erdjument-Bromage, H., Warren, M. E., Borchers, C. H., Tempst, P., and Zhang, Y. (2006). Histone demethylation by a family of jmjc domain-containing proteins. *Nature*, 439(7078):811–816.

[Uhlén et al., 2015] Uhlén, M., Fagerberg, L., Hallström, B. M., Lindskog, C., Oksvold, P., Mardinoglu, A., Sivertsson, Å., Kampf, C., Sjöstedt, E., Asplund, A., et al. (2015). Tissue-based map of the human proteome. *Science*, 347(6220):1260419.

[van Dam et al., 2017] van Dam, S., Võsa, U., van der Graaf, A., Franke, L., and de Magalhães, J. P. (2017). Gene co-expression analysis for functional classification and gene–disease predictions. *Briefings in bioinformatics*, page bbw139.

[Van Duin et al., 2007] Van Duin, D., Mohanty, S., Thomas, V., Ginter, S., Montgomery, R. R., Fikrig, E., Allore, H. G., Medzhitov, R., and Shaw, A. C. (2007). Age-associated defect in human tlr-1/2 function. *The Journal of Immunology*, 178(2):970–975.

[van Noort et al., 2012] van Noort, V., Seebacher, J., Bader, S., Mohammed, S., Vonkova, I., Betts, M. J., Kühner, S., Kumar, R., Maier, T., O'Flaherty, M., et al. (2012). Cross-talk between phosphorylation and lysine acetylation in a genome-reduced bacterium. *Molecular systems biology*, 8(1):571.

[Vellai et al., 2003] Vellai, T., Takacs-Vellai, K., Zhang, Y., Kovacs, A. L., Orosz, L., and Müller, F. (2003). Influence of tor kinase on lifespan in c. elegans. *Nature*, 426:620.

[Ventura et al., 2010] Ventura, M., Mateo, F., Serratosa, J., Salaet, I., Carujo, S., Bachs, O., and Pujol, M. J. (2010). Nuclear translocation of glyceraldehyde-3-phosphate dehydrogenase is regulated by acetylation. *The international journal of biochemistry & cell biology*, 42(10):1672–1680.

[Wahlestedt et al., 2017] Wahlestedt, M., Erlandsson, E., Kristiansen, T., Lu, R., Brakebusch, C., Weissman, I. L., Yuan, J., Martin-Gonzalez, J., and Bryder, D. (2017). Clonal reversal of ageing-associated stem cell lineage bias via a pluripotent intermediate. *Nature Communications*, 8:14533.

[Walker and Slifka, 2010] Walker, J. M. and Slifka, M. K. (2010). Longevity of t-cell memory following acute viral infection. In *Memory T Cells*, pages 96–107. Springer.

[Walsh, 2006] Walsh, C. (2006). *Posttranslational modification of proteins: expanding nature's inventory*. Roberts and Company Publishers.

[Walther et al., 2015] Walther, D. M., Kasturi, P., Zheng, M., Pinkert, S., Vecchi, G., Ciryam, P., Morimoto, R. I., Dobson, C. M., Vendruscolo, M., Mann, M., et al. (2015). Widespread proteome remodeling and aggregation in aging c. elegans. *Cell*, 161(4):919–932.

[Weichsel et al., 2007] Weichsel, A., Brailey, J. L., and Montfort, W. R. (2007). Buried s-nitrosocysteine revealed in crystal structures of human thioredoxin? *Biochemistry*, 46(5):1219–1227.

[Weiskopf et al., 2009] Weiskopf, D., Weinberger, B., and Grubeck-Loebenstein, B. (2009). The aging of the immune system. *Transplant international*, 22(11):1041–1050.

[Wiśniewski et al., 2007] Wiśniewski, J. R., Zougman, A., Krüger, S., and Mann, M. (2007). Mass spectrometric mapping of linker histone h1 variants reveals multiple acetylations, methylations, and phosphorylation as well as differences between cell culture and tissue. *Molecular & Cellular Proteomics*, 6(1):72–87.

[Wright and Dyson, 2015] Wright, P. E. and Dyson, H. J. (2015). Intrinsically disordered proteins in cellular signalling and regulation. *Nature Reviews Molecular Cell Biology*, 16(1):18–29.

[Wu et al., 2004] Wu, C. C., MacCoss, M. J., Howell, K. E., Matthews, D. E., and Yates, J. R. (2004). Metabolic labeling of mammalian organisms with stable isotopes for quantitative proteomic analysis. *Analytical chemistry*, 76(17):4951–4959.

[Wu et al., 2015] Wu, Q., Cheng, Z., Zhu, J., Xu, W., Peng, X., Chen, C., Li, W., Wang, F., Cao, L., Yi, X., et al. (2015). Suberoylanilide hydroxamic acid treatment reveals crosstalks among proteome, ubiquitylome and acetylome in non-small cell lung cancer a549 cell line. *Scientific reports*, 5:9520.

[Xiong and Guan, 2012] Xiong, Y. and Guan, K.-L. (2012). Mechanistic insights into the regulation of metabolic enzymes by acetylation. *J Cell Biol*, 198(2):155–164.

[Yanes et al., 2010] Yanes, O., Clark, J., Wong, D. M., Patti, G. J., Sánchez-Ruiz, A., Benton, H. P., Trauger, S. A., Desponts, C., Ding, S., and Siuzdak, G. (2010). Metabolic oxidation regulates embryonic stem cell differentiation. *Nature chemical biology*, 6(6):411–417.

[Yogalingam et al., 2013] Yogalingam, G., Hwang, S., Ferreira, J. C., and Mochly-Rosen, D. (2013). Glyceraldehyde-3-phosphate dehydrogenase (gapdh) phosphorylation by protein kinase cδ (pkcδ) inhibits mitochondria elimination by lysosomal-like structures following ischemia and reoxygenation-induced injury. *Journal of Biological Chemistry*, 288(26):18947–18960.

[Yoon et al., 2009] Yoon, K. W., Cho, J.-H., Lee, J. K., Kang, Y.-H., Chae, J. S., Kim, Y. M., Kim, J., Kim, E. K., Kim, S. E., Baik, J.-H., et al. (2009). Cib1 functions as a ca2+-sensitive modulator of stress-induced signaling by targeting ask1. *Proceedings of the National Academy of Sciences*, 106(41):17389–17394.

[Zeidan and Hart, 2010] Zeidan, Q. and Hart, G. W. (2010). The intersections between o-glcnacylation and phosphorylation: implications for multiple signaling pathways. *J Cell Sci*, 123(1):13–22.

[Zhang et al., 2009] Zhang, J., Sprung, R., Pei, J., Tan, X., Kim, S., Zhu, H., Liu, C.-F., Grishin, N. V., and Zhao, Y. (2009). Lysine acetylation is a highly abundant and evolutionarily conserved modification in escherichia coli. *Molecular & Cellular Proteomics*, 8(2):215–225.

[Zhang and Cheng, 2003] Zhang, X. and Cheng, X. (2003). Structure of the predominant protein arginine methyltransferase prmt1 and analysis of its binding to substrate peptides. *Structure*, 11(5):509–520.

[Zhang et al., 2012] Zhang, Y., Cooke, M., Panjwani, S., Cao, K., Krauth, B., Ho, P.-Y., Medrzycki, M., Berhe, D. T., Pan, C., McDevitt, T. C., et al. (2012). Histone h1 depletion impairs embryonic stem cell differentiation. *PLoS genetics*, 8(5):e1002691.

[Zhang et al., 2013] Zhang, Y., Fonslow, B. R., Shan, B., Baek, M.-C., and Yates III, J. R. (2013). Protein analysis by shotgun/bottom-up proteomics. *Chemical reviews*, 113(4):2343–2394.

[Zhao et al., 2012] Zhao, E., Xu, H., Wang, L., Kryczek, I., Wu, K., Hu, Y., Wang, G., and Zou, W. (2012). Bone marrow and the control of immunity. *Cellular & molecular immunology*, 9(1):11–19.

[Zhao and Jensen, 2009] Zhao, Y. and Jensen, O. N. (2009). Modification-specific proteomics: strategies for characterization of post-translational modifications using enrichment techniques. *Proteomics*, 9(20):4632–4641.

[Zheng et al., 2003] Zheng, L., Roeder, R. G., and Luo, Y. (2003). S phase activation of the histone h2b promoter by oca-s, a coactivator complex that contains gapdh as a key component. *Cell*, 114(2):255–266.

[Zhu et al., 2000] Zhu, H., Klemic, J. F., Chang, S., Bertone, P., Casamayor, A., Klemic, K. G., Smith, D., Gerstein, M., Reed, M. A., and Snyder, M. (2000). Analysis of yeast protein kinases using protein chips. *Nature genetics*, 26(3):283–289.

All the world's a stage,
And all the men and women merely players;
They have their exits and their entrances,
And one man in his time plays many parts,
His acts being seven ages. At first, the infant
Mewling and puking in the nurse's arms.
Then the whining schoolboy, with his satchel
And shining morning face, creeping like snail
Unwillingly to school. And then the lover,
Sighing like furnace, with a woeful ballad
Made to his mistress' eyebrow.Then a soldier,
Full of strange oaths and bearded like the pard,
Jealous in honour, sudden and quick in quarrel,
Seeking the bubble reputation
Even in the cannon's mouth. And then the justice,
In fair round belly with good capon lined,
With eyes severe and beard of formal cut,
Full of wise saws and modern instances
And so he plays his part. The sixth age shifts
Into the lean and slippered pantaloon,
With spectacles on nose and pouch on side;
His youthful hose, well saved, a world too wide
For his shrunk shank, and his big manly voice,
Turning again toward childish treble, pipes
And whistles in his sound. Last scene of all,
That ends this strange eventful history,
Is second childishness and mere oblivion,
Sans teeth, sans eyes, sans taste, sans everything.

As You Like It, Act II, Scene VII . W. Shakespear