The Role of Post-Translational Modifications and Protein Turnover during Neutrophil Differentiation and Maturation

Mohammed A. Akbor

St Hilda's College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Trinity 2025

Abstract

Neutrophils are frontline effectors of the innate immune system, rapidly mobilised in response to both inflammation and microbial invasion. Their development from hematopoietic progenitors is a tightly regulated process involving lineage commitment, functional maturation, and priming for rapid immune responses. Although key transcriptional and signalling pathways involved in neutrophil differentiation have been described, the contribution of post-translational modifications (PTMs) – critical modulators of protein activity, stability, and interactions – remains largely unexplored to date.

This work presents the first comprehensive, unbiased characterisation of global proteomic and PTM dynamics during neutrophil differentiation, using state-of-the-art mass spectrometry-based proteomics. By applying parallel enrichment strategies for ubiquitinated (di-Gly-modified) and phosphorylated peptides, over ~6,000 proteins, ~30,000 ubiquitination sites, and ~14,000 phosphorylation events across defined stages of neutrophil development were quantified. Moreover, potential regulators of ubiquitination changes during neutrophil differentiation were interrogated using CRISPR-Cas9 technology to target the deubiquitinases (DUBs) in a phenotypic screen.

These datasets reveal stage-specific patterns of PTM regulation, implicating a coordinated remodelling of signalling networks and effector functions during differentiation. Integrative bioinformatic analyses identified potential candidate master regulators, including previously unrecognised E3 ligases, and deubiquitinases (DUBs), whose activity or modification state strongly correlates with distinct differentiation phases. The targeted CRISPR-Cas9 phenotypic screen identified candidate DUBs that may participate in the regulation of neutrophil differentiation. These included YOD1, BAP1, COPS6, VCPIP1, USP7, USP47, OTUB2, and USP44. Several of these hits are currently under experimental validation to assess their mechanistic role in neutrophil lineage specification and function. Moreover, increasing levels of NLRP3 inflammasome components were detected during neutrophil differentiation, suggesting that NOD-like receptor signalling pathways may contribute to neutrophil development.

These findings provide a high-resolution dynamic molecular PTM landscape of neutrophil differentiation, highlighting the underappreciated role of protein phosphorylation and ubiquitination in shaping innate immune cell fate. This work opens new avenues towards better understanding neutrophil biology and developing strategies to therapeutically modulate immune responses in inflammatory and infectious diseases.

Acknowledgements

First, I would like to thank my supervisor Prof. Benedikt Kessler for warmly welcoming me to the group, for his brilliant mentorship, and for his thoughtful supervision and support throughout my time at Oxford. I extend appreciation to my co-supervisor Prof. Irina Udalova, for her scientific guidance, invaluable advice, and eager collaboration. And to my other co-supervisor Dr. Andreas Damianou, I thank him for his unwavering support day in day out, consistent enthusiasm and energy, and sincere camaraderie. It is an extreme honour and great pleasure to have worked under the mentorship of such exemplary scientists.

I am grateful to all the research facilities that contributed to the experimental work in this thesis. I thank Svenja Hester, Dr. Sarah Flannery, George Berridge, Dr. Iolanda Vendrell, and Prof. Roman Fisher from the Discovery Proteomics team. I highlight my gratitude to Svenja Hester for her expertise in phosphoproteomic analysis. I thank Humaira Islam from the Biotechnology research facility. Lastly, I extend gratitude to Katie Holden and Dr. James Bancroft from the Cellular Imaging Core Facility for all their support with microscopy, and Dr. Maciej Kliszczak for his help with image analysis.

I wish to thank members of the Udalova research group, particularly Dr. Ananda Mukherjee for his collaboration, help with experiments, and advice and suggestions surrounding the biology of the work. And to Dr. Abhinandan Devaprasad for providing valuable data for this thesis.

Many thanks to Dr. Vincenzo D'Angiolella for sharing DUB sgRNAs from his CRISPR library, and Verena Gautsch for helping prepare the bespoke library.

Thank you to my first supervisee, Si Yui Charis Pang, who helped with validation experiments of the CRISPR screen during her summer internship.

I thank all the current and previous members of the Kessler research group and members of the CMD who have helped, advised, and nurtured a friendly and enjoyable atmosphere during my DPhil. Thanks go out to Dr. Zhu Liang, Dr. Hannah Jones, Dr. Simon Davis, Dr. Darragh O'Brien, Dr. Frederik Lassen, Dr. Vassi Sharlandjieva, and Dr. Zhanru Yu.

I would like to acknowledge my DPhil Transfer and Confirmation of Status examiners for their time and input regarding my thesis project.

I thank the Wellcome Trust for funding this research and financially supporting my time in Oxford. To all the staff and students associated with the DPhil Genomic Medicine and Statistics course, from teaching leads to administration staff, thank you for the kindness you showed throughout my time at Oxford.

Last but not least, I am deeply grateful to all my friends and family who have been supportive, understanding, and patient over the course of my DPhil and thesis writing period. In particular, I would like to thank my partner Sam Brandon for his companionship and endless support throughout.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

Abbreviation	Definition
AHR	Aryl hydrocarbon receptor
AMFR	Autocrine motility factor receptor
AML	Acute myeloid leukaemia
APL	Acute promyelocytic leukaemia
ATRA	All-trans retinoic acid
BAHD1	Bromo adjacent homology domain containing 1
BH	Benjamini-Hochberg
BIN2	Bridging integrator 2
BP	biological process
C/EBP	CCAAT-enhancer-binding protein
C3	Complement component 3
C5a	Complement component 5a
C5aR1	C5a receptor 1
CASP1	Caspase 1
CC	cellular component
CCL5	Chemokine (C-C motif) ligand 5
CDC42	Cell division control protein 42 homolog
CDK	Cyclin dependant kinase
CMP	Common myeloid progenitor
COPS5	COP9 signalosome subunit 5
CSM	Cell surface membrane
CSN	COP9 signalosome
CTSG	Cathepsin G
CXCL1	Chemokine (C-X-C motif) ligand 1
CXCR2	Chemokine (C-X-C motif) receptor 2
CYBB	Cytochrome b(558) subunit beta
CYLD	Cylindromatosis
DAP	Differentially abundant protein
DAPep	Differentially abundant modified peptide
dHL60	Differentiated HL60
DMSO	Dimethyl sulfoxide
DTX2	Deltex E3 ligase 2
DUB	Deubiquitinase
E2F1	Transcription factor E2F1
ELANE	Neutrophil elastase
ER	Endoplasmic reticulum
ERAD	ER-associated degradation
ERBD	Estrogen receptor binding domain
FA	Formic acid
FBXO11	E3 ubiquitin ligase FBXO11
FCER1G	High affinity immunoglobulin epsilon receptor subunit gamma FCER1G

Abbreviation	Definition
FCγR	High affinity immunoglobulin gamma Fc receptor
FDR	False discovery rate
FGR	Non-receptor tyrosine-protein kinase FGR
FSC	Forward scatter
FSC-A	FSC-Area
FSC-H	FSC-Height
G-CSF	Granulocyte-colony stimulating factor
GCSFR, CD114	G-CSF receptor
GFI1	Growth factor independence 1
GMP	Granulocyte-monocyte progenitor
GO	Gene ontology
GPCR	G-protein coupled receptors
GSDMD	Gasdermin D
GSEA	Gene set enrichment analysis
HCD	Higher-energy collisional dissociation
HL60	Human leukaemia cell line 60
HoxB8	Homeobox B8
HSC	Haematopoietic stem cell
IRF5	Interferon regulatory factor 5
IRF8	Interferon regulatory factor 8
ITCH	HECT-type E3 ubiquitin ligase ITCH
ITGAM, CD11b	Integrin alpha M
JAG1	Jagged-1
JAK	Janus kinase
JUNB	Transcription factor JUNB
KEGG	Kyoto Encyclopedia of Genes and Genomes
KIT	Receptor tyrosine kinase KIT
KO	Knockout
LC-MS/MS	Liquid chromatography tandem mass spectrometry
LCN2	Lipocalin-2
LDN	Low-density neutrophil
LEF1	Lymphoid enhancer factor 1
Log_2FC	Log_2(fold change)
LSP1	Lymphocyte specific protein 1
LTF	Lactoferrin
LYN	Lck/yes-related novel tyrosine kinase
MAPK	Mitogen-activated protein kinase
MF	Molecular function
MMP8	Matrix metalloproteinase-8
MPO	Myeloperoxidase
MS	Mass spectrometry
mTOR	Mammalian target of rapamycin
NADPH	Nicotinamide adenine dinucleotide phosphate

Abbreviation	Definition
NCE	Normalised collision energy
NEDD4	E3 ubiquitin-protein ligase NEDD4
NELFB	Negative elongation factor complex member B
NES	Normalised enrichment scores
NET	Neutrophil extracellular trap
NF-κB	Nuclear factor kappa-light-chain-enhancer of activated B cells
NLC	Neutrophil-like cell
NLRP3	NOD-, LRR- and pyrin domain-containing protein 3
NOX	NADPH oxidase complex
ORA	Over representation analysis
PCA	Principal component analysis
PCC	Pearson correlation coefficient
PCD	Programmed cell death
PFA	Paraformaldehyde
PFK2	6-phosphofructo-2-kinase
PI3K/Akt	Phosphoinositide 3-kinase/protein kinase B
PPP3C	Calcineurin
PTM	Post-translational modification
PTPN2	Protein tyrosine phosphatase non-receptor type 2
PU.1	Transcription factor PU.1
PYCARD or ASC	Apoptosis-associated speck-like protein containing a CARD
QC	Quality control
RA	Rheumatoid arthritis
RAC1	Ras-related C3 botulinum toxin substrate 1
RB1	Retinoblastoma
RELB	Transcription factor RELB
RHOA	Transforming protein RHOA
ROS	Reactive oxygen species
RUNX1	Runt-related transcription factor 1
SCF	Stem cell factor
SENP5	Sentrin-specific protease 5
sgRNA	Single guide RNA
SIRT3	NAD-dependent protein deacetylase sirtuin-3, mitochondrial
SLE	Systemic lupus erythematosus
SMG1	Serine/threonine-protein kinase SMG1
SPD	Samples per day
SSC	Side scatter
STAT	Signal transducer and activator of transcription
SUMO	Small ubiquitin-like modifier
SYK	Spleen tyrosine kinase
TAN	Tumour-associated neutrophil
TF	Transcription factor
TFA	Trifluoracetic acid

Abbreviation	Definition
TGM2	Transglutaminase 2
TIN	Tissue infiltrating neutrophil
TRIM33	E3 ligase TRIM33
TTK	Protein kinase TTK
Ub	Ubiquitin
UPS	Ubiquitin-proteasome system
USP	Ubiquitin-specific protease
VCP	Valosin-containing protein
VRK2	Serine/threonine-protein kinase VRK2
VST	Variance-stabilising transformation
WCNA	Weighted correlation network analysis
WGA	Wheat germ agglutinin
WT	Wildtype
YBX1	Y-box binding protein 1
YOD1	Deubiquitinase YOD1

Contributions

This thesis is submitted in support of my application for the degree of Doctor of Philosophy (DPhil) in Genomic Medicine and Statistics at the University of Oxford. The work presented herein was performed and written by myself, Mohammed A. Akbor, under the supervision of Prof. Benedikt Kessler, Prof. Irina Udalova, and Dr. Andreas Damianou. Individuals who contributed to the data presented in this thesis are as follows:

1. Dr. Ananda Mukherjee (Kennedy Institute of Rheumatology, University of Oxford) performed all of the HoxB8 cell culture and sample preparation steps and provided cell pellets and lysates for analysis.
2. Dr. Abhinandan Devaprasad (Kennedy Institute of Rheumatology, University of Oxford) provided pre-processed epigenomic and transcriptomic data for the HoxB8 model.
3. Dr. Iolanda Vendrell, Svenja Hester, Dr. Sarah Flannery, and George Berridge (Discovery Proteomics, Centre for Medicines Discovery, University of Oxford) operated the mass spectrometers for proteomic, ubiquitomic, and phosphoproteomic analyses.
4. Katie Holden (Cellular Imaging Core Facility, Centre for Human Genetics, University of Oxford) performed the high-throughput microscopy imaging and initial image processing.
5. Si Yui Charis Pang (University of Cambridge) contributed to Western blot validation of the candidate hits from the HL60 CRISPR-Cas9 DUB KO screen.

Chapter 1: Introduction

1.1. Neutrophil biology

Neutrophils, also referred to as polymorphonuclear leukocytes, are a type of white blood cell that primarily contributes to the innate immune system. They are generally the largest population of leukocytes, comprising approximately 50% of all white blood cells, and perform a variety of functional responses to different sterile and microbial stimuli (Figure 1.1). These activities include phagocytosis, reactive oxygen species (ROS) production, neutrophil extracellular trap (NET) generation, and cytokine production and detection. They are typically the first leukocytes recruited to sites of inflammation to clear pathogens and/or cellular debris. As such, this implicates neutrophils in both inflammation and the adaptive immune system. There is growing evidence that suggests neutrophils are transcriptionally active and can adapt an appropriate response, thus forming heterogenous populations[1]. Moreover, neutrophils are implicated in various immune disorders and diseases. However, the molecular mechanisms that underpin these activities and roles of neutrophils remain unclear.

1.1.1. Neutrophil effector functions

1.1.1.1. Neutrophil migration and infiltration

Circulating neutrophils are recruited to inflamed or infected tissue. Neutrophils express several G-protein coupled receptors (GPCRs) on their cell surface membrane (CSM) that allow detection of different chemokine classes, enabling chemotaxis[2]. Neutrophils must first detect and interpret extracellular inflammatory signals such as chemokines and cytokines. Neutrophils will tether, roll, and adhere to the walls of blood vessels in response to such inflammatory agents, priming the neutrophil for transmigration and tissue infiltration[3]. This involves various biomechanical actions, such as cytoskeletal rearrangement and regulation of adhesion molecules, that allow neutrophils to migrate across boundaries and invade tissue[4], where they subsequently perform their effector functions at target sites.

Neutrophils integrate various exogenous cues via intracellular signalling networks. This typically involves activation of nuclear factor kappa-light-chain-enhancer of activated B cells (NF-κB), phosphoinositide 3-kinase/protein kinase B (PI3K/AKT), mitogen-activated protein kinase (MAPK), Janus kinase/signal transducer and activator of transcription (JAK/STAT), and/or small GTPase signalling pathways[5,6]. These signalling pathways involve proteins that undergo expression and/or post-translational modifications (PTMs) to achieve an appropriate cellular response. For example, protein

phosphorylation, ubiquitination, acetylation, and methylation play key roles in determining the biological consequences of NF-κB signalling[7].

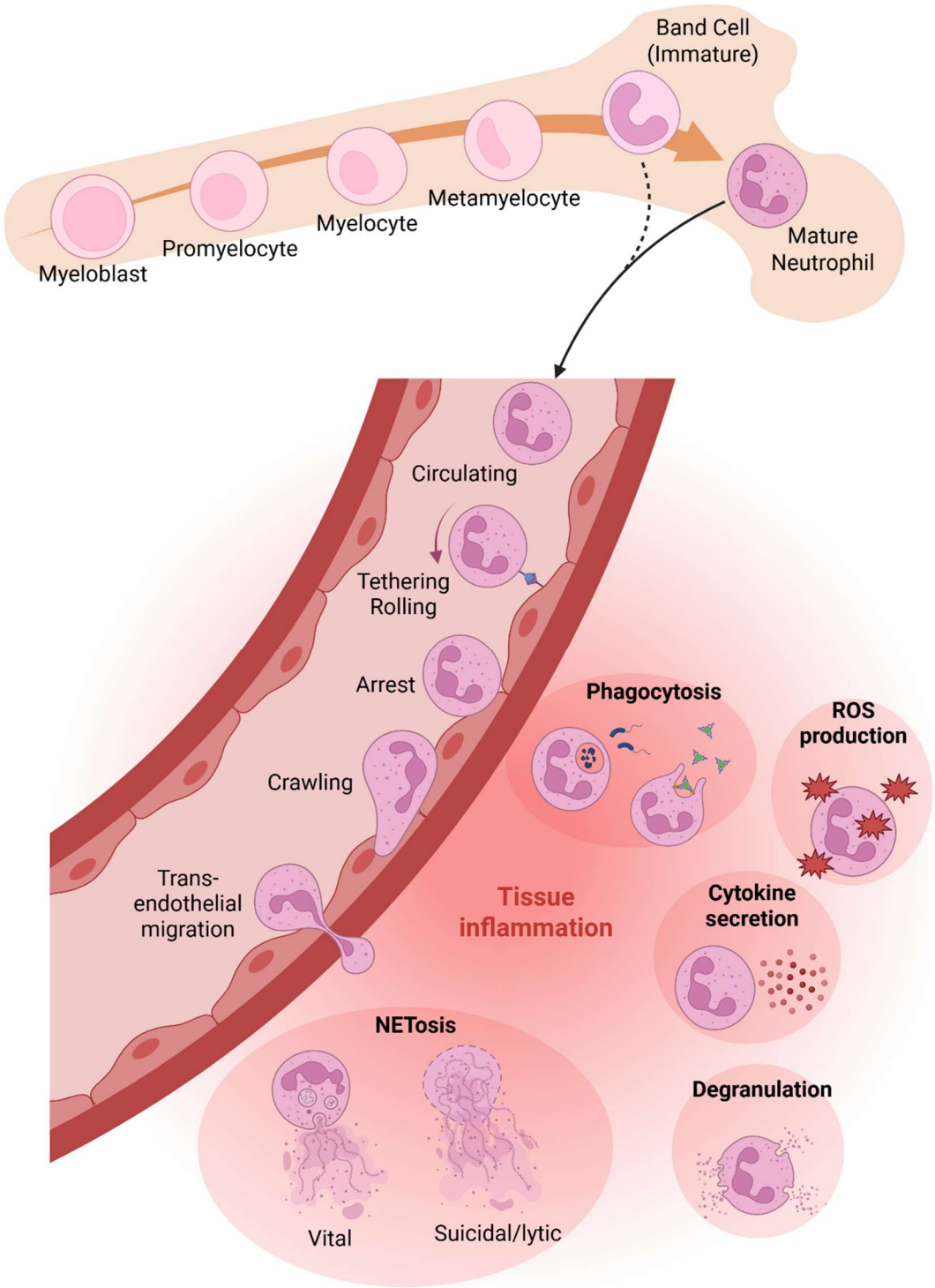


Figure 1.1 ***Overview of neutrophil development, circulation, and effector functions.*** Neutrophil differentiation occurs in the bone marrow. Typically, mature neutrophils egress from the bone marrow and enter circulation. In response to chemotactic stimuli, neutrophils migrate to sites of inflammation. Circulating neutrophils transmigrate across vascular epithelium, infiltrating the tissue space. In response to microbial threats and sterile challenges, neutrophil can perform several effector functions. This figure was created using BioRender.

As migratory cells, neutrophils typically become polarised and form a leading and lagging edge[8]. This involves the asymmetric distribution of cytoskeletal components that provides biomechanical forces and allows for directed cellular movement in response to chemotactic stimuli. Various pathways mediate the arrangement of cytoskeletal and signalling proteins, notably those involving the Rho family of small GTPases such as Ras-related C3 botulinum toxin substrate 1 (RAC1), cell division control protein 42 homolog (CDC42), and transforming protein RHOA[9-11].

1.1.1.2. Phagocytosis and antimicrobial activities

Upon recognition of a microbial or sterile challenge, neutrophils can engulf and eliminate pathogens and cellular debris. Neutrophils engage pathogenic targets through various receptors, leading to signalling pathways that promote phagocytosis. Exploiting their flexible morphology, neutrophils form a phagosome, a distinct organelle that engulfs and internalises pathogenic material. Phagosomes can then fuse with endosomal and lysosomal compartments, which contain digestive enzymes and reactive molecular species, resulting in the degradation of the phagosome contents. Again, this involves precise signalling, mediated by receptors, and coordinated biomechanical changes for successful phagocytosis[12,13].

Stimulation of neutrophils via phagocytosis, pathogenic material, or other immune stimuli can activate the nicotinamide adenine dinucleotide phosphate (NADPH) oxidase complex (NOX) and produce reactive oxygen species (ROS) that aid in the killing of phagocytosed or external threats[14]. Various reactions occur, resulting in a diverse range of ROS types that can target a broad spectrum of pathogens.

Moreover, neutrophils can produce anti-microbial peptides that have unique properties and can modulate immune response and signalling[15]. In combination with various proteins that perform antimicrobial activities, such as the proteolytic enzymes neutrophil elastase (ELANE) and cathepsin G (CTSG), the neutrophil has a diverse repertoire of agents at its disposal for eliminating threats.

Neutrophils are granulocytes, named so due to the presence of membrane-bound vesicles called granules. These specialised organelles contain combinations of antimicrobial and cytotoxic molecules and proteins important in neutrophil effector functions. Granules can either fuse with phagosomes to facilitate elimination of phagosomal cargo or neutrophils can release granule contents via exocytosis in order to destroy extracellular pathogens and coordinate inflammatory responses[16-18].

1.1.1.3. NETosis

Neutrophil extracellular traps (NETs) are a distinct antimicrobial behaviour exhibited by neutrophils. NETs consist of neutrophil-derived DNA (chromatin), RNA, and antimicrobial proteins, which are then

released from the neutrophil via a process known as NETosis, to capture pathogenic microbes. Neutrophils performing NETosis can either maintain cellular viability or undergo programmed cell death.

Classical or suicidal NETosis is a type of programmed cell death (PCD) that involves neutrophils releasing granule contents, decondensing chromatin, and permeabilising the cell surface membrane[19]. Distinct from other types of PCD, such as apoptosis, NETosis does not involve caspase activation. However, NETosis appears mechanistically similar to pyroptosis and necroptosis, maintaining features such as permeabilisation mechanisms and signalling pathways that lead to PCD[20,21].

Vital NETosis involves neutrophils releasing NETs and maintaining cell viability, resulting in neutrophil cells lacking nuclei yet capable of performing their effector functions, such as phagocytosis and chemotaxis[22].

1.2. Neutrophil differentiation

To become specialised cells and professional phagocytes, neutrophils must first undergo differentiation. Haematopoietic stem cells (HSCs) in the bone marrow develop into common myeloid progenitors (CMPs), then granulocyte-monocyte progenitors (GMPs), which eventually differentiate into neutrophils[23], as well as monocytes, dendritic cells, macrophages, eosinophils, or basophils. An overview of the neutrophil differentiation process, including key transcription factor (TF) and select protein marker/receptor changes, is depicted in Figure 1.2.

Transcriptional programs underpinning differentiation towards neutrophil lineages begin in progenitor cells. The CCAAT-enhancer-binding protein (C/EBP) family of TFs play a pivotal role in modulating transcriptional activity as neutrophils mature form GMPs. Lymphoid enhancer factor 1 (LEF1) expression in progenitor cells transactivates C/EBPα, which, in combination with growth factor independence 1 (GFI1), promote expression of granulocyte-specific genes, such as the granulocyte-colony stimulating factor (G-CSF) receptor (GCSFR, CD114)[24,25]. GFI1 also reinforces response to G-CSF via the MAPK signalling pathway[26]. C/EBPε expression is crucial during the preneutrophil/myelocyte stage, interacting with retinoblastoma (RB1) and transcription factor E2F1 proteins to prevent proliferation and encourage terminal differentiation, and is downregulated in mature neutrophils[27]. Closer to mature neutrophils, C/EBPα levels decrease while C/EBPβ and C/EBPδ expression increases. Transcription factor PU.1, C/EBPγ, and C/EBPδ are continuously expressed as neutrophils mature. Furthermore, C/EBPβ has been shown to be important in emergency granulopoiesis[28], the rapid production of neutrophils in response to severe inflammation or infection. Other TFs, such as runt-related transcription factor 1 (RUNX1) and interferon regulatory factor 8 (IRF8), modulate gene expression programs that determine neutrophil commitment[1,29,30]. Transcription factor RELB,

interferon regulatory factor 5 (IRF5), and transcription factor JUNB also demonstrate increased expression in mature neutrophils[1].

Despite the identification and characterisation of key transcription factors and target genes that are essential for neutrophil differentiation, there remains a limited understanding of the molecular details on the protein level that underpin neutrophil differentiation and maturation.

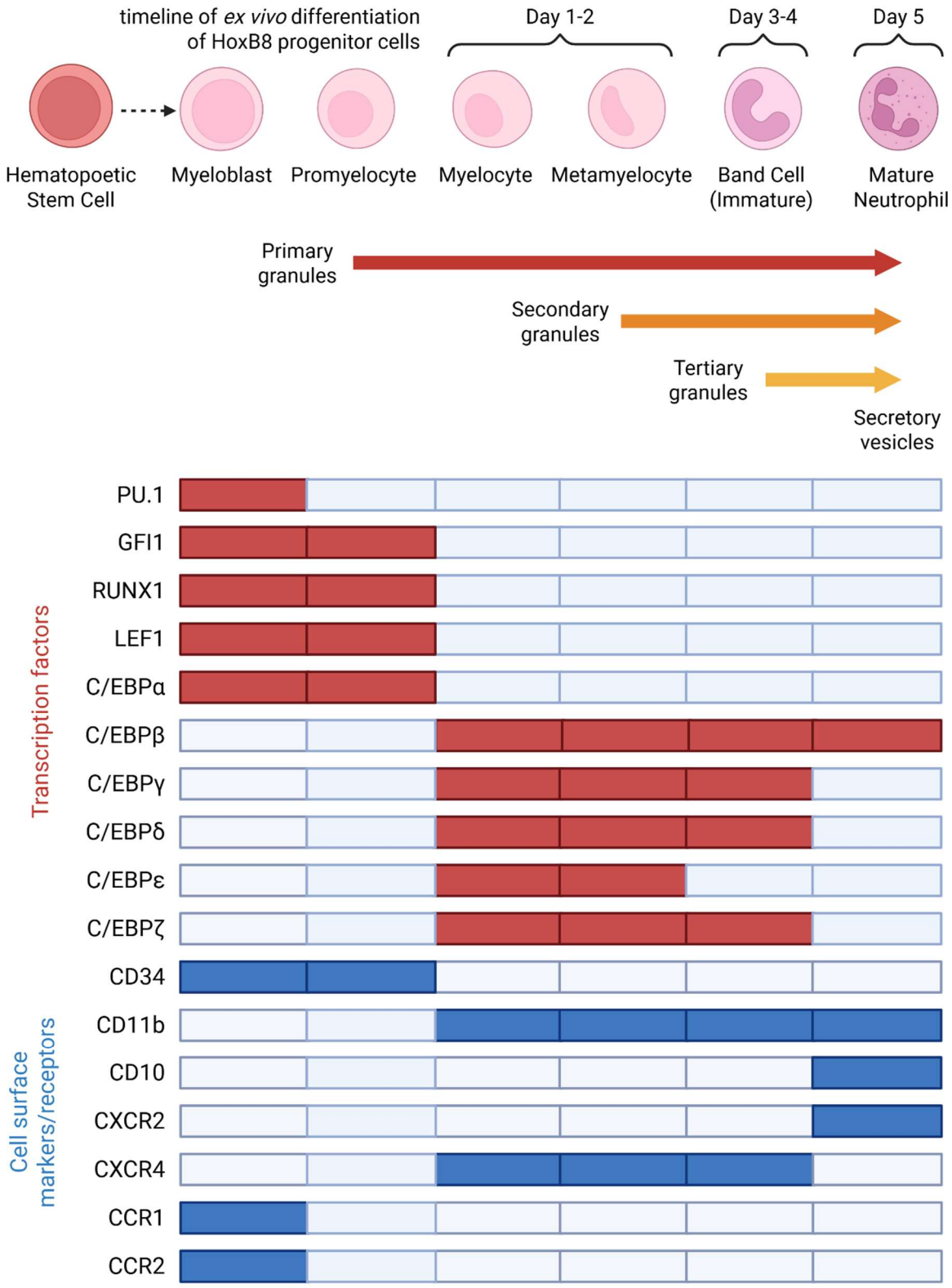


Figure 1.2 ***Stages of neutrophil differentiation and corresponding timeline of ex vivo differentiation of HoxB8 progenitor cells.*** Neutrophil differentiation primarily occurs in the bone marrow where hematopoietic stem cells (HSCs) undergo transcriptional and molecular changes towards granulocyte-monocyte progenitors (GMPs), subsequently myeloblasts, and finally mature neutrophils. Both band cells (immature) and mature neutrophils are capable of egress from the bone marrow. Key TFs participate in programmes that drive commitment towards the neutrophil lineage. The different stages of neutrophil differentiation are associated with the expression of protein markers and cell surface receptors to facilitate the biological functions of neutrophils. Coloured boxes indicate the stage in which proteins are more expressed/active. This figure was created using BioRender.

1.2.1. Neutrophil cell models of differentiation

Neutrophils are generally a complex cell type to examine due to their short lifespan and tendency to become activated in response to cellular debris and artificial extraction. The neutrophil differentiation and maturation process primarily occurs in the bone marrow, and so studying this process poses several challenges, from isolating sufficient material to recapitulating the environment in which neutrophil differentiation takes place. As such, *in vitro* and *ex vivo* cellular models are required[31].

1.2.1.1. HoxB8

A model system of immortalised mouse myeloid HoxB8 progenitor cells can be used to investigate neutrophil differentiation[32]. These cells contain a modified version of the TF homeobox B8 (HoxB8) that is regulated by an estrogen receptor binding domain (ERBD). Estradiol can bind to the ERBD and maintain HoxB8 myeloid progenitors in a proliferative state. Following estradiol removal and G-CSF-induced *ex vivo* differentiation for several days, the progenitor cells become non-proliferating neutrophils. The stages of *ex vivo* differentiation follow a similar timeline to primary neutrophils (Figure 1.2). The HoxB8 model provides a means to not only generate more biological material through expansion, but also to introduce genetic perturbations through molecular biology approaches such as lentiviral transduction[33]. This facilitates characterisation of the neutrophil differentiation process and subsequent neutrophil effector functions such as phagocytosis, NETosis, and migration, in physiologically relevant and genetically perturbed scenarios[34]. For example, HoxB8 neutrophils have been shown to recapitulate high-affinity immunoglobulin gamma Fc receptor (FCγR) and integrin signalling pathways in response to appropriate stimulation, leading to ROS production, degranulation, and chemotaxis, comparable to primary neutrophils[35].

1.2.1.2. HL60

The human leukaemia cell line HL60 shows neutrophilic promyelocytic morphology. Differentiation of HL60 cells can be induced by various chemicals, such as dimethyl sulfoxide (DMSO) and all-trans retinoic acid (ATRA), resulting in neutrophil-like cells (NLCs) that share functional properties with those of primary neutrophils[31,36]. Despite attenuated capabilities of HL60-derived NLCs compared to primary neutrophils, differentiated HL60 (dHL60) cells have been used to investigate several neutrophil behaviours, for example, chemotaxis[37], degranulation[38], NETosis[39], and phagocytosis[40]. HL60 cells are relatively easy to manipulate in culture and can be genetically modified to introduce perturbations[41,42], interferences[43], or overexpression/activation[44,45] of genes, through various molecular biology techniques such as lentiviral transduction and liposome transfection.

HL60 cells prove to be a valuable model of neutrophil biology for *in vitro* experimentation. Research involving HL60 cells has primarily focused on immune cell effector functions, comparing differentiation

protocols and robustness of the model to physiological neutrophils, or simply used HL60 cells as a model of promyelocytic leukaemia[46]. Characterisation of HL60 differentiation into NLCs has demonstrated changes at the transcriptional level, utilising approaches such as RNA-seq[47,48]. Time-resolved proteomic characterisation is the logical next step, being able to capture the diversity and complexity of proteoforms that may be relevant to neutrophil differentiation, and, more broadly, haematopoiesis. Recent work has focused on proteomic profiling of NETs produced by dHL60 cells[49]. Validation of effector proteins identified using such proteomic approaches would improve understanding of the underlying molecular mechanisms.

1.3. Post-translational protein modifications

Proteins are the most diverse biological macromolecule, owing to functional and structural variation that arises from different mechanisms such as alternative splicing. PTMs are alterations that are introduced on amino acid side chains or protein termini, leading to further diversity in structure and function (Figure 1.3). PTMs are typically added to proteins via enzymatic reactions. This can involve non-proteinaceous modifications, such as acetylation, methylation, glycosylation, and phosphorylation. In contrast, other peptides or protein components can be conjugated to a target protein. For example, ubiquitination and SUMOylation involve the conjugation of a small protein to a protein of interest, via an isopeptide bond.

PTMs can affect many aspects of protein characteristics, such as structure, function, cellular localisation, and stability. PTMs such as deamidation, citrullination, acetylation, and methylation have been demonstrated to be induced by pathological stimuli in neutrophils, regulating distinct signalling pathways[50].

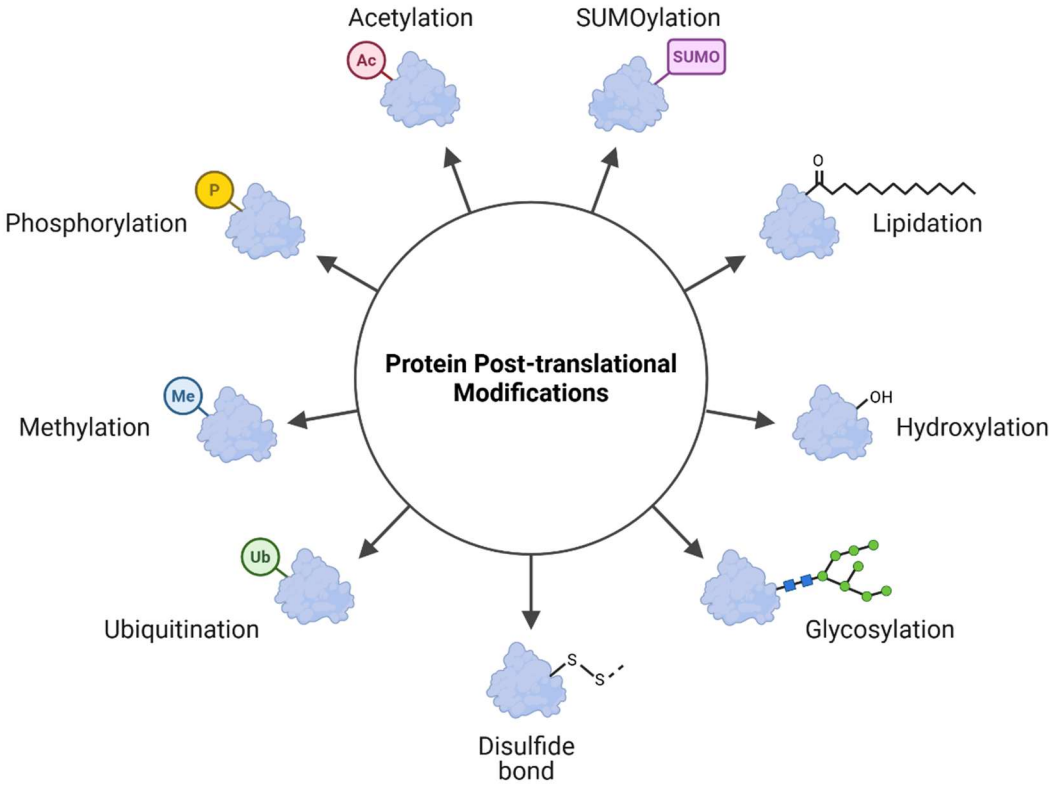


Figure 1.3 ***Examples of protein post-translational modifications (PTMs).*** This figure was created using BioRender.

1.3.1. The ubiquitin system

1.3.1.1. The role of ubiquitination in protein stability and turnover

Ubiquitin (Ub) is a small protein that is expressed in virtually all eukaryotic cells and can be conjugated to other proteins and itself, forming either monoubiquitinated sites or Ub chains. Conjugation to target proteins occurs via an enzymatic cascade consisting of E1 ubiquitin-activating, E2 ubiquitin-conjugating, and E3 ubiquitin-protein ligase enzymes, whereby Ub becomes covalently linked to typically lysine side chains (Figure 1.4A). Further ubiquitination can occur on already-conjugated Ub, resulting in chains with diverse architectures and constituents; additional ubiquitination can occur on several lysine residues on the surface of ubiquitin protein, and ubiquitin-like proteins such as small ubiquitin-like Modifier (SUMO) can be conjugated to proteins or incorporated into chains. These variable Ub modifications can be cleaved from protein substrates via deubiquitinases (DUBs). The opposing activities between E3 ligases and DUBs can determine the ubiquitination status of a protein and its fate, through modulated structure, function, and stability.

Ubiquitination is a PTM that typically marks a protein for degradation, but can also affect a protein's cellular localisation, activation status, and protein-protein interactions, depending on where the

ubiquitination is attached and the type of Ub linkages and chains present on the protein (Figure 1.4B). Polyubiquitin chains can consist of single or mixed linkage types, in linear or branched configurations, and certain linkage types are generally associated with particular biological consequences[51,52]. For example, K11 and K48 Ub linkages are typically associated with protein degradation, K6, K27, and K63 with DNA damage, and K27 and K33 with cellular signalling. K27 and K63 are also involved in anti-viral immunity and anti-bacterial autophagy, respectively[53]. Utilisation of different linkage types and patterns acts as a code that can be deciphered and interpreted by other cellular components to bring about an appropriate response. Furthermore, although lysine side chains are the canonical site for Ub conjugation, there is a newfound appreciation of non-lysine ubiquitination on serine, cysteine, and threonine side chains and N-terminal amino groups, as well as ubiquitination of non-proteinaceous substrates, all of which may have diverse biological significance[54].

1.3.1.2. Proteasome- versus lysosome- mediated degradation

The primary role of protein ubiquitination is to target proteins for degradation. This can occur through two main degradation pathways: the proteasomal and lysosomal pathways (Figure 1.4C). The latter can be further divided into the endolysosomal and autophagosome pathways. Ubiquitination plays fundamental roles in protein degradation, irrespective of the degradation pathway[55]. Protein degradation facilitates the recycling of protein components, as well as maintaining homeostatic balance within the cell.

The proteasome (26S) is a large protein complex consisting of different protein subunits forming a core (20S) and a regulatory (19S) particle. Proteins marked for degradation (typically K48 polyubiquitination) are unfolded and inserted into the core particle, where multiple proteolytic sites are found. Alternatively, the lysosomal pathway involves fusion with lysosomal compartments that contain proteolytic enzymes. Cargo marked for degradation via the lysosomal pathway can either originate from endocytosis, where proteins in the plasma membrane can be recycled, or from autophagosomes, organelles that form around protein aggregates or other organelles, which can be otherwise incompatible with the proteasomal or endolysosomal pathways[56].

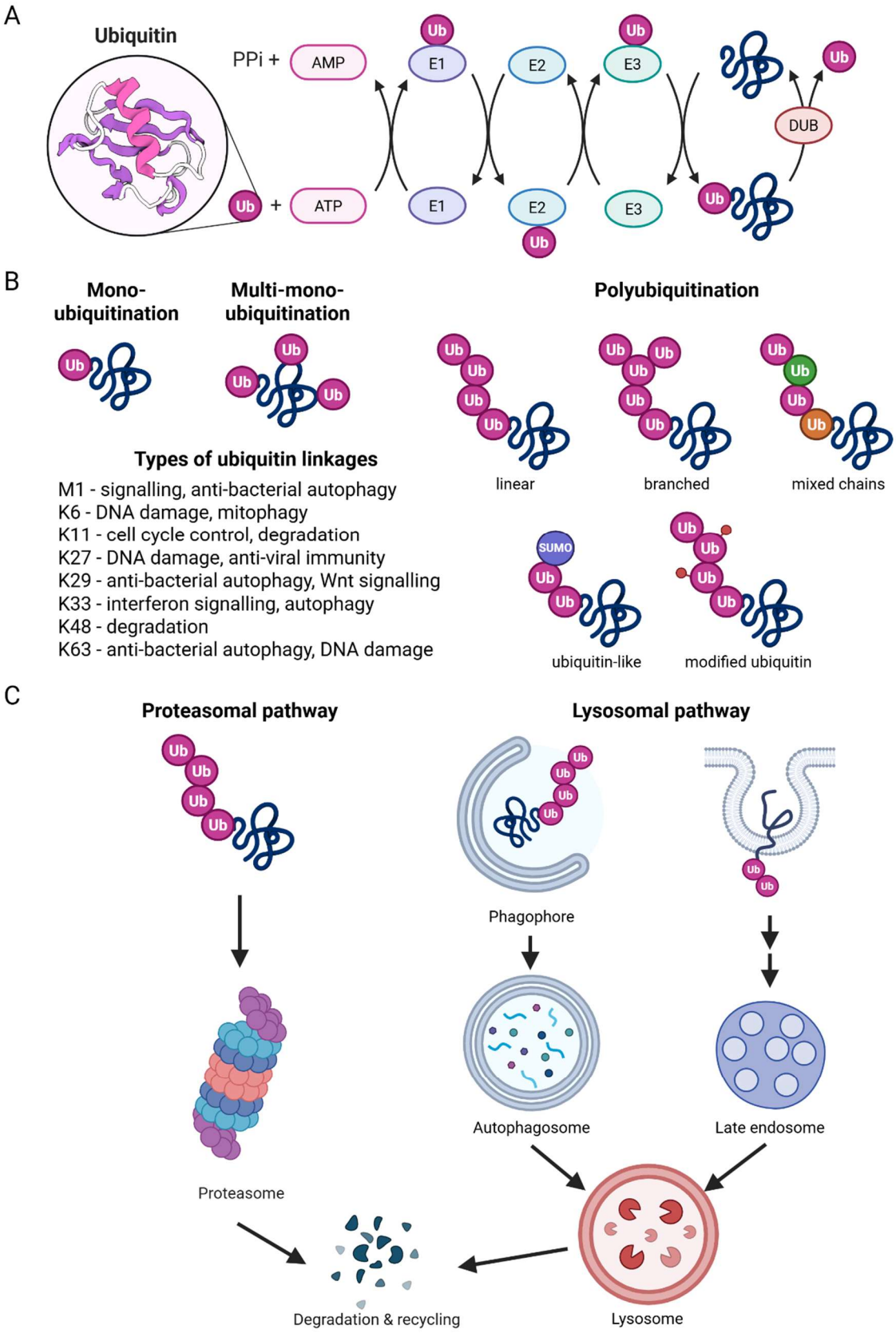


Figure 1.4 ***Ubiquitination is a complex post-translational modification.*** (**A**) Ubiquitin is conjugated to protein targets via an enzymatic cascade, ending in E3 ubiquitin-protein ligases. Ubiquitin is removed form substrates via deubiquitinases (DUBs). (**B**) Ubiquitin can be added to proteins in various configurations and combinations. Different linkage types can lead to various biological consequences. (**C**) Degradation-associated ubiquitination can involve the proteasomal or lysosomal degradation pathways. This figure was created using BioRender.

1.3.1.3. Targeting the ubiquitin system and deubiquitinases for therapeutic intervention

The ubiquitin-proteasome system (UPS) and its abnormal activity have been linked to several pathologies, including neurological and neoplastic diseases. Inhibitors of the proteasome, such as Bortezomib, showed initial success in the treatment of multiple myeloma and mantle cell lymphoma[57]. However, prolonged use can lead to toxicity and drug resistance, and drug usage was limited to haematological malignancies. Due to the ubiquitous nature of the UPS in virtually all cell types, targeting and inhibiting the proteasome is likely to have undesirable side effects.

DUBs are a large family of proteases that can recognise and cleave Ub from ubiquitinated substrates to regulate their behaviours and associated cellular pathways. The main classes of DUB enzymes include ubiquitin carboxyl-terminal hydrolases (UCH), ubiquitin-specific proteases (USP), ovarian tumour domain proteins (OTU), Josephin Ataxin-3–like proteins (MJD), MIU-containing new DUB family (MINDY), zinc-finger ubiquitin protease 1 (ZUP1/ZUFSP), and JAB1/MPN/Mov34 metalloenzymes (JAMMs)[58]. DUB activity is dependent on several factors, such as protein-protein interactions, subcellular localisation, expression levels, PTMs, and response to intracellular and extracellular cues[59,60]. DUBs may also show preference for specific substrates, types of ubiquitin linkages, and/or ubiquitin chain architectures[58,61], the details of which are unclear for many DUBs. Moreover, DUBs are essential in maintaining the available Ub pool by processing Ub precursors and recycling Ub chains, necessary for cellular homeostasis and response to stimuli[62,63].

DUBs have recently become attractive drug targets relevant to various diseases[58,64]. Genetic perturbation and/or chemical inhibition of DUBs have been shown to alleviate pathological phenotypes in various disease contexts such as cancer[65]. Despite the limited understanding of DUB substrate specificity for the majority of DUBs[66], selective inhibitors for certain DUBs have been developed, some of which are currently in clinical trials[67]. Unique structural features and defined catalytic domains within a DUB facilitate the development of selective and potent chemical inhibitors, providing opportunity for the advancement of therapeutic agents for DUB-associated disease contexts. Furthermore, drug selectivity can leverage tissue- or cell-specific expression or activity of a target DUB[68]; for example, a specific DUB or DUB variant may exact its pathological activity within a certain type of cell, and so its inhibition may reduce the risk of unwanted off-target effects.

Furthermore, new technologies that leverage the functions of Ub processing enzymes demonstrate innovative therapeutic opportunities to target proteins for which effective drugs are unavailable or challenging to develop. Molecular glue degraders and proteolysis-targeting chimeras (PROTACs) facilitate the interaction between a protein substrate and E3 ligase, promote substrate ubiquitination, and enable protein degradation via the UPS system[56]. In contrast, deubiquitinase-targeting chimeras

(DUBTACs) stabilise their target protein via the recruitment of DUB(s) that remove degradative ubiquitination[69]. Development of these molecules presents novel approaches in alleviating disease phenotypes by modulating aberrant, pathological protein function or stability[70].

1.3.2. The phosphorylation system

1.3.2.1. The role of phosphorylation in protein regulation and function

Protein phosphorylation can result in diverse biological responses (Figure 1.5). Phosphorylation typically occurs on serine, threonine, and tyrosine residues within proteins, introducing negative charges that can affect the local and global conformation of a protein. Phosphorylation can occur on histidine and aspartate residues; however, this is less common due to instability. Phosphorylation can have context- and protein-specific consequences, such as altering the activation status of a protein or modulating protein-protein interactions[71,72]. In some instances, protein phosphorylation can affect the stability of a protein and potentially mediate its degradation[73].

Similar to the Ub system, there are PTM enzymes that facilitate the addition or removal of phosphate groups to amino acid side chains. Kinases catalyse protein phosphorylation, while phosphatases cleave phosphate groups from protein substrates. These processes have been shown to occur in an enzyme-substrate specific manner, where enzymes recognise amino acid motifs contained within protein targets and deposit or remove phosphate groups[74].

Kinases participate in several cellular processes such as cell cycle regulation, cellular growth, division, and proliferation. For example, cyclin-dependant kinases (CDKs) are key enzymes that participate in phosphorylation cascades and signalling pathways that control cell cycle stage[75]. During differentiation, these pathways can be modulated through various means such as extracellular cues binding to surface receptors, resulting in signal transduction that fine-tunes appropriate pathway activation to enable cell specialisation programmes and modulate cell potency. The balance between kinase and phosphatase activity is necessary to determine the fate of protein activities within the cellular environment suitable to the developmental stages of a cell.

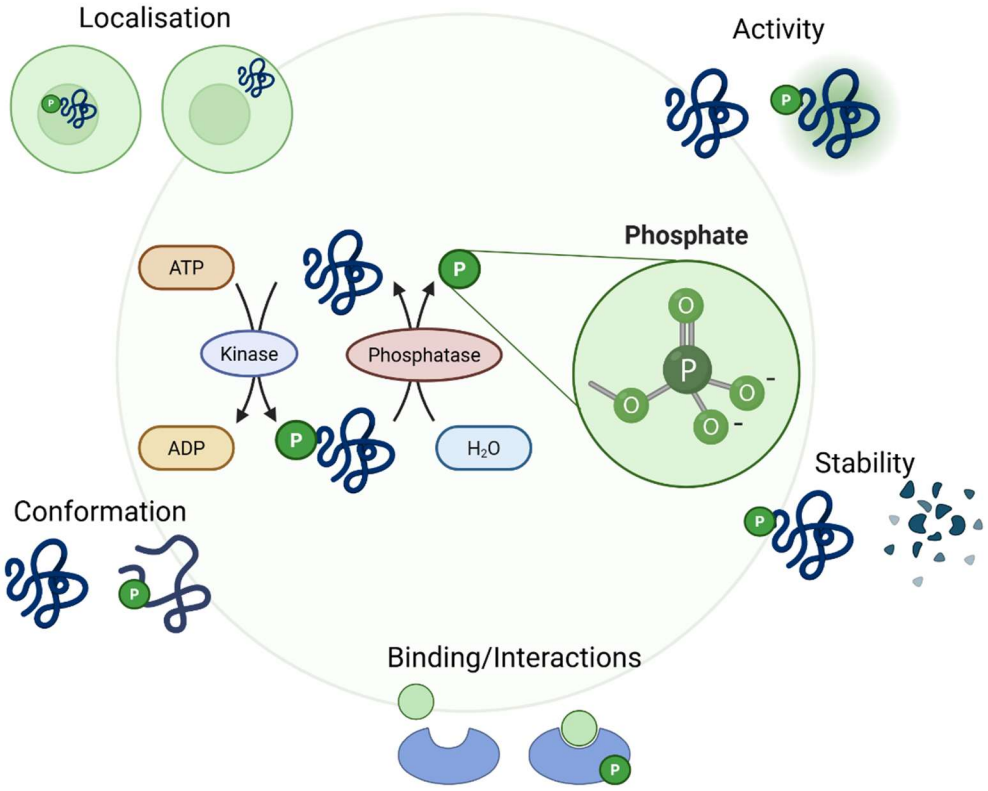


Figure 1.5 ***Phosphorylation can modulate protein structure and function.*** Phosphate groups are covalently attached to amino acid side chains, typically serine, threonine, and tyrosine, via the action of kinases, and removed through phosphatases. Phosphorylated proteins can show diverse biological consequences. This figure was created using BioRender.

1.3.3. The role of PTM in neutrophil biology

1.3.3.1. Ubiquitination

There are several biological processes involved in neutrophil biological functions wherein protein ubiquitination contributes.

Neutrophils are migratory cells. They require cytoskeletal remodelling to allow them to tether, roll, adhere, crawl, and transmigrate across epithelial membranes to infiltrate inflamed tissues in response to chemotactic signals[3]. These biomechanical processes require coordination of several structural and functional proteins, such as actin/actin-associated proteins and matrix-digesting proteases[76], respectively. These functional responses can, in part, result from the activation of signalling pathways that involve small GTPases[4,77]. Studies have shown ubiquitination to play a key role in regulating small GTPase activity[78]. For example, ubiquitination of RHOA and RAC1 can regulate protein abundance and consequently affect cytoskeleton dynamics and cell migration[79,80].

Neutrophils become activated in the presence of different stimuli such as cytokines and pathogens. One study demonstrated increased ubiquitin-specific protease (USP) 17 expression upon IL-8

treatment in neutrophils[81], suggesting the UPS may play a role in neutrophil activation. USP17 was shown to regulate chemokine-induced subcellular relocalisation of the small GTPases RHOA, RAC, and CDC42[81]. The DUB OTU domain-containing protein 1 (OTUD1) was identified to regulate the polarisation and secretory phenotype of tissue-infiltrating neutrophils (TINs) in the context of periodontitis[82].

Neutrophils release NETs as a means to sequester pathogenic threats. Ubiquitination was demonstrated to play a role in NETosis. The degradation of peptidylarginine deiminase 4 (PAD4) through increased levels of ubiquitination and targeting to the proteasome prevented NETosis, as observed in a pyelonephritis mouse model of *Escherichia coli* infection[83]. The role of PAD4 ubiquitination in NETosis inhibition was also observed in the context of sepsis[84]. The association between (defects in) ubiquitination and NETosis has been investigated in the context of other human disease[85], with examples including proliferative glomerulonephritis and systemic lupus erythematosus (SLE). Protein components of NETs such as myeloperoxidase (MPO) were demonstrated to be ubiquitinated, and these NETs or free ubiquitin were shown to stimulate monocyte-derived macrophages and modulate cytokine production[86]. The release of ubiquitinated proteins during NETosis has been suggested to play a role in immune signalling.

Neutrophil cells naturally consist of an abundance of specialised compartments known as granules, which are functionally similar to lysosomes. These compartments contain hydrolytic enzymes and ROS-producing enzymes, which can be released into the extracellular environment in response to microbial threats. Ubiquitinated proteins are detected in primary (azurophilic) granules, but not secondary (specific) granules in mature neutrophils[87]. Ubiquitination may be a means for neutrophils to differentiate between different granule types. How ubiquitination contributes to granule functions and neutrophil differentiation remains unclear.

Key stem cell transcription factors such as c-Myc, Oct3/4, Sox2, and Klf4 have been shown to be regulated through ubiquitination[88]. The UPS also contributes to regulating stem cell function in muscle tissue[89] and bone formation[90]. Within the context of red blood cells, the E3 ubiquitin ligase FBXO11 was shown to facilitate erythropoiesis through degradation of bromo adjacent homology domain containing 1 (BAHD1), inhibiting transcriptional repression of erythroid genes[91]. A study identified a SUMO protease (the DUB equivalent for SUMOylated proteins) affecting ATRA-induced differentiation of an acute myeloid leukaemia (AML) cell model. NB4 is a human acute promyelocytic leukaemia (APL) cell line that is used in myeloid research. Stable knockdown of sentrin-specific protease 5 (SENP5), a deSUMOylating enzyme, in NB4 cells demonstrated decreased CD11b expression and cell viability upon ATRA-induced differentiation[92].

1.3.3.2. Phosphorylation

Phosphorylation has been implicated in various contexts associated with neutrophil biology, from cell signalling pathways to transcriptional regulation. For example, stimulation of neutrophils with G-CSF leads to activation of JAK/STAT, MAPK, PI3K/AKT, and mammalian target of rapamycin (mTOR) signalling pathways[93]. The JAK/STAT pathway participates in innate immune functions, response to cytokines and interferons, and is involved in neutrophil activation[94]. The mTOR signalling pathway, which involves phosphorylation of several key proteins such as AKT, protein kinase C, and other kinases[95], was shown to affect NET formation in activated neutrophils[96]. Other pathways involved in innate immune signalling include the NF-κB signalling pathway, which consists of several kinases participating in inflammatory signal transduction[97]. NF-κB signalling was demonstrated to be involved in neutrophil activation and gene expression[98]. Furthermore, a family of essential regulatory TFs in neutrophil differentiation and function are the C/EBP family. Most of these TFs have been previously reported to be phosphorylated, with more studies focusing on C/EBPβ phosphorylation. These phosphorylation events are thought to regulate C/EBP DNA binding[99-102]. C/EBPβ is found downstream of MAPK pathway proteins, which can also elicit various transcriptional responses in neutrophils to a variety of stimuli involving protein phosphorylation events[103].

Moreover, 6-phosphofructo-2-kinase (PFK2) can be activated through protein phosphorylation, and its association with NOX2 – a protein complex involved in ROS production and innate immunity – was seen to affect glycolysis and support energy metabolism[104,105]. These are a few examples of how protein phosphorylation contributes to neutrophil effector functions and biological processes.

1.4. Neutrophil pathophysiology

Due to their abundance and specialised functions, neutrophils are essential cells in health and physiology. Aberrations to their usual functions and behaviours can have detrimental systemic effects leading to disease and disorder (Figure 1.6).

Neutropenia occurs when there is a reduction in the levels of functional neutrophils within the body due to changes in neutrophil production, elimination, distribution in the body, or a combination thereof. This leads to an increased susceptibility of an individual to pathogens and infection. Neutropenia is a complex disease that may be congenital, potentially caused by germline or *de novo* mutations, or acquired through the development of autoimmune disorders, nutritional deficiencies, or infection[106]. Due to the complexity of neutropenia pathogenesis, genetic causes or predisposition may not explain all cases of neutropenia, and may involve epigenetic regulation, transcriptional dysregulation, or proteomic alterations that contribute to disease development.

Within the context of cancer and tumour biology, tumour-associated neutrophils (TANs) have been shown to participate in the tumour microenvironment. Mature neutrophils promote anti-tumour activities through clearance of cellular debris and immune signalling to recruit other immune cell types, such as T-cells, to cancerous tissue. Contrarily, immature neutrophils exhibit immunosuppressive activities and hinder anti-tumour activities. Deltex E3 ligase 2 (DTX2) in hepatocellular carcinoma was shown to promote recruitment of TANs exhibiting tumour-promoting phenotypes[107], indicating interplay between the UPS, cancer biology, and immune modulation. Neutrophils have also been shown to be reprogrammable by mesenchymal cells, leading to an immunosuppressive state thereby promoting metastisis[108].

Characteristic of a leukocyte, neutrophils participate in the development and exacerbation of various immune and autoimmune disorders and diseases. Neutrophils are highly abundant in the body and are recruited to sites of inflammation. Upon activation, they release anti-pathogenic and immunomodulatory agents. However, excessive or inappropriate activation can lead to unintended detrimental effects on nearby tissues. A subset of neutrophil populations called ‘low-density neutrophils’ (LDNs) were identified in patients with rheumatoid arthritis (RA) and SLE. These LDNs have characteristics of both immature and mature neutrophils, leading to abnormal neutrophil behaviours, notably altered NETosis, ROS production, and apoptosis[109]. NETs are typically associated with sequestration and protection against pathogenic microbes. However, over-production and release of NETs can have adverse physiological effects in autoimmune disorders, leading to thrombosis, ‘cytokine storms’, and collateral tissue damage[19].

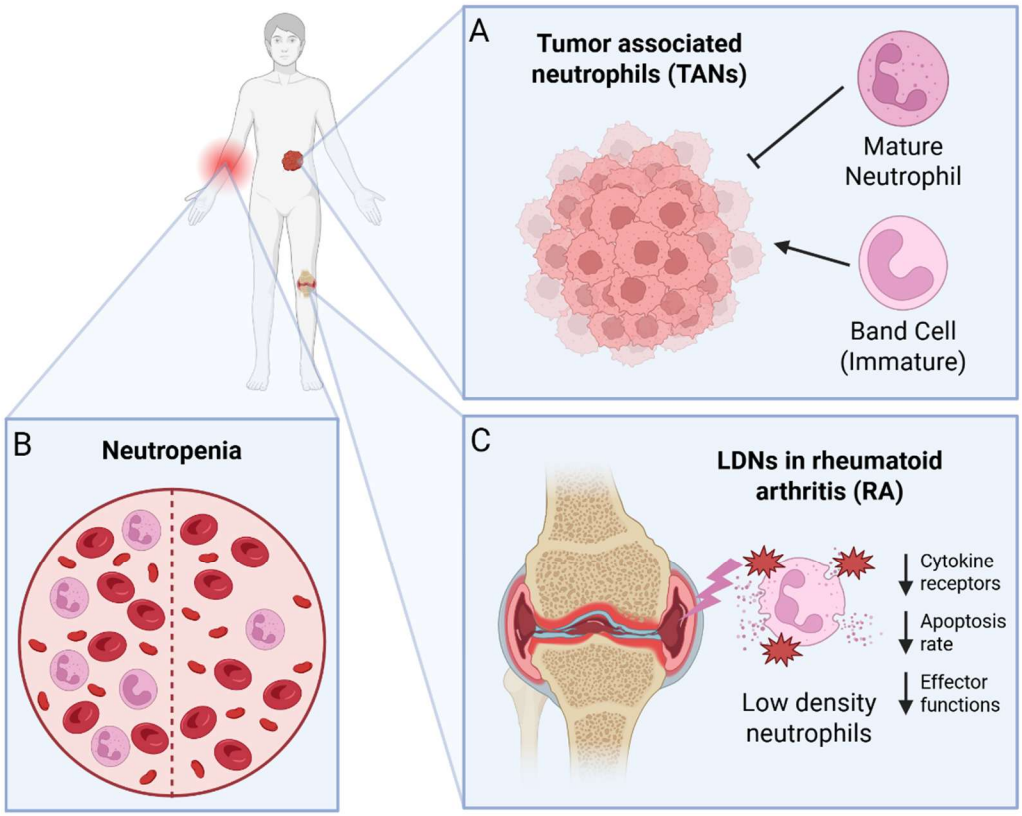


Figure 1.6 ***Neutrophil roles in disease and disorder.*** (**A**) Neutrophils can be associated with the tumour microenvironment in cancer. Mature neutrophils show anti-tumour behaviours, while immature neutrophils demonstrate immunosuppressive, tumour promoting behaviours. (**B**) Reduction in circulating neutrophils leads to neutropenia (Left is healthy, right is neutropenic). (**C**) Low density neutrophils are associated with autoimmune disorders such as rheumatoid arthritis. LDNs have characteristics of both mature and immature neutrophils, leading to abnormal effector functions. Excessive activation of neutrophils can also cause collateral damage to surrounding tissue.

1.5. Thesis aims

This thesis aimed to explore the molecular details of neutrophil differentiation, primarily focusing on the roles PTMs such as phosphorylation and ubiquitination. Investigating the role of PTMs in neutrophil differentiation can provide better understanding of how proteins are regulated and altered, potentially identifying new mechanisms through which neutrophil differentiation occurs. This could provide opportunities to target specific proteins or pathways, modulating their activities for therapeutic benefit. Moreover, in the context of ubiquitination, the extent of DUB contribution to neutrophil development was also examined. Accordingly, the aims of this thesis were as follows:

1. To characterise the proteomic changes that occur during neutrophil differentiation, using the HoxB8 model of neutrophil differentiation and MS-based approaches.
2. To explore the ubiquitomic and phosphoproteomic changes that occur during HoxB8 neutrophil differentiation, using modified peptide enrichment paired with MS-based approaches.
3. To integrate epigenomic, transcriptomic, proteomic, and ubiquitomic data from the HoxB8 neutrophil differentiation model to identify potential regulatory relationships between different modalities, using a systems biology approach.
4. To identify candidate DUBs that may regulate neutrophil differentiation, using the HL60 neutrophil model and a CRISPR-Cas9 KO phenotypic screen.

Chapter 2: Materials and Methods

2.1. Materials

Table 2.1 ***List of cell culture reagents.***

Reagents	Identifier	Source
HL60	CCL-240	ATCC
RPMI-1640 medium	R5886	Sigma-Aldrich
DMEM – high glucose medium	D5671	
Foetal Bovine Serum	F4135	
L-glutamine	G7513	
Sterile-filtered DMSO	D2438	
β-estradiol	E2758	
G-CSF	G8160	
Puromycin	P4512	
Polybrene Infection / Transfection Reagent	TR-1003	
Dulbecco's Phosphate Buffered Saline	D8537	
Blasticidin S Ready Made Solution	SBR00022	
FuGENE® HD Transfection Reagent	HD-1000	FuGENE®
Lenti-X™ GoStix™ Plus	631280	Takara Bio
Lenti-X™ Concentrator	631232	

Table 2.2 ***List of plasmids and molecular biology reagents.***

Reagents	Identifier	Source
LentiArray™ Human CRISPR Library	A32167 – A32185 (inclusive)	Invitrogen
psPAX2	12260	addgene
pMD2.G	12259	
QIAGEN Plasmid Plus Midi kits	12943	QIAGEN
QIAGEN Plasmid Plus 96 kits	16181	
LB Broth (Lennox)	L3022	Sigma-Aldrich
Ampicillin sodium salt (Ready Made Solution)	A5354	

Table 2.3 ***List of reagents and materials for Western blot analysis.***

Reagents	Identifier	Source
Pierce™ BCA Protein Assay Kit	23225, 23227 and A65453	ThermoFisher
Nitrocellulose Membrane, 0.45 µm	88018	Life Technologies
cOmplete™, EDTA-free Protease Inhibitor Cocktail	COEDTAF-RO	Roche
PhosSTOP™	PHOSS-RO	
NuPAGE™ 4 to 12%, Bis-Tris, 1.0 mm, Midi Protein Gel	WG1402BOX	Invitrogen
NuPAGE™ LDS Sample Buffer (4X)	NP0008	
NuPAGE™ MES SDS Running Buffer (20X)	NP000202	
NuPAGE™ Transfer Buffer (20X)	NP00061	

Table 2.4 ***List of antibodies used for Western blot and flow cytometry.***

Reagents	Identifier	Source
β-Actin Antibody	4967	Cell Signalling Technologies
CD11b/ITGAM (D6X1N) Rabbit mAb	49420	
CD11b/ITGAM (E6E1M) Rabbit mAb	17800	
SirT3 (D22A3) Rabbit mAb	5490	
CD34 (E2J1K) Rabbit mAb	26233	
RhoA (67B9) Rabbit mAb	2117	
TGM2 (D11A6) XP® Rabbit mAb	3557	
Caspase-1 (E9R2D) Rabbit mAb	83383	
NLRP3 (D4D8T) Rabbit mAb	15101	
FCER1G Polyclonal Antibody	PA5-115222	ThermoFisher
IRDye® 680RD Goat anti-Mouse IgG Secondary Antibody	926-68070	Li-COR
IRDye® 800CW Goat anti-Rabbit IgG Secondary Antibody	926-32211	
CD11b Antibody, anti-human, APC, REAfinity™	130-110-612	Miltenyi Biotec
CD182 (CXCR2) Antibody, anti-human, APC, REAfinity™	130-120-133	
CD184 (CXCR4) Antibody, anti-human, APC, REAfinity™	130-120-778	
REA Control Antibody (S), human IgG1, APC, REAfinity™	130-113-434	

Table 2.5 ***List of reagents and materials for mass spectrometry sample preparation and analysis.***

Reagents	Identifier	Source
Trypsin, TPCK treated	20233	ThermoFisher
Acclaim™ PepMap™ 100 C18 HPLC Columns, 300µm x 5mm, 5µm particle size	160454	
S-Trap™ micro spin column	C02-micro	Protifi LLC
S-Trap™ midi spin column	C02-midi	
AssayMap 5 µL Titanium Dioxide TiO2 cartridges	G5496-60016	Agilent
PTMScan® Ubiquitin Remnant Motif (K-ε-GG) Kit	5562	Cell Signalling Technologies
Acclaim™ PepMap™ 100 C18 HPLC Column, 300µm x 5mm, 5µm particle size	160454	ThermoFisher
15 cm x 150 µm x 1.5 µm analytical column	EV1137	Evosep

Table 2.6 ***List of miscellaneous reagents and materials.***

Reagents	Identifier	Source
NaCl	S9888	Sigma-Aldrich
Glycerol	G9012	
Triton™ X-100	X100	
DL-Dithiothreitol (DTT)	D0632	
Sodium deoxycholate	D6750	
Sodium lauroylsarcosinate	Y0001772	
Benzonase® Nuclease	E1014	
SDS, 20% Solution, RNase-free	AM9820	Invitrogen
EDTA (0.5 M), pH 8.0, RNase-free	AM9260G	
HEPES	15630080	Gibco
Tris hydrochloride (Tris-HCl)	10812846001	Roche
EGTA, Molecular Biology Grade	324626	Millipore
NP-40 lysis buffer	J60766-AP	ThermoFisher
RIPA Lysis and Extraction Buffer	89900	
Wheat Germ Agglutinin Alexa Fluor 647	W32466	
Hoechst 33342, trihydrochloride trihydrate	H3570	

2.2. Cell culture

2.2.1. HL60 culturing

HL60 cells (parental line: CCL-240, ATCC) were cultured in RPMI-1640 medium containing L-glutamine (R8758, Sigma-Aldrich) supplemented with 10% FBS (F4135, Sigma-Aldrich) at 37°C in a humidified, 5% CO_2 incubator. Stable transfected cell lines (HL60-Cas9 and HL60-Cas9-KO) were cultured in medium supplemented with 10µg/ml of blasticidin (SBR00022, Sigma-Aldrich), and HL60-Cas9-DUB KO cell culture medium additionally contained 10µg/ml puromycin (P4512, Sigma-Aldrich), for selection and maintenance.

2.2.1.1. HL60 differentiation

HL60 cells with >95% viability in the logarithmic growth phase were collected, counted, and pelleted by centrifugation. Cells were washed once with PBS and resuspended in complete medium supplemented with 1.3% sterile-filtered DMSO (D2438, Sigma-Aldrich) at a density of 5×10^5cells/mL. Cultures were maintained for at least 5 days (unless otherwise stated) to allow differentiation. Samples were collected at daily intervals; cells were washed in PBS, and either pelleted and stored at -80°C or fixed using 3.2% paraformaldehyde (PFA).

2.2.2. HoxB8 proliferation, culturing, and differentiation

HoxB8 C57BL/6N myeloid progenitors[32] were received from Dr. David Sykes's group at the Harvard Stem Cell Institute, Harvard University, and cultured by Dr. Ananda Mukherjee from the Udalova Group (The Kennedy Institute of Rheumatology, University of Oxford), following established protocols[110]. Culture medium consisted of RPMI-1640 medium supplemented with 10% FBS, 30mM β-mercaptoethanol (21985023, Gibco), 1% supernatant from stem cell factor (SCF)-producing CHO cells, and 1µM β-estradiol (E2758, Sigma-Aldrich). Passage numbers between 2 and 5, inclusive, from fresh revival were used for experiments, with biological replicates originating from separate passages. HoxB8 differentiation was induced by removal of β-estradiol and addition of 20ng/mL G-CSF (G8160, Sigma-Aldrich) to culture medium. Cells were cultured at 37°C in a humidified, 5% CO_2 incubator, and samples were collected at daily intervals; cells were washed in PBS, pelleted, and stored at -80°C, ready for downstream analyses.

2.2.2.1. Nuclear enrichment

All steps for nuclear enrichment were performed at 4°C or on ice, and buffers were prepared in H_2O, unless otherwise stated.

The cytoplasmic lysis buffer was prepared with 50mM HEPES-KOH pH 7.5, 140mM NaCl, 1mM EDTA, 10% glycerol, 0.5% NP-40, 0.25% Triton X-100, and 1mM DTT. The wash buffer contained 10mM Tris-

HCl pH 8.0, 200mM NaCl, 1mM EDTA, and 0.5mM EGTA. The nuclear lysis buffer consisted of 10mM Tris-HCl pH 8.0, 100mM NaCl, 1mM EDTA, 0.5mM EGTA, 0.1% sodium deoxycholate, 0.5% sodium lauroylsarcosinate, 1% Triton X-100, 0.1% SDS, and 1mM DTT. All buffers were supplemented with protease and phosphatase inhibitors.

To isolate cytoplasmic and nuclear fractions, the cell pellet was washed with cold PBS and centrifuged for 5min at 1,500rpm, for a total of 3 washes. Cells were then resuspended in 500µL cytoplasmic lysis buffer and incubated on ice for 5min. Lysates were centrifuged for 10min at 1,500rpm, and the supernatant (cytoplasmic fraction) was removed without dislodging the pellet (isolated nuclei). The pellet was washed once with 1 mL wash buffer, centrifuged for 5min at 1,500rpm, then resuspended in 200µL nuclear lysis buffer and incubated on ice for 15min. A final centrifugation of 10min at 13,000rpm, yielded the supernatant as the nuclear fraction lysate.

2.3. Molecular biology

2.3.1. Plasmids

Plasmids encoding single guide RNAs (sgRNAs) targeting various DUBs were kindly donated from Vincenzo D'Angiolella's Group (Institute of Genetics and Cancer, University of Edinburgh). Glycerol stocks of Stbl3™ cells containing sgRNAs that target DUBs were selected from a pre-defined Invitrogen™ LentiArray™ Human CRISPR Library (A32167, A32168, A32169, A32170, A32171, A32172, A32173, A32174, A32175, A32176, A32177, A32178, A32179, A32180, A32181, A32182, A32183, A32184, and A32185, Invitrogen). Plasmids were verified by DNA sequencing.

Plasmids used for lentivirus packaging were provided by Dr. Ilknur Sur Erdem (The Chinese Academy of Medical Sciences Oxford Institute, University of Oxford). A Cas9 overexpression plasmid was donated by Wanlin He from Sir Andrew McMichaels's group at the Centre for Immuno-Oncology, University of Oxford.

For plasmid isolation, Stbl3™ cells containing plasmids were grown in autoclaved LB broth (L3022, Sigma-Aldrich) supplemented with 100µg/mL ampicillin (A5354, Sigma-Aldrich) overnight in a shaking incubator at 37°C. Overnight cultures were then processed using QIAGEN Plasmid Plus Midi kits (12943, QIAGEN) or QIAGEN Plasmid Plus 96 kits (16181, QIAGEN), following manufacturers protocols. Plasmid concentrations were determined by NanoDrop.

2.3.2. Generation of HL60-Cas9 and HL60-Cas9-KO cell lines

HEK293T cells (CRL-1573, ATCC) were cultured in high glucose DMEM (D5671, Sigma-Aldrich), supplemented with 10% FBS, and 2% L-glutamine (25030081, Gibco) at 37°C in a humidified, 5% CO_2 incubator.

Lentiviral particles were generated by transfection of HEK293T cells (plated on 6-well plates) with the Cas9 overexpression plasmid or DUB-targeting sgRNA plasmids, and lentivirus packaging plasmids (psPAX2, 12260; pMD2.G, 12259; addgene), using FuGENE® HD Transfection Reagent (HD-1000, FuGENE®), in ratio amounts of 9:7:5, pX : psPAX2 : pMD2.G, where pX is the plasmid of interest, for a total of 2.1μg of plasmid DNA per well of a 6-well plate. For sgRNA plasmids, at least 3 unique plasmids encoding sgRNAs that target the same DUB were combined in equal proportions. At 2-day intervals for 4 days, culture medium was collected (with replacement on day 2). Cellular debris was removed by centrifugation for 5min at 500rpm, 4°C, from collected media, and supernatants were tested using Lenti-X™ GoStix™ Plus (631280, Takara Bio). Lentivirus-containing supernatant was concentrated using Lenti-X™ Concentrator (631232, Takara Bio), following manufacturer's protocol. Lenti-X™ Concentrator was added to supernatants in volume ratios of 1:3 and incubated at 4°C overnight. Precipitated lentivirus particles were then collected by centrifugation for 45min at 1,500rpm, 4°C. Supernatant was removed and concentrated lentivirus pellets were resuspended in PBS. Lentivirus particles were either used immediately for transduction or stored in aliquots at -80°C.

The HL60-Cas9 cell line was generated by transduction, with polybrene (TR-1003, Sigma-Aldrich), of HL60 cells using lentiviral particles containing Cas9 plasmids. Cells and virus particles were centrifuged for 30 mins at 800rpm, 37°C, to improve contact between virus particles and HL60 cells. After 24h, 10μg/mL blasticidin (SBR00022, Sigma-Aldrich) was added to culture medium for selection and maintenance. Cells were immediately cultured in 'recovery' medium – RPMI-1640 supplemented with 50% filter-sterilised medium from WT HL60 cells, an additional 10% FBS, and 50μM β-mercaptoethanol – for 2 days, to help reduce oxidative stress. Then, regular culture medium was used once cells recovered from antibiotic selection.

To generate KO cell lines, HL60-Cas9 cells were transduced (using polybrene) with lentivirus containing DUB sgRNA plasmids and selected with 10μg/ml puromycin. Again, similar transduction procedure was employed and cells were then cultured in 'recovery' medium (see above) to minimise oxidative stress. Transduced HL60-Cas9-DUB KO cells were maintained with 10μg/mL blasticidin and 10μg/ml puromycin supplemented in culture medium.

2.3.3. SDS-PAGE and Western blot

Samples were resuspended in RIPA Lysis and Extraction Buffer (89900, ThermoFisher) supplemented with cOmplete™, EDTA-free Protease Inhibitor Cocktail (COEDTAF-RO, Roche) and PhosSTOP™ (PHOSS-RO, Roche), and 0.1μL of Benzonase® Nuclease (Sigma-Aldrich, E1014). If lysates were viscous, they were sonicated in a water bath for 5min to sheer DNA. Protein quantification was performed using the Pierce™ BCA Protein Assay Kit (23225, 23227 and A65453, ThermoFisher). Lysates were mixed with

NuPAGE™ LDS Sample Buffer (4X) (NP0008, Invitrogen) to a final concentration of 1X Sample Buffer and 50 mM DTT, and boiled for 10min at 97°C. 10-20µg of protein from samples were loaded onto a NuPAGE™ 4 to 12%, Bis-Tris, 1.0mm, Midi Protein Gel (WG1402BOX, Invitrogen) and separated at a constant voltage of 130V for approximately 1-1.5h at room temperature. Separated proteins were transferred to nitrocellulose membranes (88018, Life Technologies Ltd) at a constant current of 390mA for 1h at 4°C. Membranes were then blocked with 5% milk for 1h at room temperature, washed 3 times in TBST, and incubated with primary antibody of interest overnight at 4°C with agitation. Primary antibody was then removed and the membrane was washed 3 times in TBST. The membrane was then incubated with an appropriate secondary antibody, for 1h at room temperature with agitation, then washed 3 times. The Li-Cor detection system was used for imaging blotted membranes.

2.4. Mass spectrometry sample preparation

Buffers for resuspending peptides were prepared using HPLC-grade reagents. Input samples for S-Trap digestion include cell pellets and lysates from nucleus-enriched fractions.

2.4.1. S-Trap protein digestion

Samples were resuspended in RIPA lysis and extraction buffer supplemented with 0.1µL of Benzonase® Nuclease. Lysates were sonicated in a water bath for 5min to sheer DNA. SDS was added to lysates to a final concentration of 5%. Samples were then processed using the S-Trap™ micro/midi spin column digestion protocol (C02-micro/C02-midi, Protifi LLC), depending on protein amounts. Manufacturer instructions were followed, using 5mM tris-2-carboxyethyl-phosphine (TCEP) and 20mM iodoacetamide (IAA) as the reducing and alkylating agents, respectively. Proteins were digested overnight using TPCK-treated trypsin (20233, ThermoFisher), using quantities recommended by S-Trap instructions (1:25, protein : trypsin). Resultant peptides were eluted and dried using a SpeedVac concentrator centrifuge. Dried peptides were resuspended in an appropriate buffer for mass spectrometry (MS) analysis or peptide-level enrichment.

2.4.2. Phosphopeptide enrichment

For phosphopeptide enrichment, S-Trap protein digestion used citric acid instead of phosphoric acid to acidify proteins (pH ≤ 1) before loading onto columns. Following S-Trap protein digestion, peptides derived from ~300µg of protein were aliquoted, dried, and resuspended in a buffer containing 50% ACN, 2% trifluoracetic acid (TFA), 1M glycolic acid in H_2O.

Phosphopeptide enrichment was performed using the pre-defined Agilent Bravo AssayMap liquid handler workflow for TiO_2 cartridges (G5496-60016, Agilent). The prime and syringe wash solution consisted of 50% ACN, 5% NH_3 in H_2O and the equilibration and cartridge wash solution was 50% ACN, 2% TFA, 1M glycolic acid in H_2O. Phosphopeptides were eluted in 15% ACN, 5% NH_3. Samples were

loaded onto the cartridges at 2.5µL/min. Eluted phosho-enriched peptides were dried down and resuspended in 5% DMSO and 5% formic acid (FA).

2.4.3. Ubiquitinated peptide enrichment

For ubiquitin-modified peptide enrichment, S-Trap protein digestion used citric acid instead of phosphoric acid to acidify proteins (pH ≤ 1) before loading onto columns. Following S-Trap protein digestion, peptides derived from ~1mg of protein were aliquoted and dried for downstream enrichment. Ubiquitinated peptides contain a K-GG (diGly) remnant motif upon trypsin digestion. These peptides were enriched using the PTMScan® Ubiquitin Remnant Motif (K-ε-GG) Kit (5562, Cell Signalling Technologies). Manufacturers protocol was followed, using 10µL of beads instead of 20µL.

2.5. Mass spectrometry analysis

2.5.1. Q-Exactive

Peptides were resuspended in 5% formic acid and 5% DMSO and then trapped on an Acclaim™ PepMap™ 100 C18 HPLC Columns (300µm x 5mm, 5µm particle size, 160454, ThermoFisher) using solvent A (0.1% FA in H_2O) at a pressure of 60 bar and separated on an Ultimate 3000 UHPLC system (ThermoFisher) coupled to a Q-Exactive (QE) mass spectrometer (ThermoFisher). The peptides were separated on an Easy Spray PepMap RSLC column (75µm i.d. x 2µm x 50mm, 100 Å, ThermoFisher) and then electro-sprayed directly into the QE mass spectrometer. An EASY-Spray nano-electrospray ion source (ThermoFisher) using a linear gradient of 5% to 35% solvent B (0.1% FA in ACN and 5% DMSO), with length of 120min/60min, depending on sample complexity, and flow rate of 250nL/min. The raw data was acquired in the mass spectrometer in data-independent acquisition (DIA) mode. Full scan MS spectra were acquired in the Orbitrap (inclusion list with scan range 495-995 m/z, 20 m/z increments, with an overlap of +/- 2 Daltons, resolution 35000, AGC target 3e6, maximum injection time 55ms). After the MS scans peaks were selected for higher-energy collisional dissociation (HCD) fragmentation at 28% of normalised collision energy (NCE) / stepped NCE. HCD spectra were also acquired in the Orbitrap (resolution 17500, AGC target 1e6, isolation window 20 m/z).

2.5.2. TimsTOF

Peptides were chromatographically separated on a 15cm x 150µm x 1.5µm analytical column (EV1137, Evosep) using the 30 samples per day (SPD) gradient method of an Evosep One LC system. Eluted peptides were directed for MS analysis on a TimsTOF Pro mass spectrometer (Bruker). Phosphopeptide MS data were acquired in ddaPASEF mode. The ion mobility window was 0.60-1.60Vs/cm2, with an accumulation and ramp time of 100ms. The mass range of MS and MS/MS scans was m/z 150–1,700. MS/MS spectra were acquired in 10 PASEF ramps with a 4-frame overlap, giving a cycle time of 1.17s. Ions were selected for PASEF MS/MS if they met an intensity threshold of 2,500 and were sampled

multiple times until a target intensity of 20,000 was reached. A polygon filter was used to exclude singly charged ions from MS/MS selection. Ions selected for fragmentation were isolated by the quadrupole and fragmented using an ion mobility-dependent collision energy that increased nonlinearly over the ion mobility range as follows: a collision energy (eV) of 20 at a 1/K0 of 0.60Vs/cm2, 22 at 0.70, 25 at 0.75, 30 at 0.80, 35 at 0.85, 40 at 0.90, 45 at 0.95, 50 at 1.00, 55 at 1.10, 60 at 1.20, 65 at 1.30, 70 at 1.40, 75 at 1.50, and 80 at 1.60. A dynamic exclusion time of 24 sec was used. Total proteome and ubiquitome MS data were acquired in diaPASEF mode using 16 diaPASEF scans per TIMS-MS scan with an accumulation and ramp time of 100 msec, for a total cycle time of 1.80s. The ion mobility range was set to 0.6-1.6Vs/cm^2. Each mass window isolated was m/z 26 wide, ranging from m/z 400-1200 with an ion mobility-dependent collision energy that increased linearly from 20eV to 59eV between 0.6-1.6Vs/cm^2.

2.6. Mass spectrometry data analysis

2.6.1. Data processing

Raw DIA MS data from proteomic analysis was analysed using DIA-NN v1.9.2 or v2.1[111] with library-free search and default settings. Ubiquitomic analysis of raw MS data included 'K-GG' as a variable modification, with 2 variable modifications per peptide and 2 missed cleavages. Database search of raw phosphoproteomic DIA MS data was performed using either DIA-NN, with 3 variable 'phospho' modifications and 2 missed cleavages, or FragPipe v23.1[112-118], with the default 'DIA-phospho' workflow using phosphorylation (STY) as a variable modification and diaTracer spectral deconvolution[119]. Raw DDA MS phosphoproteome data was analysed using FragPipe with phosphorylation (STY) as a variable modification, using the default settings in the 'LFQ-phospho' workflow. 3 variable modifications per peptide were enabled for phosphoproteomic analysis. Methionine oxidation and cysteine carbamidomethylation were fixed modifications for all analyses.

2.6.2. Bioinformatics

Bioinformatic analysis and visualisation of data was primarily performed in RStudio[120] 2023.03.0+386 (R, v4.3.2) using ggplot2[121]. Session information and packages can be found in Appendices Section 1 (Appendices 'Chapter 2').

Data quality control (QC) involved principal component analysis (PCA), Pearson correlation coefficient (PCC) calculation, and inspection of the distribution of intensity values. Data was log2-transformed and normalised by median subtraction within each replicate. Filtering, imputation, and coefficient of variation (CV) adjustment were performed depending on the dataset (Table 2.7).

2.6.2.1. Differential abundance analysis

For differential abundance analysis, the moderated t-test was used to shrink sample variances and increase statistical power by implementing an empirical Bayes method. Benjamini-Hochberg procedure (BH) for false discovery rate (FDR) control was applied to moderated t-test p-values. Time points were compared to day 0 and results were visualised using volcano plots.

2.6.2.2. Clustering

Clustering approaches included hierarchical clustering and weighted correlation network analysis (WCNA), the latter using the corresponding R package[122]. For hierarchical clustering, median values for each protein or peptide, per time point, were Z-score normalised across all time points to determine an abundance profile. Clustered data were then interrogated for functional enrichment and enzyme-substrate interactions (ESI), the latter of which focused on E3 ligases, DUBs, kinases, and phosphatases. ORA used proteins identified in the experiment as the background. Additionally, associations between TFs and TF target genes were assessed. Relevant databases and methods are shown in Table 2.8.

Table 2.7 **Details of data processing steps**.

Dataset	Data acquisition type	Filtering	Imputation
Initial proteomic analysis of HoxB8 neutrophil differentiation			
Proteome	DIA	Protein group quantified in 3/3 biological replicates for at least 1 time point	Random sampling from downshifted-median normal distribution
Proteomics, ubiquitomics, and phosphoproteomics of HoxB8 neutrophil differentiation			
Proteome	DIA	Protein group quantified in at least 3/4 biological replicates for at least 1 time point	Random sampling from downshifted-median normal distribution
Ubiquitome		Unique modified peptide quantified in at least 3/4 biological replicates for at least 1 time point	Absent IDs: Random sampling from downshifted-median normal distribution. Missing IDs: K-nearest neighbours (KNN) imputation
Phosphoproteome	DDA/DIA		

Table 2.8 **Details of databases and methods used in downstream analyses.**

Analysis		Database/Implementation	Ref
Functional enrichment (over representation)		clusterProfiler https://bioconductor.org/packages/devel/bioc/html/clusterProfiler.htmlt	123
Gene set enrichment		clusterProfiler https://bioconductor.org/packages/devel/bioc/html/clusterProfiler.html	123
ESI	Kinases	PhosphoSitePlus https://www.phosphosite.org/	124
	Phosphatases	DEPOD https://depod.bioss.uni-freiburg.de/index.php	125
	E3 ligases	UbiBrowser 2.0 http://ubibrowser.bio-it.cn/ubibrowser_v3/	126
	Deubiquitinases		
Transcription factor-target gene		TFlink https://tflink.net/	127
Loss of function (LoF) intolerance		Genome Aggregation Database (gnomAD) v4.1.0 https://gnomad.broadinstitute.org/	128
Cell line gene dependency		DepMap https://depmap.org/portal	129
Cell line RNA expression		Human Protein Atlas https://www.proteinatlas.org/	130,131

2.6.2.3. Functional enrichment

Over representation analysis (ORA) used the enrichGO and enrichKEGG functions from the clusterProfiler package[123], and gene set enrichment analysis (GSEA) used the gseGO and gseKEGG functions. All functional enrichment analyses used genes corresponding to proteins identified in the datasets as background, as opposed to the whole genome. Ranking metrics used for GSEA are described in experimental chapters. Permutation-based FDR threshold for significance was 0.25.

2.6.2.4. Enzyme-substrate interaction enrichment

Using information regarding enzyme-substrate interactions from corresponding databases, clustered data were subjected to hypergeometric testing to assess statistical significance of identifying substrates of an enzyme within a cluster. BH FDR correction was applied to resulting p-values.

2.6.2.5. Linear regression

Linear regression analysis used the 'lm' function in the 'stats' R package to predict estimates for coefficients of each predictor variable. Unless otherwise stated, coefficients were filtered to maintain only those that showed significant FDR-adjusted p-values (FDR < 0.05) for all predictor variables.

2.7. Flow cytometry

The Attune™ NxT Acoustic Focusing Cytometer (ThermoFisher) and CytKick Max autosampler (ThermoFisher) were used to count and analyse (fixed) cells. Instrument settings were maintained for all experiments after initial calibration using HL60 WT cells. Forward scatter (FSC) versus side scatter (SSC) dot plots with gating were used to count HL60 and dHL60 cells. Cells were labelled with APC-conjugated antibodies that bind CXCR2, CXCR4, CD11b, and IgG (isotype control), following the manufacturers protocol (130-110-612, 130-120-133, 130-120-778, 130-113-434, Miltenyi Biotec). FSC-Height (FSC-H) versus FSC-Area (FSC-A) plots were used to gate single cells, and labelling of antibody was measured using the default APC-detection option (638 nm red laser, RL1 far red channel, 670/14 filter).

2.8. Microscopy

2.8.1. Image acquisition

Fixed cells were stained with Hoechst 33342 (nuclear stain) and/or wheat germ agglutinin (WGA)-AlexaFluor647 (WGA-AF647, cell membrane stain).

Images were either acquired on a ZEISS Axio Observer Z1 microscope or a ZEISS Celldiscoverer 7 microscope for hight-throughput imaging.

Images acquired on ZEISS Axio Observer Z1 used a ZEISS EC Plan-Neofluar 40X, 0.75 NA M27 objective, 1.6x tube lens, and an Axiocam 506 mono camera with a 0.071 µm pixel size. Acquisition details are listed in Table 2.9.

Images acquired using a ZEISS Celldiscoverer 7 widefield fluorescence microscope used a ZEISS Plan-Apochromat 20X, 0.95 NA objective and an Axiocam 807 mono CMOS camera with a 4.5 µm pixel size. 31 Z stacks with 0.5 µm spacing were captured for at least six randomly selected positions per well. Specific exposure times and individual channel and filter details are listed in Table 2.9. Acquisition and deconvolution were performed in ZEN Blue v3.12. Fast iterative deconvolution was performed with default settings in ZEN.

Table 2.9 ***Image acquisition details for microscopy.***

Microscope	ZEISS Axio Observer Z1	ZEISS Celldiscoverer 7 widefield fluorescence microscope	
Target/ Fluorophore	Hoechst 33342 (nucleus)	Hoechst 33342 (nucleus)	WGA-AF647 (cell membrane)
Excitation Laser	357-393 nm	385 nm	630 nm
Emission filters	QBP 432 + 519 + 590 + 680 (HE)	QBP 425/30 + 514/30 + 592/25 + 709/100	QBP 425/30 + 514/30 + 592/25 + 709/100
Exposure time (ms)	80	10	20
LED power	50%	10%	50%

2.8.2. Image analysis

Z-stacks were merged by max intensity projection for each scene and channel (DNA and membrane stains). Images were then analysed using CellProfiler v4.2.8[132]. A gaussian filter was applied to the membrane channel image, and objects (cells) were identified with diameters of 50-120 pixels, using the global 'Otsu' thresholding method and 'declumping' based on shape. Identified objects were filtered based on form factor values between 0.8 and 1.0 and compactness between 1 and 1.5, to identify round cells. Filtered objects were then used to mask the DNA channel image. Nuclei were identified in a similar fashion to cells, using object identification with diameters of 10-80 pixels, the global 'Otsu' thresholding method, and declumping based on shape. Due to the lobed nature of certain nuclei, occasionally multiple objects were identified in a single cell. Therefore, nearby objects within 15 pixels of each other were merged into a single object to represent each nucleus. The shape, size, and intensity distributions of each nuclei object were then measured.

Nuclei measurements were then analysed and processed using R and RStudio and visualised using ggplot2. Distributions of form factor values were compared to HL60-WT cells for each time point using the Kolmogorov–Smirnov test with BH FDR correction. Proportion of dHL60 cells were calculated using a form factor threshold value, as described in Chapter 6.

Chapter 3: Proteomic changes during HoxB8 mouse neutrophil differentiation

3.1. Introduction

In this chapter, the HoxB8 mouse neutrophil model was analysed at different stages of differentiation using MS-based proteomic approaches.

Regulatory mechanisms that govern neutrophil differentiation on the protein level are unclear. Working with primary neutrophils poses different challenges due to their short lifespan and tendency to become activated in response to cellular debris and artificial extraction. The differentiation niche of neutrophils is within the bone marrow, and so isolating sufficient material and recapitulation of the appropriate environment presents several limitations. As such, *in vitro* and *ex vivo* cellular models are required.

The Udalova group at the Kennedy Institute of Rheumatology, University of Oxford, have established the HoxB8 neutrophil model in the lab. This system takes mouse myeloid progenitor cells and immortalises them through genetic manipulation of the HoxB8 TF[32]. HoxB8 progenitors can then be maintained in a proliferative state in the presence of estradiol, and upon estradiol removal and G-CSF induction, *ex vivo* differentiation of HoxB8 progenitors occurs over several days, resulting in non-proliferating neutrophils. This provides plenty of biological material through expansion, facilitating characterisation of the neutrophil differentiation process.

The work in this chapter aims to characterise the proteomic changes that occur during HoxB8 neutrophil differentiation to assess the quality of the data, the reproducibility of the model, and ability to enrich for nuclear proteins to specifically identify and quantify TFs during HoxB8 differentiation. Additionally, different bioinformatic approaches were investigated to process the proteome data, glean biological insights, and make predictions about potential regulatory mechanisms.

3.2. Results

HoxB8 progenitor cells were cultured *ex vivo* and induced to differentiate into neutrophils. Whole cell lysates (total proteome) and nuclear-enriched fractions of the HoxB8 neutrophils were collected at daily intervals post-induction for 5 days, with 3 biological replicates, and subsequently subjected to MS-based proteomic analysis to identify protein changes that occur during the differentiation process.

Raw MS data were processed using DIA-NN v1.9.2 for identification and label-free quantitation of peptides, and corresponding protein inference, for each time point. Analysis focused on protein group-level data. The DIA-NN output was filtered to remove protein IDs that did not appear in all replicates

(3/3), in at least one time point (sample). Filtered data were log_2-transformed and median normalised, and missing values were imputed from a downshifted normal distribution to emulate low-abundant proteins (henceforth referred to as 'processed data').

An overview of the experimental workflow is depicted in Figure 3.1.

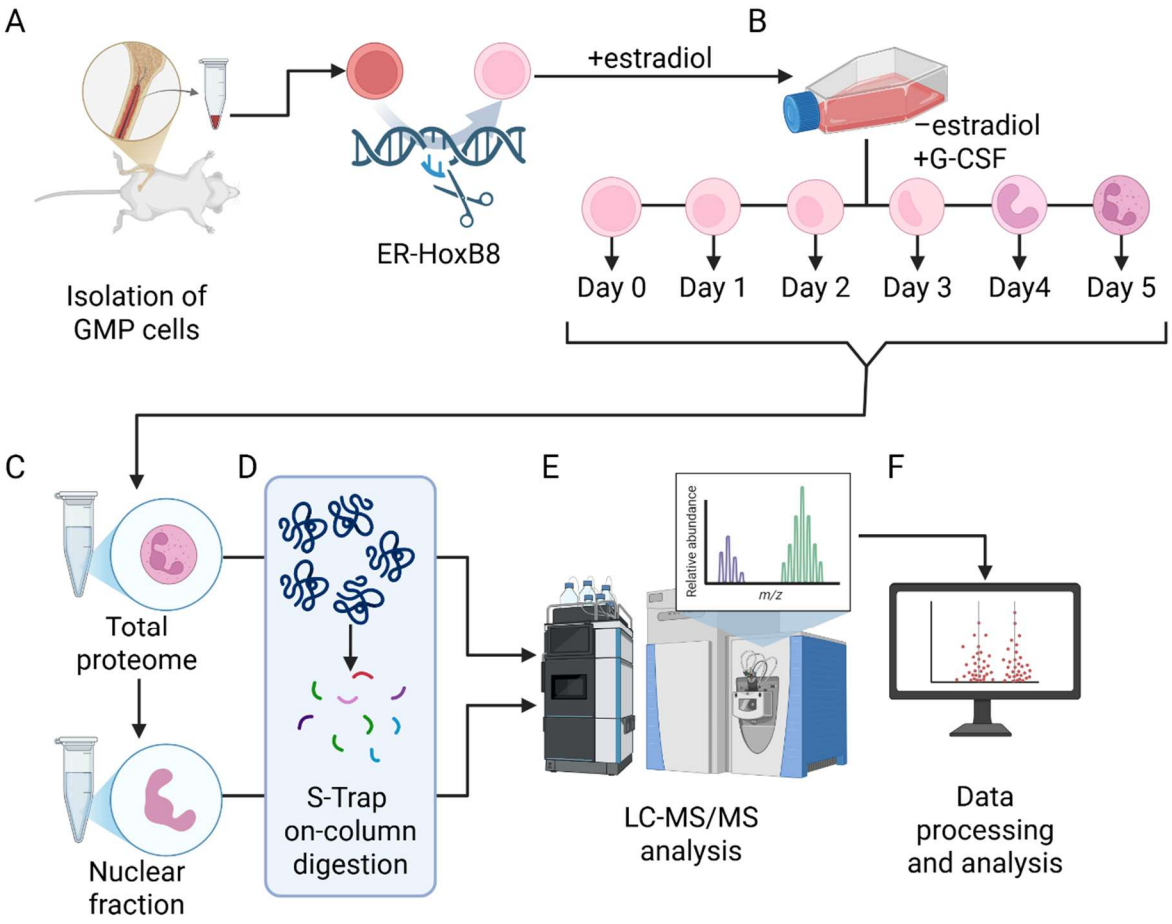


Figure 3.1 ***Experimental workflow for proteomic analysis of HoxB8 neutrophil differentiation.*** (**A**) Granulocyte-monocyte progenitor (GMP) cells were isolated from murine bone marrow and genetically modified with estrogen receptor (ER)-HoxB8 fusion protein. In the presence of estradiol, HoxB8 GMPs can be cultured *ex vivo*. (**B**) Upon estradiol removal and granulocyte colony stimulating factor (G-CSF) induction, HoxB8 GMPs can be differentiated into HoxB8 neutrophils. (**C**) Cell pellets and nuclear-enriched fractions were collected at daily intervals for 5 days. (**D**) All samples were processed using S-Trap on-column digestion. (**E**) Resultant peptides were analysed using liquid chromatography tandem mass spectrometry (LC-MS/MS). (**F**) Raw mass spectrometry data were processed and analysed using bioinformatic approaches.

A total of 7235 protein groups were identified in the total proteome, and 5497 in the nuclear fraction, across all time points. After filtering for valid values in all replicates (3/3) for at least one time point, the number of protein group IDs were 7082 and 5286 for the total proteome and nuclear fraction, respectively. Initial QC of the datasets involved calculating coefficient of variation (CV) values for each protein per time point; inspecting the distributions of log2(protein intensities) and processed data; calculating PCCs between biological replicates for each time point using the dataset with imputed values; and performing PCA on the data with imputed values to assess the impact of normalisation and imputation.

The moderated t-test with FDR multiple correction testing was used to identify differentially abundant proteins (DAPs) at each time point, compared to day 0. Downstream analyses included GSEA using a list of proteins ranked based on log_2(fold change) (log_2FC); identifying DAPs, generating abundance profiles, and clustering profiles; and performing over representation analysis (ORA) and enzyme-substrate interaction (ESI) analyses on clustered data.

3.2.1. Nuclear fraction enrichment provided limited information compared to total proteome data

Proteomic analysis of nuclear fractions enriched from HoxB8 neutrophils during differentiation was intended to improve the identification and quantitation of proteins that localise to the nucleus. It was hypothesised that decreasing the complexity of the protein-derived peptides that are subjected to MS analysis would potentially improve the detection of low-abundant, nuclear-localised proteins, such as TFs.

Initial QC of the data showed generally consistent protein IDs across all replicates with low coefficient of variation for each protein across all time points, suggesting that the biological replicates are consistent and protein abundances are changing rather than protein presence/absence (Figure 3.2A, B). Normal distributions of log_2(protein intensity) values per replicate were observed, showing consistency between replicates, and normalised data with imputed values introduced a ‘shoulder’ to the normal distributions, which was expected due to the imputation method (Figure 3.2C, D). High correlation was observed between replicates, and lower correlations between time points, with the lowest correlation between consecutive time points occurring between days 2 and 3 (Figure 3.2E). This suggests a ‘switch point’ during the differentiation process where proteomic changes are more pronounced as cells progress towards a more specialised state. This is in concordance with the PCA (Figure 3.2F), which shows principal component 1 (Dim1) explaining 41% of the variance, indicating large separation between days 0-2 and days 3-5.

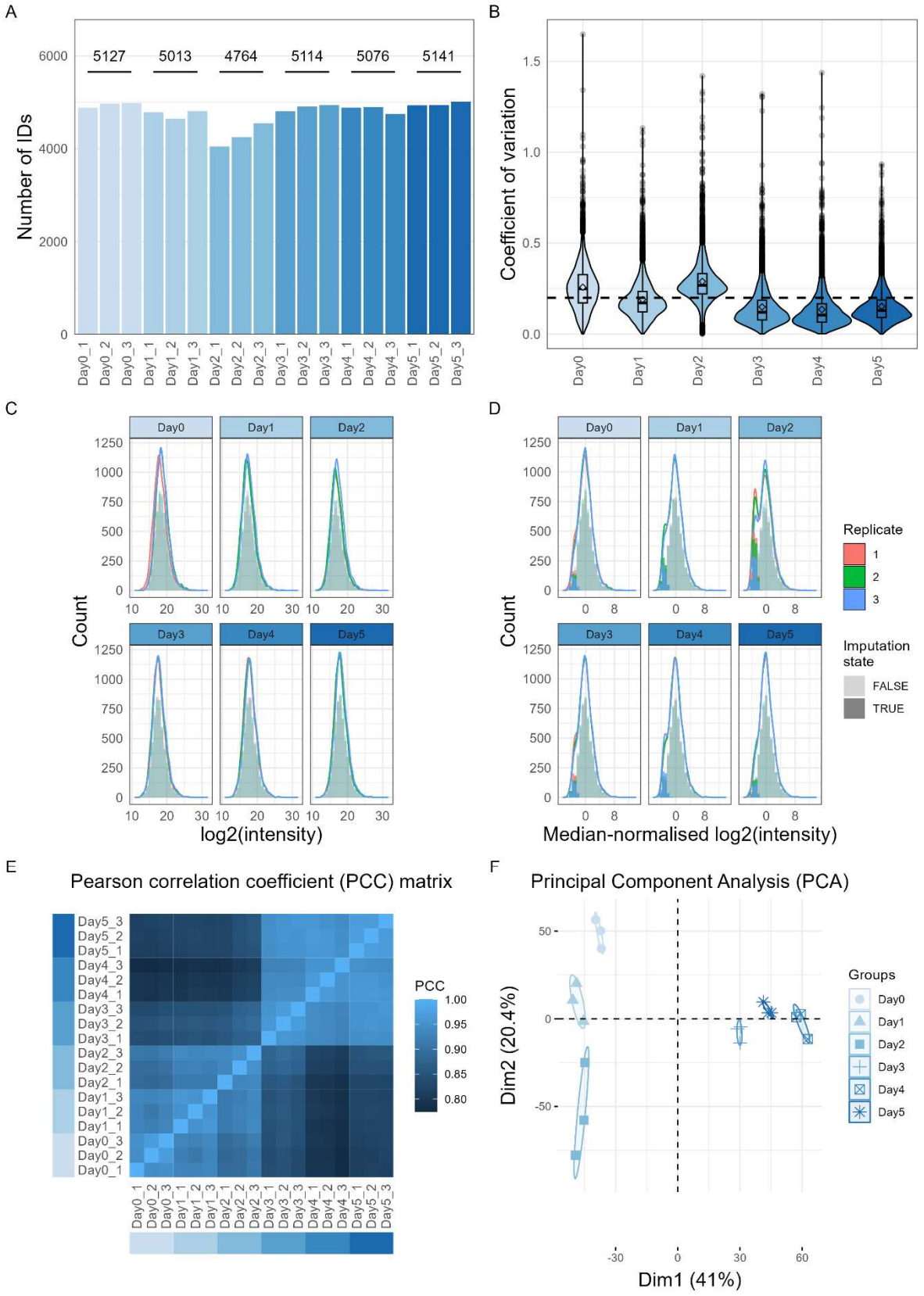


Figure 3.2 ***QC of proteomic data derived from nuclear fractions of HoxB8 neutrophils during differentiation.*** (A) Number of protein group IDs per biological replicate, with total IDs per time point shown above each group (n = 3 biological replicates). (B) Violin plots showing the distribution of coefficient of variation (CV) values per time point. (C) Distribution of log2(protein intensity) values for each time point. (D) Distribution of median-normalised, log2-transformed protein intensities with imputed missing values (processed data). Imputed values were sampled from a downshifted normal distribution to emulate low protein abundances (these values are shown as more opaque sections of their corresponding histogram) (E) Pearson correlation coefficient between biological replicates and time points of the processed data. (F) Principal component analysis of the processed data.

Moderated t-test of the data revealed that key TFs in neutrophil differentiation and effector functions were expressed in stage specific patterns[103]; notably, the C/EBP family exhibited differential abundance during early stages of HoxB8 neutrophil differentiation, consistent with literature pertaining to physiological neutrophils (Figure 3.3). GSEA of nuclear fraction data can be found in Supplementary Figures 1-8.

To assess the nuclear fraction enrichment efficiency, median protein intensity values from nuclear fractions and total proteomes, for each respective time point, were determined and their differences calculated. GSEA was performed on a pre-ranked list of these proteins – based on the difference between processed datasets – to interrogate the gene ontology (GO) cellular component (CC) gene sets and whether nuclear-related terms were enriched (Figure 3.4). Although terms such as nucleolus and nucleosome were enriched, terms associated with other cellular components and compartments, such as the mitochondria, cytoskeleton, and extracellular matrix, were also positively enriched. Additionally, more TFs were identified in the total proteome compared to the nuclear fraction. These observations suggest that the nuclear fraction enrichment did not improve the coverage of nuclear proteins, possibly due to the sample preparation, which may require further optimisation.

Nevertheless, the total proteome dataset boasted a coverage of 7082 protein groups (after data filtering), therefore further analysis focused primarily on this dataset.

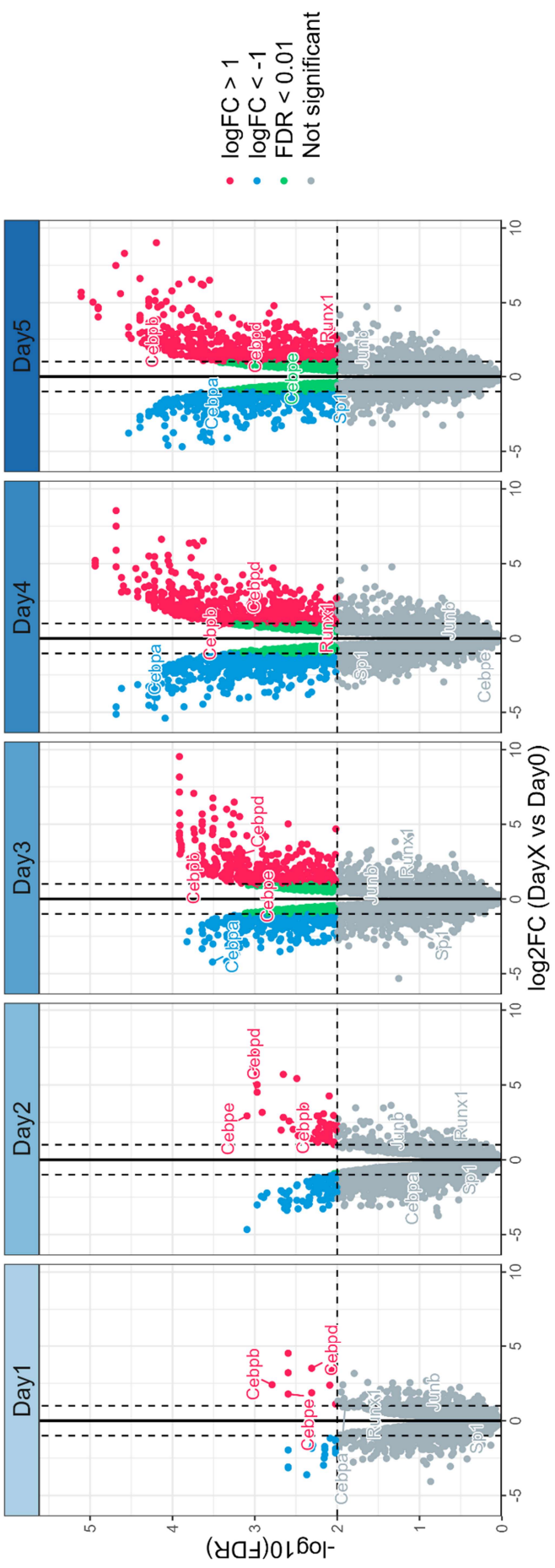


Figure 3.3 ***Volcano plots of differentially abundant proteins (DAPs) identified in the nuclear fractions of different stages of HoxB8 neutrophil differentiation and day 0.*** Moderated t-test was used to compare groups (n = 3 biological replicates), and the Benjamini-Hochberg procedure was used to control false discovery rate (FDR). Proteins with FDR < 0.01 and | log2(fold change)| > 1 were considered statistically significant and differentially abundant. Upregulated and downregulated proteins are shown in red and blue, respectively. Green represents proteins with FDR < 0.01 and |log2(fold change)| < 1.

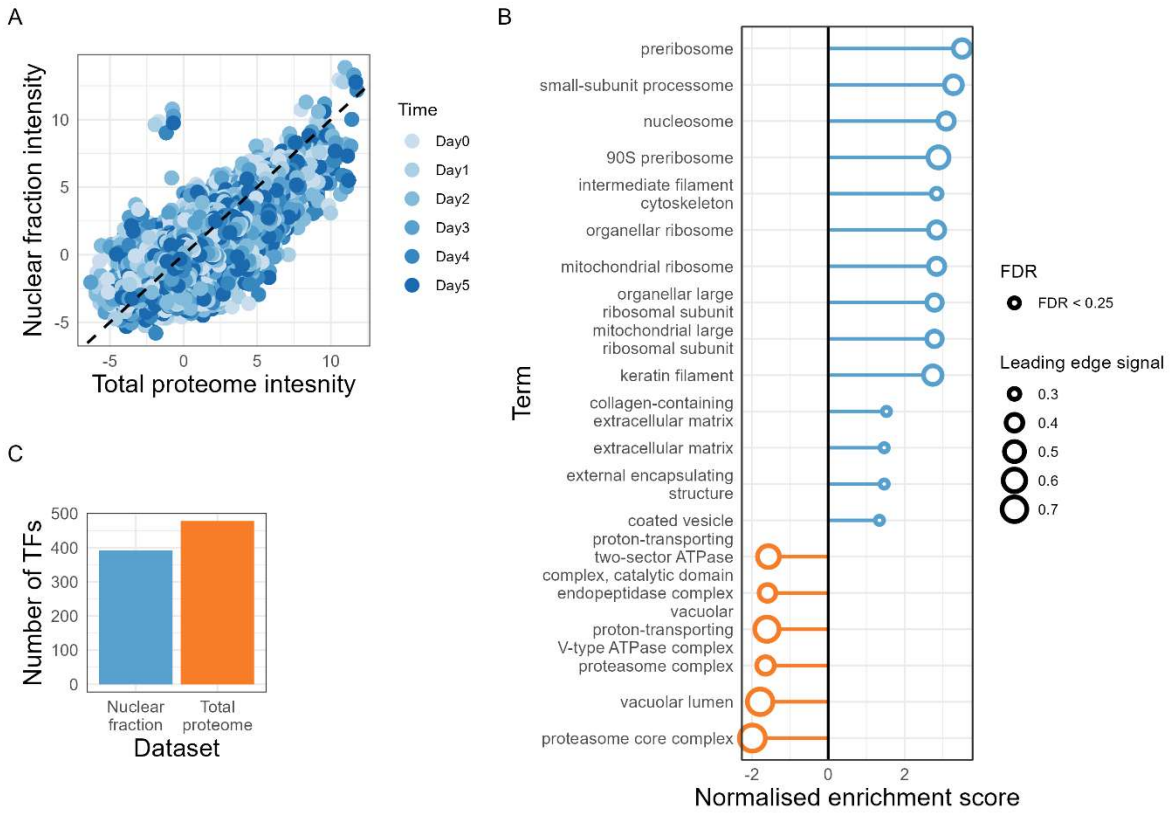


Figure 3.4 ***Comparison of proteomic analysis of nuclear fractions and total proteomes of HoxB8 neutrophils.*** (A) Scatter plot of nuclear fraction intensities against total proteome intensities. Each point corresponds to a protein, coloured based on differentiation stage (time point). Values correspond to the median protein abundance (n = 3 biological replicates). (B) Gene set enrichment analysis (GSEA) of proteins common to both data sets. Proteins were ranked based on the largest difference between median protein intensities for a given time point (nuclear fraction – total proteome). Stroke colour represents enrichment in the nuclear fraction (blue) or total proteome (orange), and size indicates leading edge signal. Significant terms are shown (FDR < 0.25). (C) Total number of transcription factor (TF) proteins identified in each dataset.

3.2.2. Total proteome analysis identifies altered protein abundances and biological processes during neutrophil differentiation

The total proteome dataset was processed identically to the nuclear fraction data. QC of the data (Figure 3.5) revealed higher CV values for proteins identified in day 5, possibly due to reduced number of protein IDs in one of the biological replicates (Figure 3.5A, B). Processed data generally showed normally-distributed data, with certain replicates showing more missing values that introduced a larger 'shoulder' to the distribution curve (Figure 3.5C,D). PCC calculations suggest the most extensive proteomic changes occurred between day 1 and 2 (Figure 3.5E). PCA and projection of data onto PC1 and PC2 were representative of the differentiation trajectory (Figure 3.5F).

Proteins which have been previously reported to be crucial for neutrophil maturation and effector functions showed clear upregulation during the differentiation process (Figure 3.6). Chemokine (C-X-C motif) receptor 2 (CXCR2) is a marker of mature neutrophils and essential for neutrophil physiology[133]. Lactoferrin (LTF) and lipocalin-2 (LCN2) exhibit antibacterial activities[134,135]. Matrix metalloproteinase-8 (MMP8) degrades the extracellular matrix, which is vital for neutrophil migration and infiltration[136]. Lastly, cytochrome b(558) subunit beta (CYBB) contributes to ROS generation[137].

GSEA was performed using a pre-ranked list of the genes based on the $\log_2$FC of the corresponding proteins for each time point compared to day 0. The GO biological process (BP) terms showed increasing normalised enrichment scores (NES) for neutrophil-associated genes during HoxB8 differentiation (Figure 3.7). For example, granulocyte/myeloid leukocyte/neutrophil migration/chemotaxis proteins are all significantly and positively enriched, particularly during day 5 of differentiation. Additionally, Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways associated with phagocytosis (phagosome, endocytosis, lysosome) and cytoskeletal remodelling (regulation of actin cytoskeleton, adherens junction, motor proteins) are positively enriched. Conversely, processes associated with DNA, RNA, and ribosomes are significantly negatively enriched, suggesting these pathways are downregulated during the differentiation process. Moreover, cell cycle proteins and proteins in p53 signalling pathways are downregulated, indicative of cell cycle modulation. These findings suggest that the proteomic landscape within neutrophil cells undergoes dynamic changes to increase synthesis and/or stabilisation of proteins relevant to neutrophil effector functions and divest resources from general cellular processes. This would be consistent with the idea that neutrophils are typically frontline, shorter-lived effector cells of the innate immune system[138].

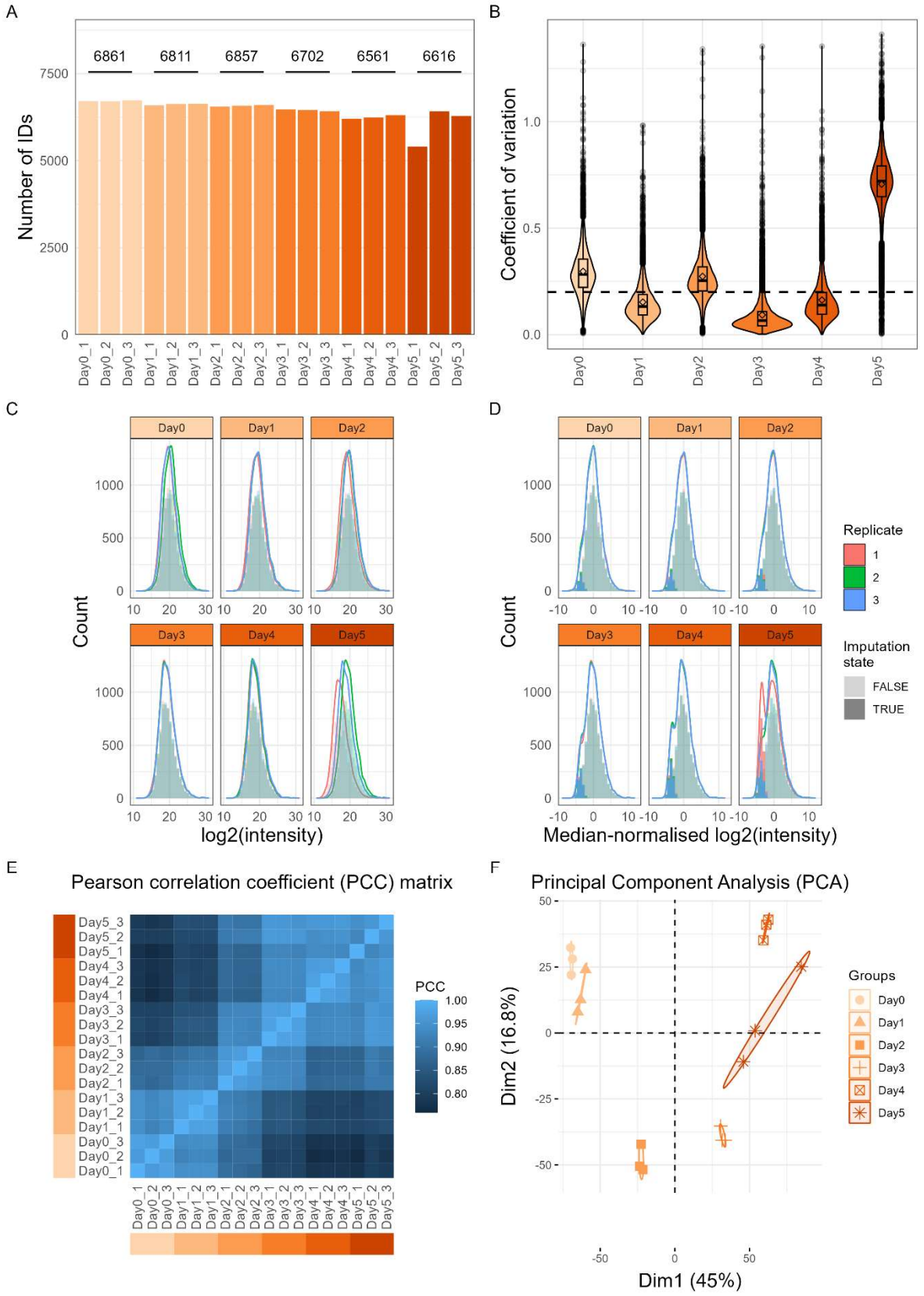


Figure 3.5 ***QC of proteomic data derived from total proteomes (whole cell lysates) of HoxB8 neutrophils during differentiation.*** (A) Number of protein group IDs per biological replicate, with total IDs per time point shown above each group (n = 3 biological replicates). (B) Violin plots showing the distribution of coefficient of variation (CV) values per time point. (C) Distribution of log2(protein intensity) values for each time point. (D) Distribution of median-normalised, log2-transformed protein intensities with imputed missing values (processed data). Imputed values were sampled from a downshifted normal distribution to emulate low protein abundances (these values are shown as more opaque sections of their corresponding histogram) (E) Pearson correlation coefficient between biological replicates and time points of the processed data. (F) Principal component analysis of the processed data.

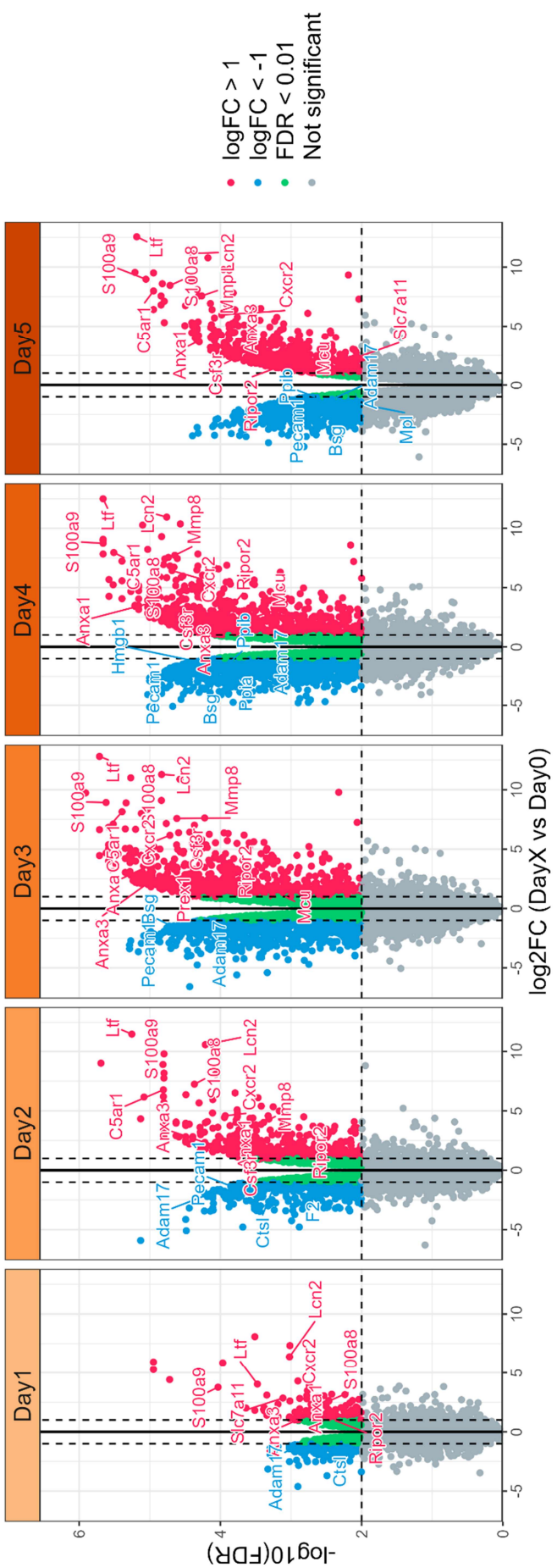


Figure 3.6 ***Volcano plots of differentially abundant proteins (DAPs) identified in the total proteomes of different stages of HoxB8 neutrophil differentiation and day 0.*** Moderated t-test was used to compare groups (n = 3 biological replicates), and the Benjamini-Hochberg procedure was used to control false discovery rate (FDR). Proteins with FDR < 0.01 and |log2(fold change)| > 1 were considered statistically significant and differentially abundant. Upregulated and downregulated proteins are shown in red and blue, respectively. Green represents proteins with FDR < 0.01 and |log2(fold change)| < 1.

Figure 3.7 ***Gene set enrichment analysis (GSEA) of total proteome data from HoxB8 neutrophil differentiation (GOBP).*** Genes were pre-ranked based on log2(fold change) of protein abundance compared to day 0. Stroke colour represents the time point, size indicates leading edge signal, and permutation-based FDR is indicated by fill colour. Gene ontology (GO) biological process (BP) terms are shown.

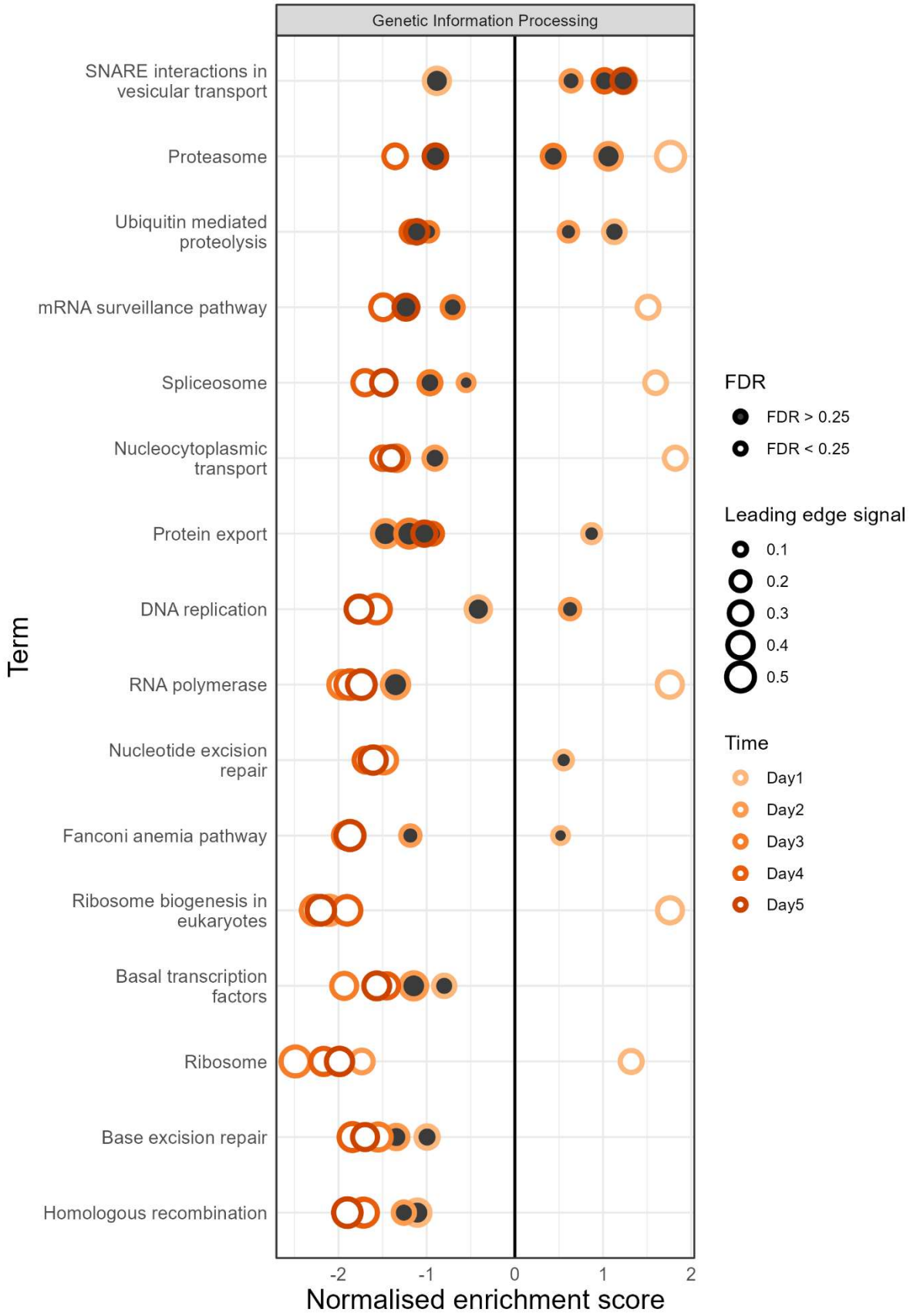


Figure 3.8 ***Gene set enrichment analysis (GSEA) of total proteome data from HoxB8 neutrophil differentiation (KEGG, Genetic Information Processing).*** Genes were pre-ranked based on log2(fold change) of protein abundance compared to day 0. Stroke colour represents the time point, size indicates leading edge signal, and permutation-based FDR is indicated by fill colour. The KEGG pathway subcategory 'Genetic Information Processing' terms are shown.

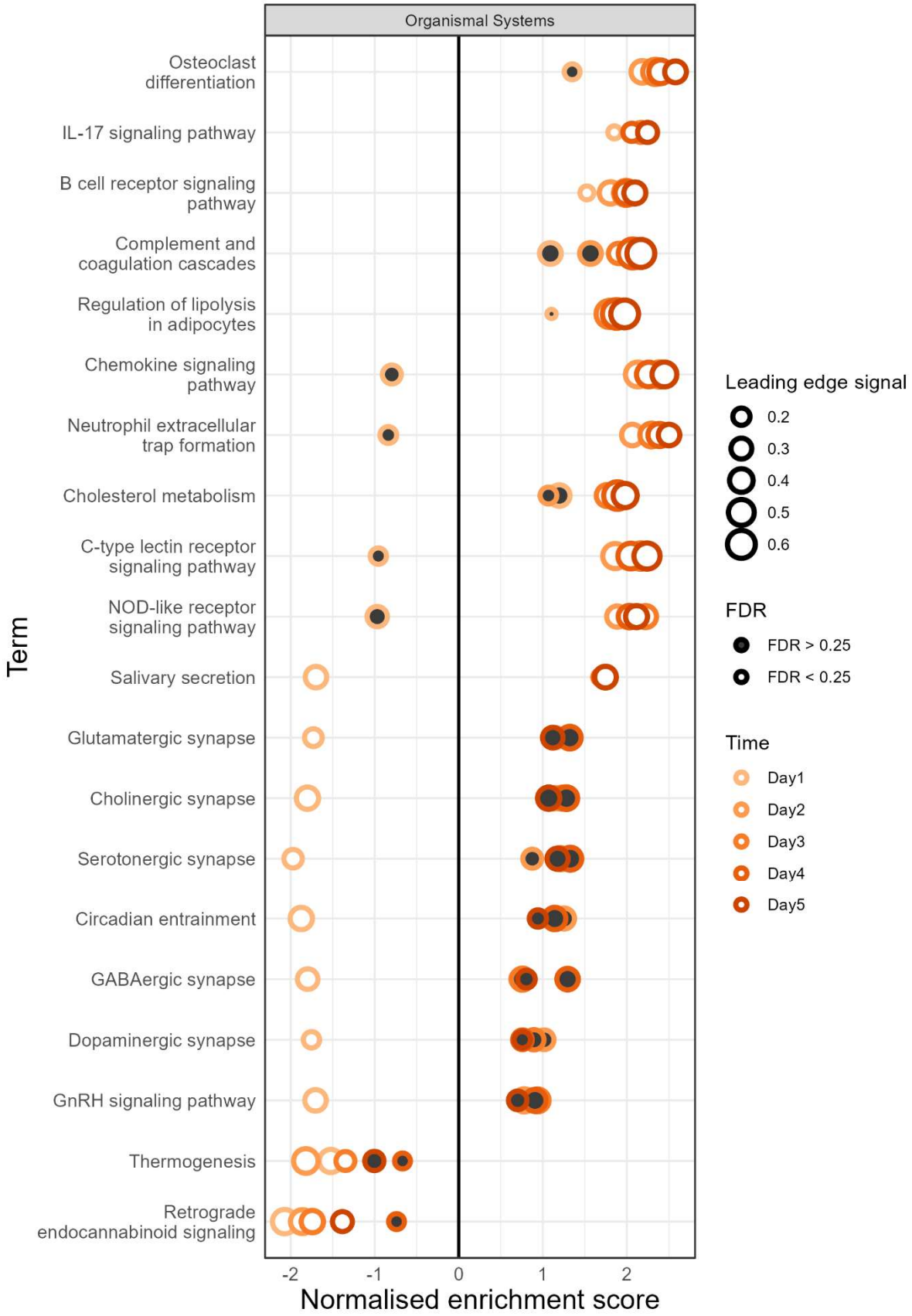


Figure 3.9 ***Gene set enrichment analysis (GSEA) of total proteome data from HoxB8 neutrophil differentiation (KEGG, Organismal Systems).*** Genes were pre-ranked based on log2(fold change) of protein abundance compared to day 0. Stroke colour represents the time point, size indicates leading edge signal, and permutation-based FDR is indicated by fill colour. The KEGG pathway subcategory 'Organismal Systems' terms are shown.

Further inspection of the KEGG subcategory ‘Genetic Information Processing’ highlighted changes in NES for terms associated with the proteasome and ubiquitin-mediated proteolysis (Figure 3.8). Interestingly, early stages (until day 2/3) of differentiation show upregulation of proteins from these gene sets and downregulation during the later stages (beyond day 3/4). Similarly, the mRNA surveillance pathway and spliceosome proteins were upregulated in day 1, and subsequently downregulated. This suggests that both post-transcriptional and post-translational regulatory mechanisms may play essential roles during neutrophil differentiation, with stage-specific control of different biological macromolecules in order to modulate genetic information processing.

Looking at KEGG ‘Organismal Systems’ terms, gene sets relevant to neutrophil behaviours, such as chemokine signalling, NET formation, and C-type lectin receptor signalling, showed increasing positive enrichment during HoxB8 differentiation (Figure 3.9). Moreover, the NES for the NOD-like receptor (NLR) signalling pathway increased as differentiation progressed. Components of the NOD-, LRR- and pyrin domain-containing protein 3 (NLRP3) inflammasome contributed most to the signal, suggesting upregulation of these components is an aspect of neutrophil differentiation. Additional GSEA of total proteome data can be found in Supplementary Figures 9-13.

3.2.3. Optimisation of protein abundance profile clustering approaches

Using statistical tests that compare two conditions does not consider the temporal aspect of protein abundance changes that may occur during HoxB8 differentiation. Protein synthesis and degradation are dynamic processes that can cause fluctuating protein levels; therefore, it would be advantageous to use an approach that considers the time resolution of the proteomics dataset. To that end, different clustering approaches were employed to identify groups of proteins with similar abundance profiles over time, which can then be interrogated to provide biological insights while incorporating time-resolved information. Biological replicates for each time point were median averaged to reduce potential noise that may influence cluster/module assignment of a protein.

3.2.3.1. WCNA

Taking inspiration from its application in transcriptomic analysis of RNA-seq data, a weighted correlation network analysis (WCNA) approach was used to assess the co-expression of proteins during HoxB8 differentiation. Using the total proteome dataset, further normalisation steps were tested and compared with the processed data to mitigate any potential effects that may influence module assignment (Figure 3.10).

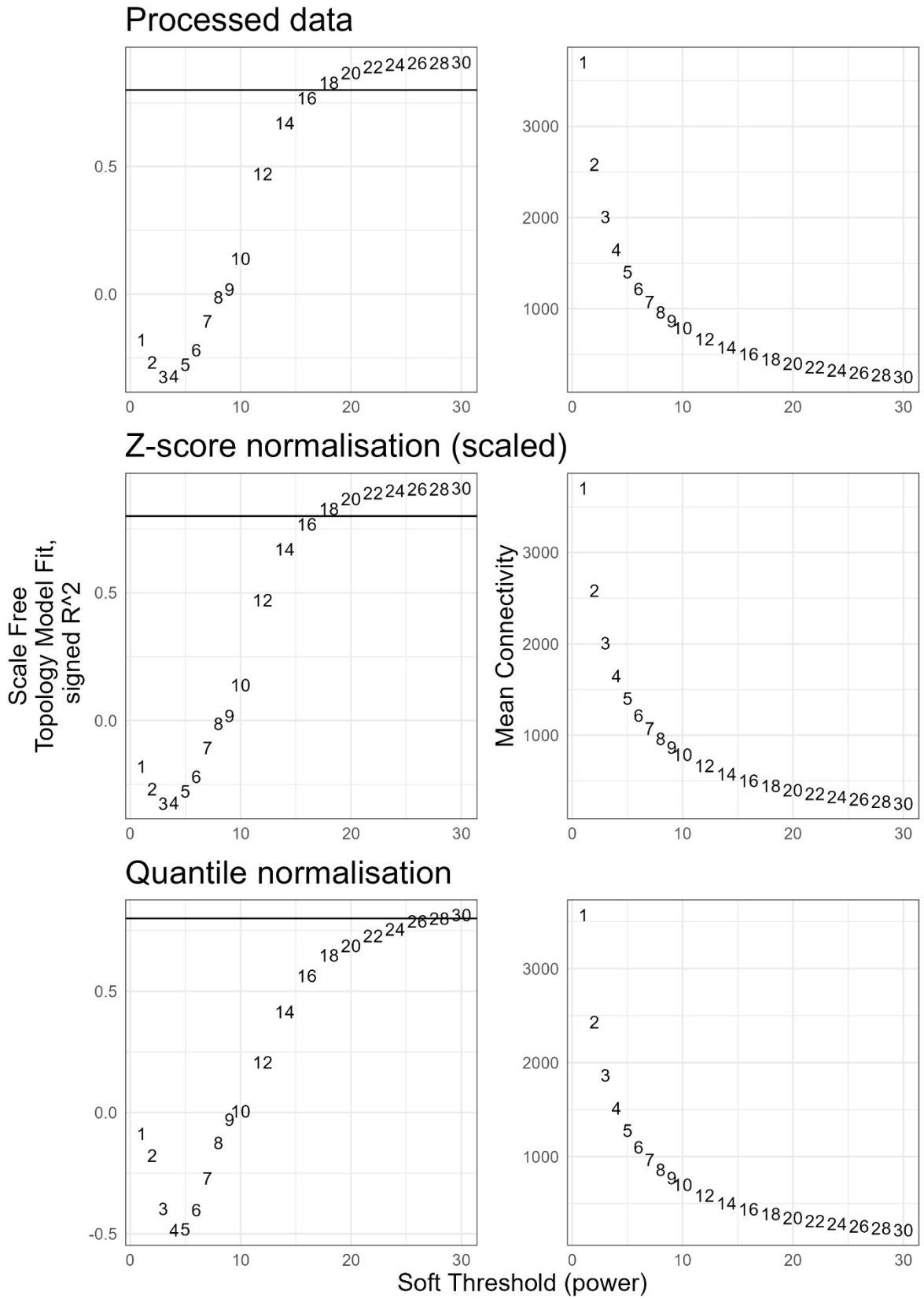


Figure 3.10 ***WCNA soft threshold 'power' parameter optimisation and comparison of different normalisation methods.*** Measures of (left) scale-free topology and (right) mean connectivity for different soft thresholding powers were calculated for the different normalised datasets: (top) processed dataset – or log2-transformed, median-normalised data with imputed values; (middle) Z-score normalisation (scaled), where each protein across all time points will have mean = 0, SD = 1; and (bottom) quantile normalisation, where mean values for each quantile are calculated across all time points, then each protein within a specific time point is assigned the value that corresponds to its quantile.

Z-score normalisation or scaling the data involved centring and adjusting protein values across each time point, resulting in a mean of 0 and standard deviation (SD) of 1 for each protein. This places them on a similar scale, and the resultant co-expression modules would reflect protein trends. However, proteins that exhibited minimal changes during differentiation may not be biologically relevant during HoxB8 differentiation. Z-score normalisation may exaggerate minor differences between time points for these proteins, resulting in protein abundance profiles that mimic those that were indeed differentially abundant during HoxB8 differentiation.

Quantile normalisation is the recommended normalisation method to apply to datasets before performing WCNA. However, the power estimation for soft-thresholding showed lower degrees of scale-free topology for power values within an appropriate range for signed (directional) networks.

The processed data (without further normalisation) was used to perform WCNA since further normalisation methods showed little to no improvement in the scale-free topology of the constructed network. 29 modules were identified using the WCNA approach, based on their temporal profiles, many of which consisted of a small number of proteins (Figure 3.11). Systematic and unbiased selection of modules required correlating modules with external features (differentiation stage, or time point post-induction of differentiation), Therefore, analysis focused primarily on abundance profiles with increasing or decreasing trends. Modules that exhibited high (PCC > 0.6) and low (PCC < -0.6) correlations with time were subjected to functional enrichment analysis using ORA.

Again, modules corresponding to increasing patterns (positive correlation with differentiation stage) demonstrated significant enrichment for terms associated with leukocyte migration, phagocytosis, and cell adhesion (Figure 3.12, Figure 3.13, Figure 3.14). Additionally, glycosylation-, glycoprotein-, and carbohydrate-associated terms were enriched, most likely corresponding to the changes in glycosylation and the glycome of neutrophils during maturation[139]. Decreasing profiles corresponded to ribosomal, mRNA, and cell cycle regulation terms. GSEA of GOCC terms are shown in Supplementary Figure 14.

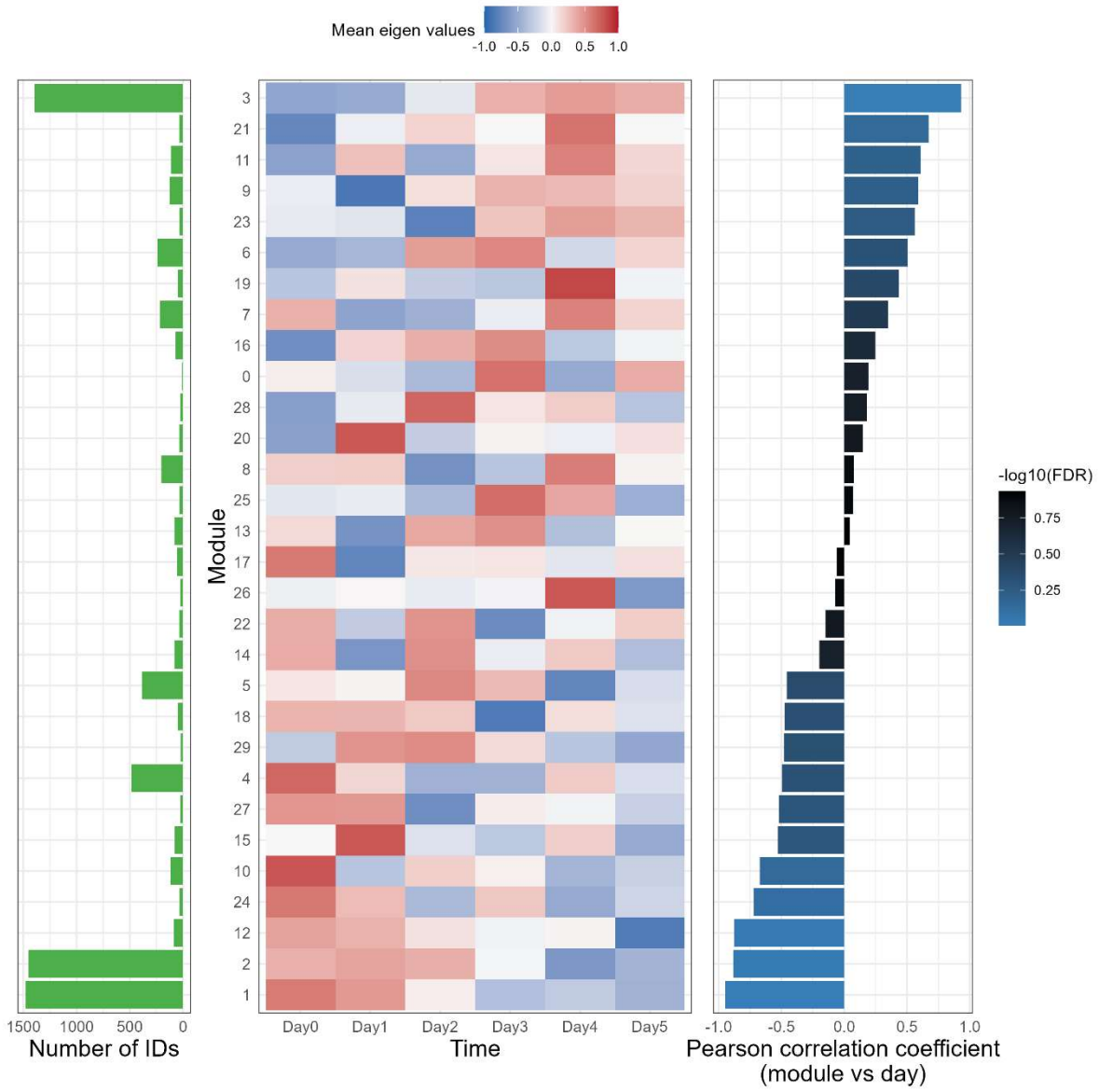


Figure 3.11 ***WCNA identifies modules containing genes with similar expression profiles during HoxB8 neutrophil differentiation.*** (left) Number of protein group IDs per module. (Middle) Eigen protein values for each module. (Right) Pearson correlation between module eigen protein and day (stage of differentiation).

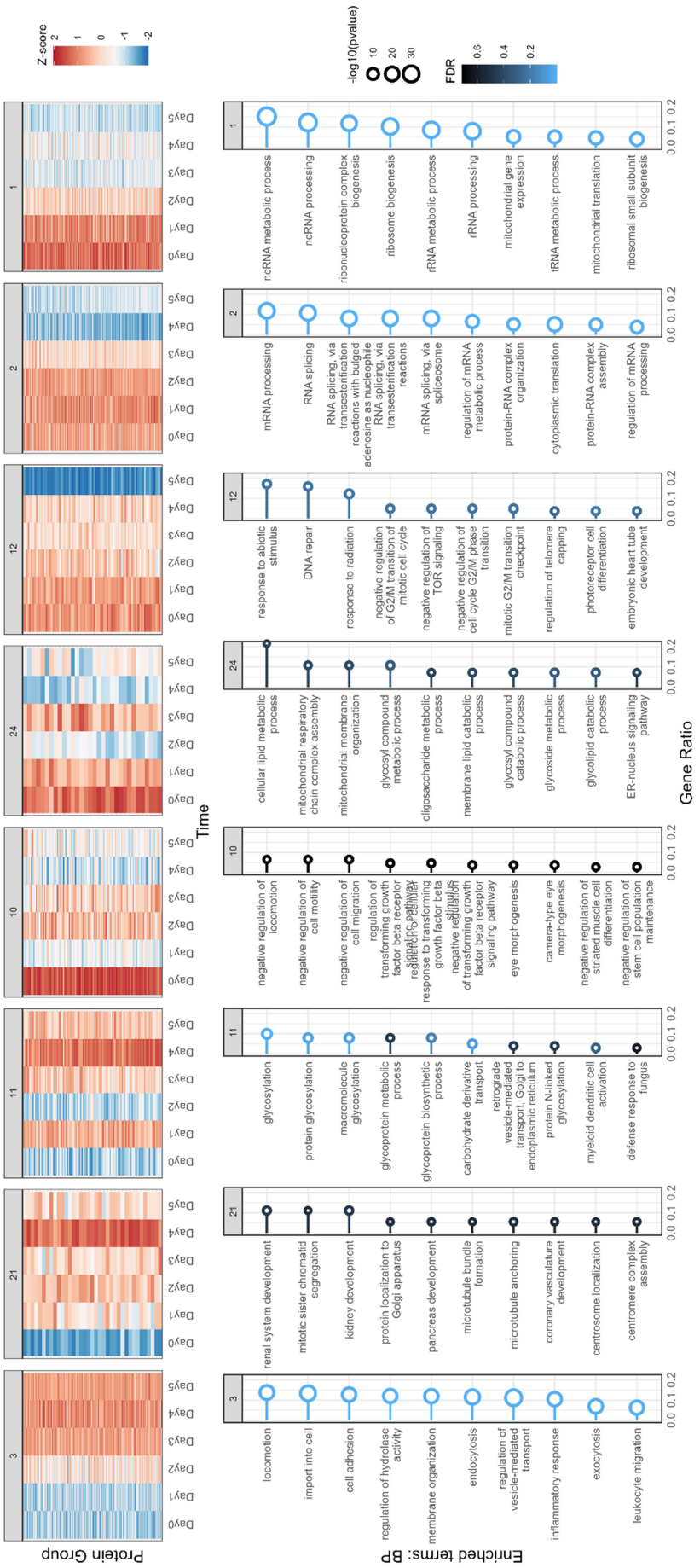


Figure 3.12 ***Functional enrichment analysis of select modules identified from WCNA show upregulation and downregulation of neutrophil-associated biological processes.*** Over-representation analysis was performed on each module identified by WCNA, using the total proteome as the background proteins. Point size and colour represent -log10(p-value) and FDR, respectively. Gene ontology (GO) biological process (BP) terms are shown.

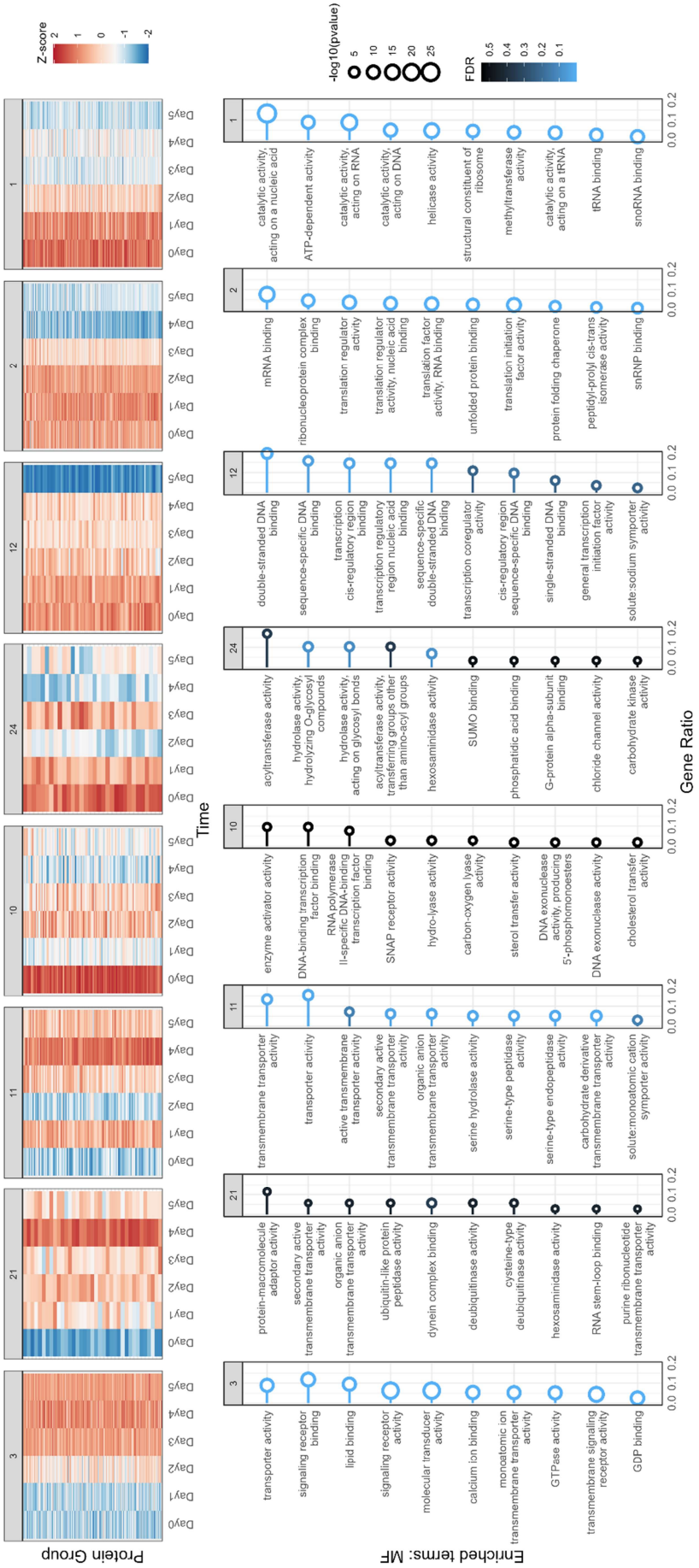


Figure 3.13 ***Functional enrichment analysis of select modules identified from WCNA show upregulation and downregulation of neutrophil-associated molecular functions.*** Over-representation analysis was performed on each module identified by WCNA, using the total proteome as the background proteins. Point size and colour represent -log10(p-value) and FDR, respectively. Gene ontology (GO) molecular function (MF) terms are shown.

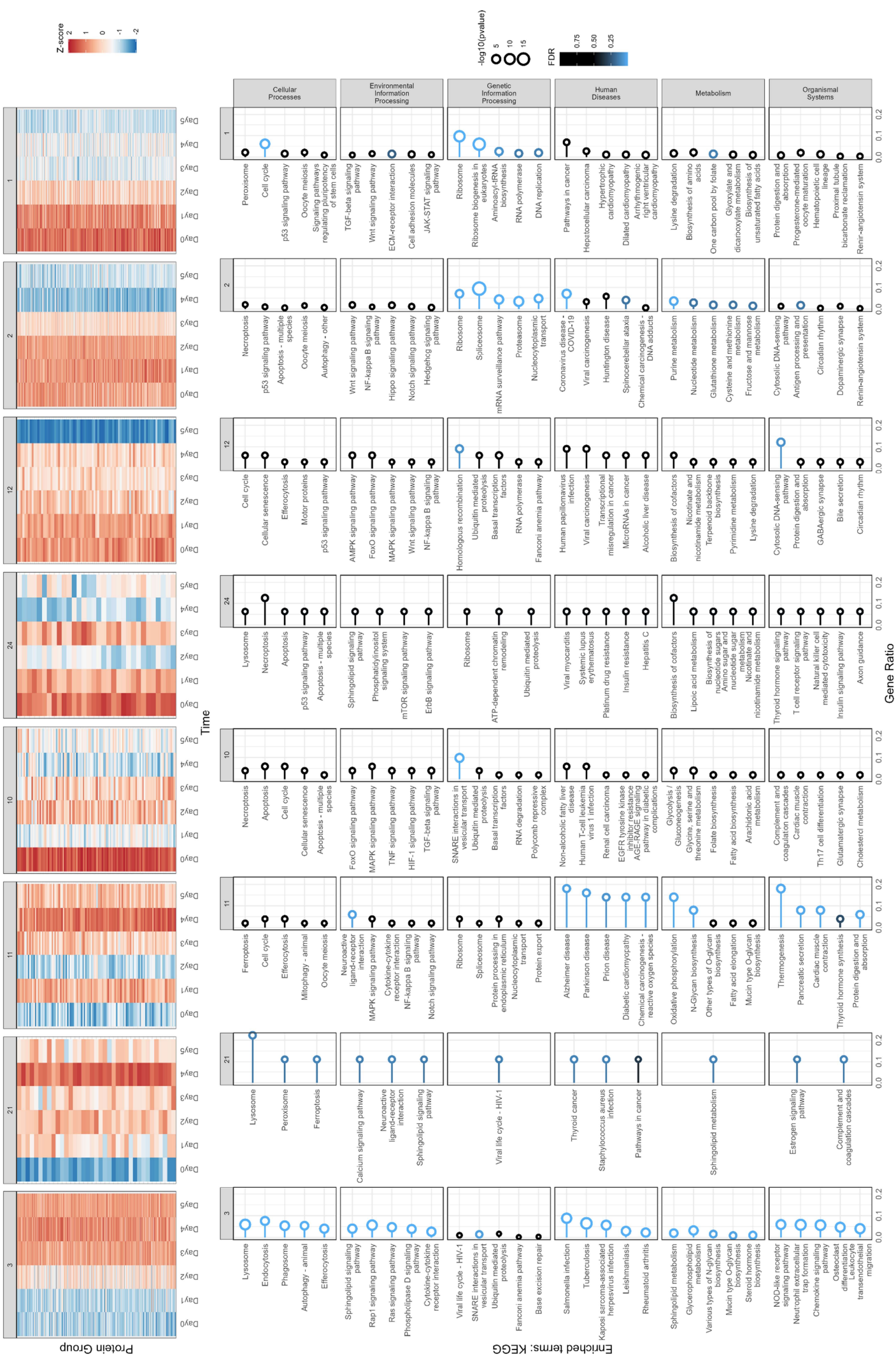


Figure 3.14 ***Functional enrichment analysis of select modules identified from WCNA show upregulation and downregulation of neutrophil-associated pathways.*** Over-representation analysis was performed on each module identified by WCNA, using the total proteome as the background proteins. Point size and colour represent -log10(p-value) and FDR, respectively. KEGG pathways and corresponding subcategories (labelled on the right) are shown.

3.2.3.2. Hierarchical clustering

Due to the potential artefacts introduced by Z-score normalisation, proteins that showed no significant changes (FDR > 0.01, |log2FC| < 1) during any stage of differentiation were removed from further analysis. Z-score profiles were calculated for the remaining proteins, maintaining those proteins that had substantial changes at some point during the HoxB8 differentiation process. Using cluster optimisation algorithms to determine an appropriate k value, a hierarchical clustering approach was used to group Z-score profiles that followed similar temporal trends (Figure 3.15). Functional enrichment analysis on the grouped profiles using ORA was then used to relate the different grouped profiles to biological processes and pathways.

Clusters of proteins showing increasing abundance profiles showed functional enrichment for terms associated with neutrophil behaviours; GOBP gene sets related to cytoskeletal reorganisation, cell migration, phagocytosis, endocytosis, and cytokine production were significantly enriched in clusters 2 and 4 (Figure 3.16). Accordingly, terms associated with ribosome biogenesis, processing, and activities were enriched in profile clusters representing decreasing abundance patterns during differentiation. Interestingly, cluster 3 proteins corresponded to terms associated with hematopoietic and myeloid progenitor cell differentiation and other chromatin- and transcription-related terms. Cluster 6 included terms associated with leukocyte activation, which showed down regulation during day 1 and, to a lesser degree, during day 5, suggesting key stages during neutrophil differentiation where inappropriate activation is prevented. Cluster 3 and 6 represent profiles that fluctuate during the differentiation process, demonstrating biological processes that may require specific modulation during neutrophil differentiation, potentially as a means to temporally regulate pathways that contribute to pre-neutrophil and mature neutrophil phenotypes. Moreover, molecular function (MF) terms associated with protein kinases and kinase binding showed significant enrichment in clusters 2 and 4 (increasing abundance during differentiation), highlighting the role of protein phosphorylation in neutrophil maturation (Figure 3.17). Additional ORA can be found in Supplementary Figures 15-16.

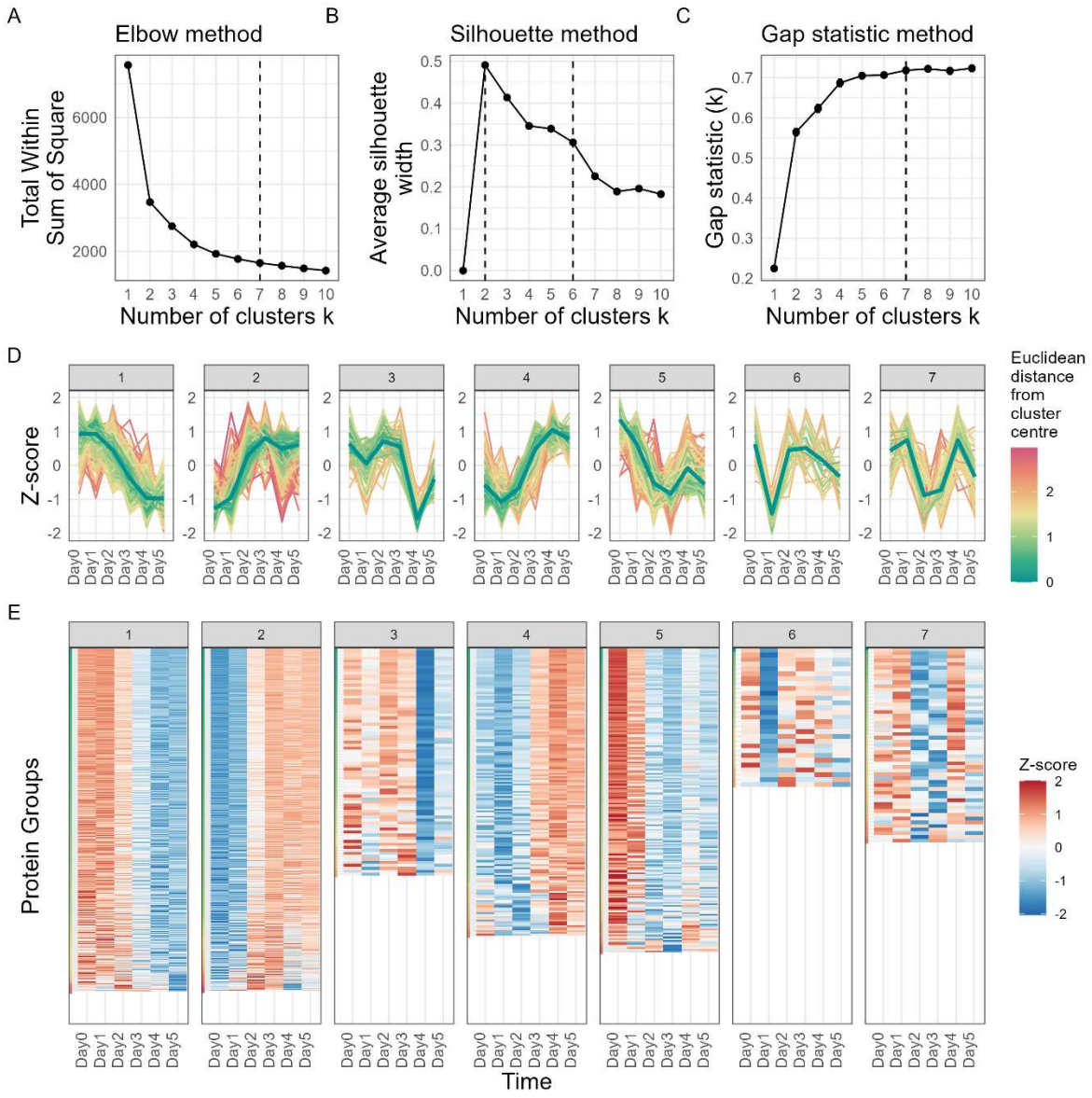


Figure 3.15 ***Hierarchical clustering approach to identify expression profiles of significantly differentially abundant proteins (DAPs).*** Protein values corresponding to the median protein abundance (n = 3 biological replicates) at each time point were Z-scored across all time points. Cluster number optimisation approaches included (A) elbow, (B) silhouette, and (C) gap statistic methods. (D) Protein expression profiles of clusters 1-7. Cluster centre is represented as a thick, dark-green line. (E) Heatmap of clustered proteins.

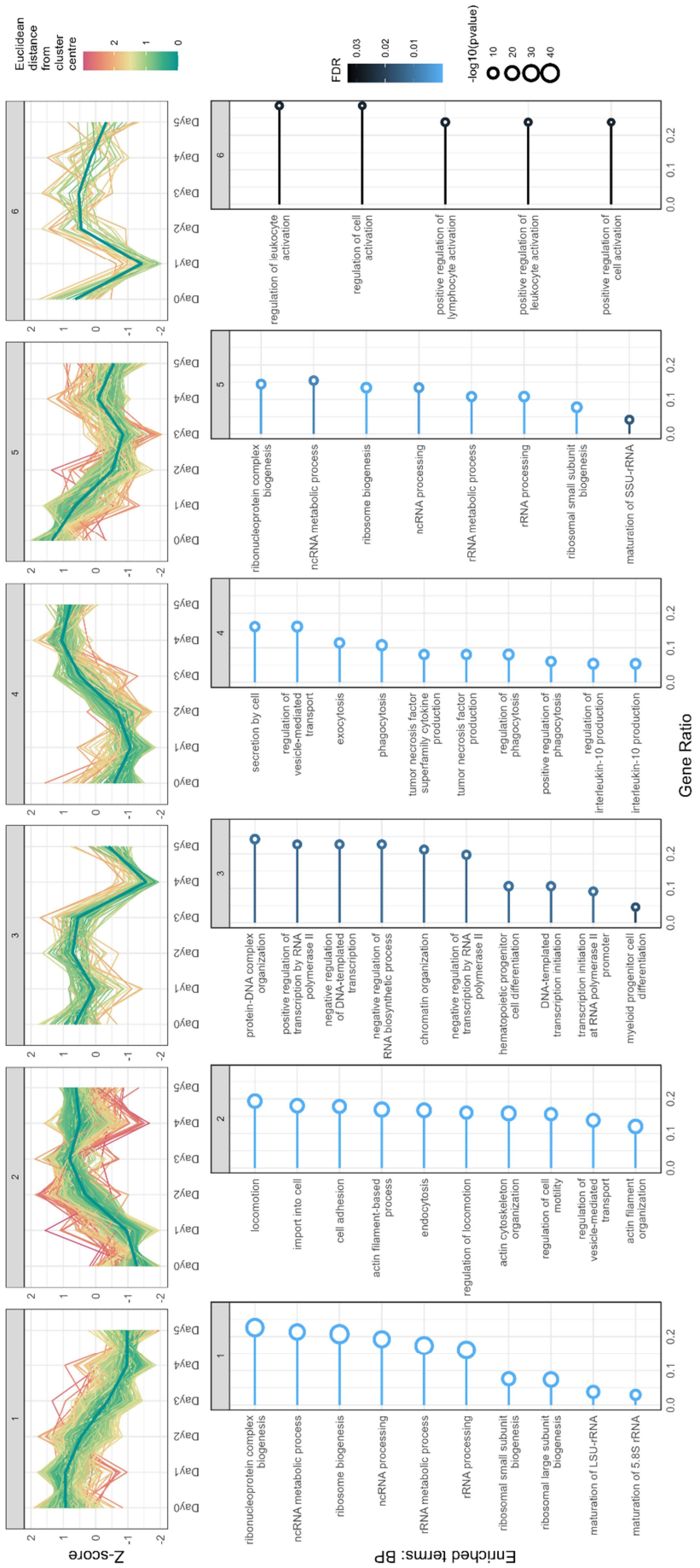


Figure 3.16 ***Functional enrichment analysis of clustered protein profiles of DAPs (GOBP).*** Over-representation analysis was performed on each cluster, using the total proteome as the background proteins. Point size and colour represent -log10(p-value) and FDR, respectively. Gene ontology (GO) biological process (BP) terms are shown.

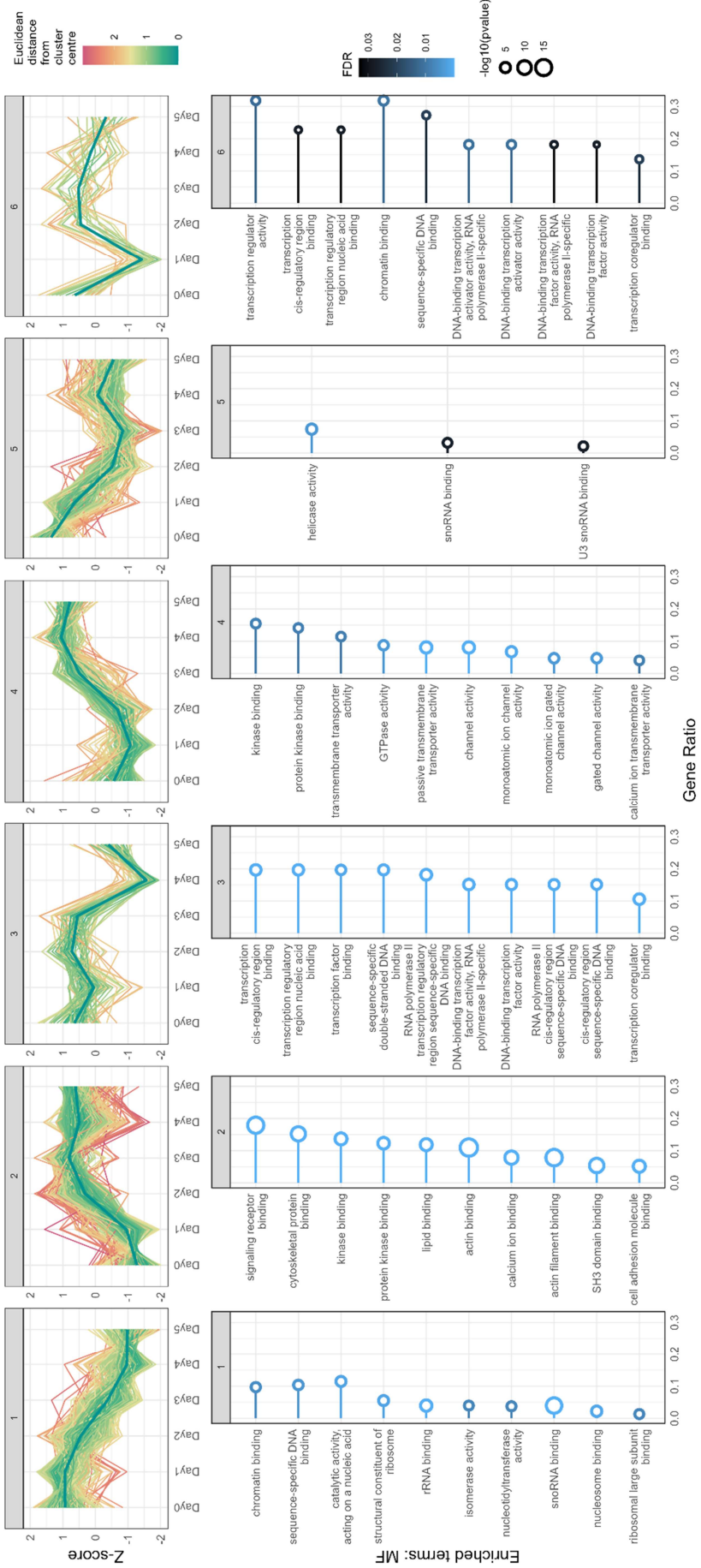


Figure 3.17 ***Functional enrichment analysis of clustered protein profiles of DAPs (GOMF).*** Over-representation analysis was performed on each cluster, using the total proteome as the background proteins. Point size and colour represent -log10(p-value) and FDR, respectively. Gene ontology (GO) molecular function (MF) terms are shown.

3.2.4. Development of enzyme-substrate interaction analysis

GSEA revealed several pathways and molecular functions involving the ubiquitin system and phosphorylation machinery. To identify potential regulators of these PTMs in the context of HoxB8 neutrophil differentiation, ESI analysis was performed (Figure 3.18). This involved hypergeometric testing against various databases containing predicted and/or reported substrates of DUBs, E3 ligases, kinases, and phosphatases, and then applying the BH procedure to control FDR. For example, looking at DUB-substrate interactions, known and predicted ESIs were identified from the UbiBrowser database for DUBs that were detected in the total proteome data. Then, these DUBs and their corresponding substrates were analysed in turn; protein substrates for the DUB of interest were 'marked' in the clustered protein abundance profiles and a hypergeometric test was applied to assess the probability of observing a certain number of marked proteins from each cluster (the 'samples') based on marked proteins within the databases (the 'population'). If there was an enrichment of substrates for a particular DUB (or other PTM enzyme) within a specific abundance profile cluster, this may indicate that an enzyme could be relevant to HoxB8 differentiation since its substrates follow coordinated abundance changes.

Due to limited availability of data pertaining to DUBs and their substrates, many DUB enzymes failed to reach reasonable significance levels (FDR ≤ 0.1), for both reported and predicted ESI databases, and demonstrated low gene ratios (a small proportion of reported/predicted substrates were identified in the corresponding cluster). Nevertheless, of the DUBs identified, a few notable examples include subunit 5 (COPS5) of the COP9 signalosome (CSN) and cylindromatosis (CYLD). The CSN is a protein complex that regulates cullin-RING-E3 ligases (CRL), thus influencing a large proportion of ubiquitination events and transcriptional programs within the cell[140]. Additionally, CYLD participates in inflammation, immunity and cell survival through regulating NF-κB signalling[141]. Despite limited evidence and statistical power, several potential DUBs that regulate proteins were identified, generating potential hypotheses about how specific DUBs may contribute to protein stabilisation, and thus increasing abundance profiles for such proteins, during neutrophil differentiation.

Having identified functional enrichment of kinase-associated MF terms in clusters 2 and 4 (as identified using the hierarchical clustering approach), kinase-substrate and phosphatase-substrate interactions were also interrogated. More information on known and validated kinase-substrate interactions for many kinases is available in the literature. As such, several kinases that have been reported in other neutrophil-associated biological processes had substrates significantly enriched in cluster 2. For example, spleen tyrosine kinase (SYK) and Lck/yes-related novel tyrosine kinase (LYN) both contribute to immune cell signalling[142,143]. The role they may play in neutrophil differentiation remains unclear.

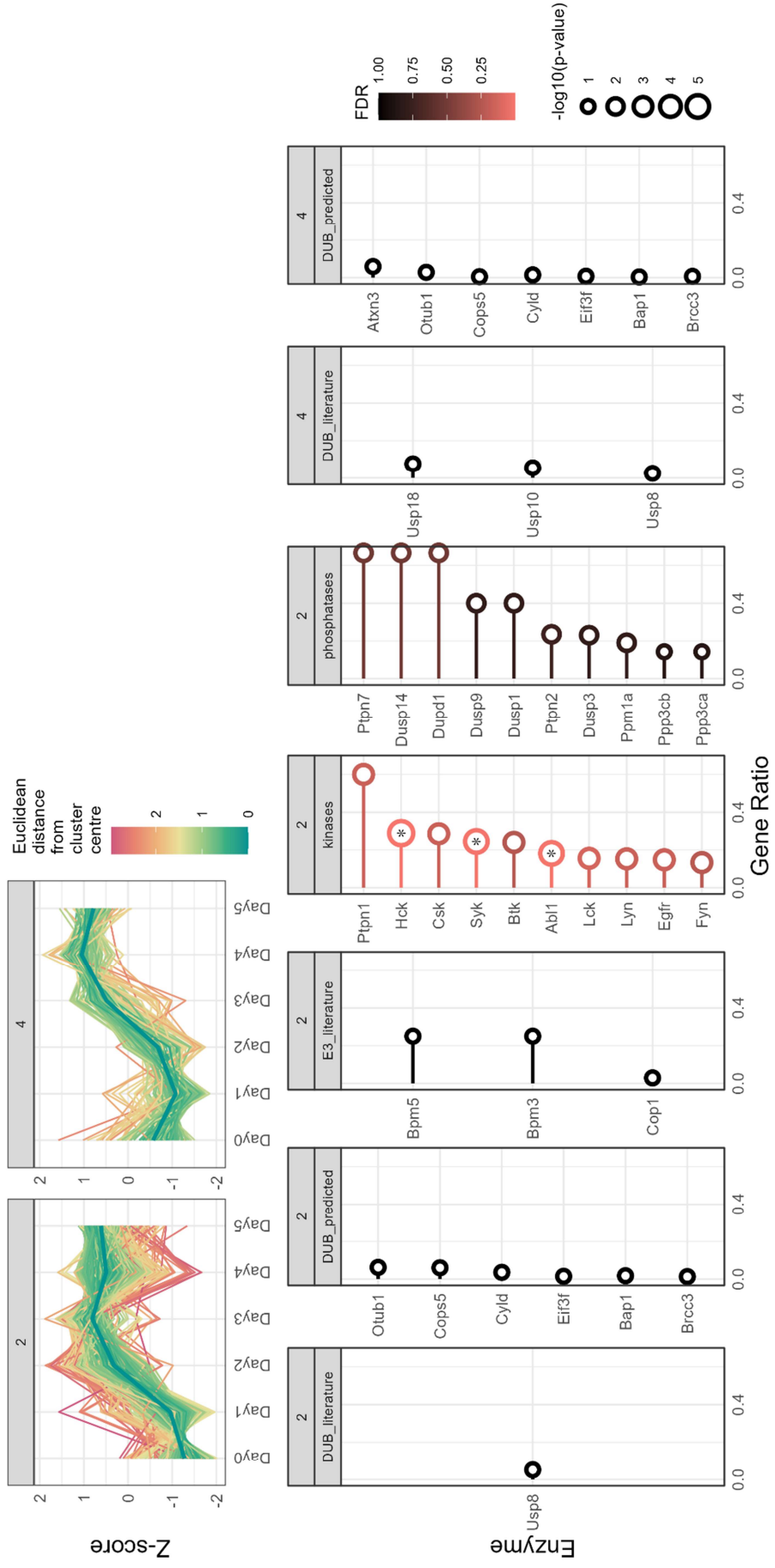


Figure 3.18 ***Enzyme-substrate interaction analysis identifies potential PTM enzymes that may regulate proteins within protein profile clusters.*** Clusters of proteins identified using the hierarchical approach were analysed. Literature/predicted DUB-, literature E3 ligase-, kinase-, and phosphatase-substrate interactions are shown for cluster 2, and only literature/predicted DUB-substrate interactions are shown for cluster 4. Hypergeometric testing with BH procedure for FDR control were used to determine statistical significance.

3.2.5. Power analysis

Power analysis was performed to assess the statistical power attainable to detect effect sizes from the data derived from the HoxB8 model, based on its variation and number of biological replicates (Figure 3.19). The method was adapted from the MSstats R package. For 3 biological replicates and an FDR threshold of 0.01, a power of 0.9 (probability of failing to reject the null hypothesis, $Pr[H_0] = 0.1$) was achieved for $log_2FC \geq 1$ (the effect size). Additionally, 3 biological replicates were sufficient to identify $log_2FC > 0.8$ at a significance threshold of FDR = 0.01 with a statistical power of 0.9, and increasing the number of biological replicates would increase the ability to identify smaller effect sizes with statistical significance.

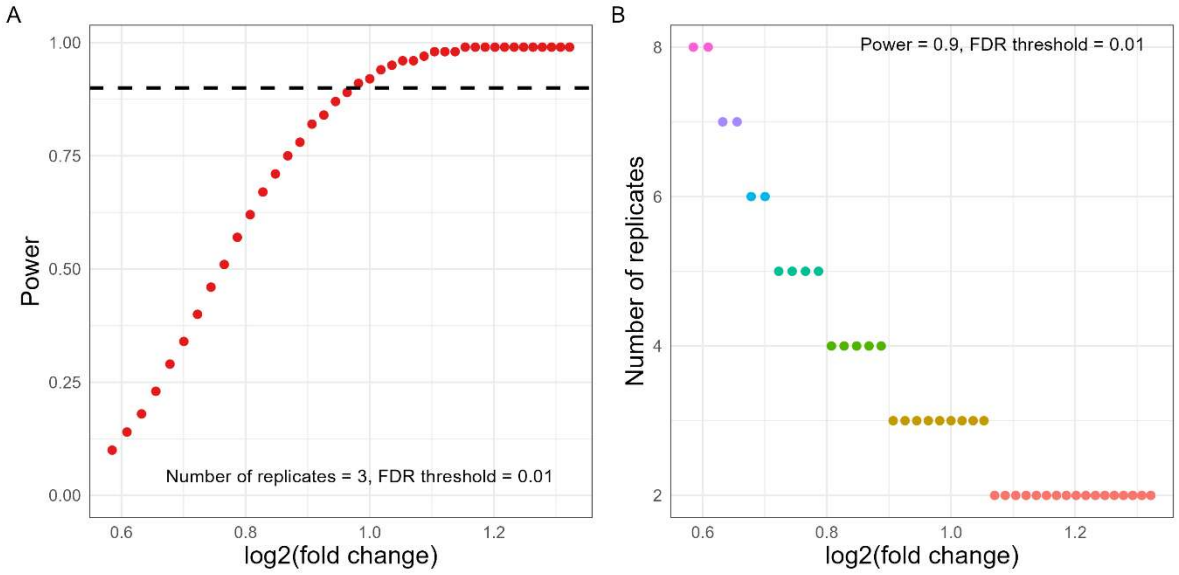


Figure 3.19 ***Power analysis of total proteome data from HoxB8 neutrophil differentiation.*** (A) Statistical power attainable at different log2(fold change) (effect sizes). (B) Number of replicates required to achieve statistical power ≥ 0.9 for different log2(fold change).

3.3. Discussion

MS-based proteomics is indispensable in characterising the protein changes that occur during cell differentiation processes in an unbiased fashion, able to identify and quantify several thousands of cellular proteins. Median normalisation of label-free MS data is a robust approach to correct for technical variation between different samples, allowing for comparisons between different samples to gain biological insight. Analysis of proteomic data derived from the HoxB8 mouse model of neutrophil differentiation captured differentially abundant proteins and modulated biological processes and pathways consistent with the current understanding of primary neutrophil biology, differentiation, and effector functions. The model is also robust with high reproducibility between biological replicates, allowing for decent statistical power. Higher variability in protein quantitation in day 5 of the total proteome analysis of HoxB8 differentiation may be indicative of certain biological replicates exhibiting higher levels of cell death and/or increased cellular debris, and subsequently potential activation of day 5 HoxB8 cells. Alternatively, despite normalisation of cell numbers and protein amounts, MS analysis may have involved injection of less material into the column leading to an increase in missing values for a particular replicate.

Identification of modulated PTM-associated pathways, particularly ubiquitination and phosphorylation, during HoxB8 neutrophil differentiation suggests that these PTM events are also controlled during, and may contribute to, the neutrophil maturation process. To better understand the potential role of ubiquitination and phosphorylation in HoxB8 neutrophil differentiation, MS-based ubiquitomic and phosphoproteomic profiling would reveal a global overview of sustained PTM changes that are characteristic of different stages of neutrophil differentiation.

Clustering methodologies were implemented to assess the suitability of the workflows to the HoxB8 model and future PTMomic approaches, while considering assumptions and limitations. WCNA identified a large number of modules (29) consisting of variable numbers of proteins. To minimise potential loss of information regarding time-specific modulation of protein abundance during the differentiation process, identification of unique or ‘interesting’ modules would necessitate visual inspection (introducing possible bias), or appropriate metrics that can capture this information in an unbiased manner. Additionally, certain proteins assigned to a module may not be changing in their relative abundances – suggesting they are not functionally relevant – during the differentiation process yet would contribute to signals in downstream analyses of such modules. Accordingly, module assignment and membership will vary based on initial parameters supplied to the algorithms, therefore optimisation may be required to coincide with the data structure and mathematical assumptions.

WCNA relies on high degrees scale-free topology and low degrees of mean connectivity among genes/proteins within the dataset. This would be applicable in shotgun approaches that do not introduce any evident skew or bias into the datasets. However, as MS-based analysis of PTMs rely on enrichment steps to capture specific PTMs or PTM-remnants, this inherently skews the data and affects the distribution of intensity values of precursors/peptides. Therefore, WCNA would be unsuitable for datasets involving peptide-level enrichment steps since assumptions of scale-free topology and mean connectivity would be violated.

Hierarchical clustering of DAP profiles proved useful in the identification of fewer profile clusters with unique and interesting temporal features, allowing for simpler computation and interpretation of downstream analyses, while maintaining biological information relevant to neutrophil differentiation. No assumption of the overall data structure is required, only a single parameter (number of clusters, k) requires optimisation, and using a deterministic approach such as hierarchical clustering allows for reproducible and consistent analysis.

ESI analysis provided predictions and testable hypotheses related to the potential roles of PTM enzymes in neutrophil differentiation. However, the ability to make such predictions and attribute statistical significance relies heavily on pre-existing data and databases that consist of well-documented and validated substrates of different PTM enzymes.

Neutrophil extraction can sometimes lead to activation, and their high turnover rate and short lifespans *in vivo* can pose several challenges during experimentation. Using the HoxB8 neutrophil model of differentiation overcame these associated challenges, enabling tractable differentiation of neutrophil progenitor cells, and sufficient material for bulk MS-based proteomics analysis. However, there are several limitations to consider. The neutrophil differentiation and maturation process primarily occurs in the bone marrow, a tissue consisting of a variety of cell types and structural, nutritional, and metabolic features that may contribute to neutrophil development. Therefore, *ex vivo* manipulation and induction with higher amounts of G-CSF compared to those found in the bone marrow cannot recapitulate the physiological niche occupied by developing neutrophils. Additionally, there may be unnoticed secondary effects resulting from the artificial expression of ER-HoxB8 and the unnatural induction of differentiation. Despite the HoxB8 model recapitulating key aspects of neutrophil differentiation and biology, tissue and organismal system features are missing.

Lastly, the proteomic data of HoxB8 neutrophil differentiation provided information about reproducibility and statistical power. This information will be useful when designing PTMomics experiments since peptide-level enrichment steps can introduce technical variation through e.g. variable enrichment efficiencies, thus augmenting the overall variability between replicates, which

already exhibit a certain degree of underlying biological variation. The number of biological samples is important to consider in order to detect statistically significant effect sizes but also to design cost-efficient experiments. This preliminary analysis of the HoxB8 model demonstrated 3 biological replicates being sufficient to reach a statistical power of 0.9 for log2FC ≥ 1. To improve the reproducibility of the data, an additional biological replicate (n=4) would provide more confidence in identification and quantitation. Fewer time points corresponding to more distinct stages of HoxB8 differentiation could be selected as a compromise to increased number of replicates. This can be inferred from PCCs between time points and the groupings on a PCA plot. Most distinct changes in the total proteome seem to occur between day 1 and 2, with notable differences between day 3 and 4. Accordingly, day 1 and 3 would represent key intermediate stages during HoxB8 differentiation, and days 0 and 5 would correspond to HoxB8 progenitor cells and mature HoxB8 neutrophils, respectively.

Chapter 4: Ubiquitomic and phosphoproteomic changes during HoxB8 mouse neutrophil differentiation

4.1. Introduction

This chapter investigates the molecular changes associated with neutrophil differentiation in the HoxB8 model using MS-based approaches, focusing on the temporal dynamics of ubiquitination and phosphorylation.

PTMs are important molecular changes that modulate protein behaviours, such as function, localisation, activation status, and stability. Ubiquitination is an example of a proteinaceous PTM, where a small 8.6 kDa protein called ubiquitin (Ub) is covalently linked to lysine side chains of target proteins. Ub conjugation occurs via a cascade of E1 ubiquitin-activating, E2 ubiquitin-conjugating, and E3 ubiquitin-protein ligase enzymes. Further ubiquitination can occur on conjugated Ub, resulting in chains with diverse architectures and constituents; additional ubiquitination can occur on several lysine residues on the surface of ubiquitin protein, and Ub-like proteins such as small ubiquitin-like modifier (SUMO) can be conjugated to proteins or incorporated into chains. This diverse PTM 'language' highlights the biological complexity of protein modification landscape. Ubiquitination plays a vital role primarily in protein degradation, marking proteins to be processed by degradation pathways. Additionally, protein ubiquitination can affect localisation, activation state, and protein-protein interactions. Ub can be removed from modified proteins through the activity of deubiquitinases (DUBs), which is typically associated with stabilisation of protein targets.

Phosphorylation is another indispensable PTM that regulates protein function. Addition of phosphate groups, typically to serine, threonine, or tyrosine side chains, can lead to diverse protein behaviours through conformational change, altered protein-protein interactions, and other mechanisms that modulate protein function. Phosphorylation is typically associated with signalling pathways but can also play a role in protein scaffolds and protein complex formation.

Establishment of MS-based proteomic analysis of the HoxB8 model of neutrophil differentiation in Chapter 3 revealed valuable data for experimental design of ubiquitomic and phosphoproteomic workflows. Enriching for ubiquitinated and phosphorylated peptides derived from trypsin digestion of protein lysates, corresponding to key stages of HoxB8 neutrophil differentiation, allows identification and quantitation of these PTMs. Data acquired provides biological insight into the PTM landscape (PTMome) associated with neutrophil development and maturation.

4.2. Results

Whole cell lysates were collected at days 0, 1, 3, and 5 post-induction of HoxB8 neutrophil differentiation, in quadruplicate (4 biological replicates). After on-column protein digestion, resultant peptides from each biological replicate were split into three aliquots for 'total proteome,' 'ubiquitome,' and 'phosphoproteome' analysis. Ubiquitome analysis utilised immunoprecipitation (IP) of modified peptides containing a remnant glycine-glycine motif on ubiquitinated lysine side chains (K-GG, diGly peptides, or Ub-peptides). Phosphoproteomic analysis involved TiO_2 enrichment of phosphorylated peptides, using a Bravo liquid handling robot. Total proteome and ubiquitome samples were analysed by MS in DIA mode, while the phosphoproteome samples were analysed in both DDA and DIA modes.

Raw DIA MS data were processed using DIA-NN v2.1 for identification and label free quantitation of Ub-peptides (ubiquitome data), phosphorylated peptides (phosphoproteome data), or peptides with corresponding protein inference (proteome data). Raw DDA and DIA phosphoproteomic MS data were analysed using FragPipe v23.1 or DIA-NN v2.1. Search outputs were filtered to remove IDs that did not appear in at least 3 out of 4 replicates, in at least one time point. Filtered data were log_2-transformed and median normalised. Missing values for PTMome datasets were imputed using a conditional approach. If intensity values were missing from all biological replicates for a given time point (considered 'absent'), values were imputed from a downshifted normal distribution to emulate low abundant proteins/peptides. Then, k-nearest neighbours (KNN) imputation was performed to impute values for the remaining missing values based on similarity with other modified peptides. Missing values in the total proteome data were replaced solely by imputation from a downshifted normal distribution. Log_2-transformed, median normalised data with imputed values are henceforth referred to as 'processed data'. QC of the data involved calculating CV values for each protein per time point; inspecting the distributions of log_2(protein intensities) and processed data; calculating PCC between biological replicates for each time point using the dataset with imputed values; and performing PCA on the data with imputed values to assess the impact of normalisation and imputation.

An overview of the experimental workflow is depicted in Figure 4.1.

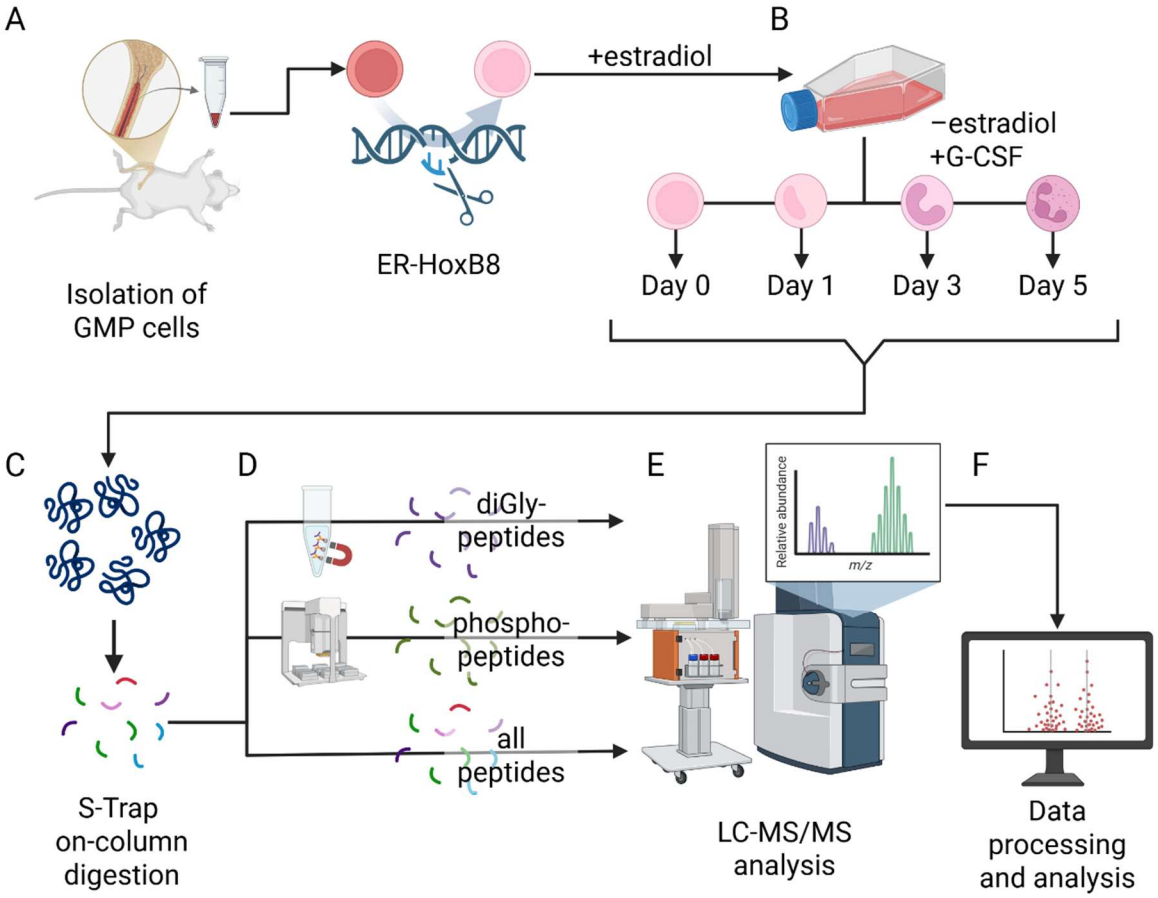


Figure 4.1 ***Experimental workflow for proteomic, ubiquitomic, and phosphoproteomic analysis of HoxB8 neutrophil differentiation.*** (**A**) Granulocyte-monocyte progenitor (GMP) cells were isolated from murine bone marrow and genetically modified with estrogen receptor (ER)-HoxB8 fusion protein. In the presence of estradiol, HoxB8 GMPs can be cultured *ex vivo*. (**B**) Upon estradiol removal and granulocyte colony stimulating factor (G-CSF) induction, HoxB8 GMPs can be differentiated into HoxB8 neutrophils. Cell pellets were collected on day 0, 1, 3, and 5 post-induction of differentiation. (**C**) All samples were processed using S-Trap on-column digestion. (**D**) Resultant peptides were split into aliquots for ubiquitomic (diGly-peptides), phosphoproteomic (phosphopeptides) or proteomic (all peptides) analysis. (**E**) Peptide samples were analysed using liquid chromatography tandem mass spectrometry (LC-MS/MS). (**F**) Raw mass spectrometry data were processed and analysed using bioinformatic approaches. This figure was created in BioRender.

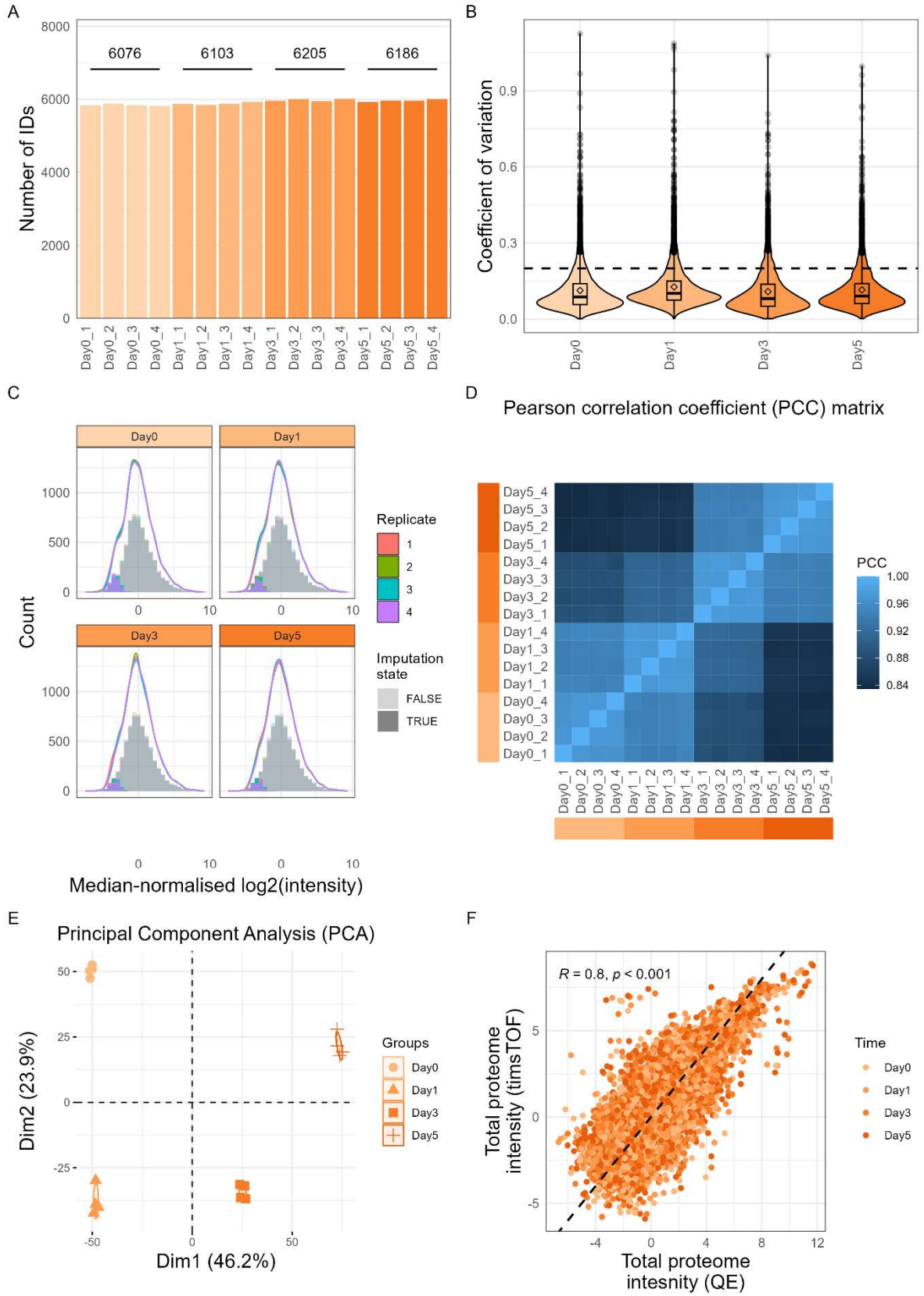


Figure 4.2 ***QC of proteome data of HoxB8 neutrophils during differentiation.*** (**A**) Number of protein group IDs per biological replicate, with total IDs per time point shown above each group (n = 4 biological replicates). (**B**) Violin plots showing the distribution of coefficient of variation (CV) values per time point. (**C**) Distribution of median-normalised, log2-transformed protein intensities with imputed missing values (processed data). Imputed values were sampled from a downshifted normal distribution to emulate low protein abundances (these values are shown as more opaque sections of their corresponding histogram) (**D**) Pearson correlation coefficient between biological replicates and time points of the processed data. (**E**) Principal component analysis of the processed data. (**F**) Comparison of HoxB8 proteomics using different MS instruments for analysis.

Proteome analysis showed similar data structure to initial proteome characterisation in Chapter 3. Protein IDs were consistent across time points and replicates, with protein quantitation showing low CVs within each time point (Figure 4.2A, B). Processed data showed normal distribution of intensity values, with minimal imputation (Figure 4.2C). PCCs between replicates and time points and PCA showed groupings of biological replicates for distinct stages of HoxB8 differentiation (Figure 4.2D, E). These findings are consistent with the previous proteome analysis and confirm the reproducibility of the HoxB8 model. MS analysis in this chapter was performed on a TimsTOF mass spectrometer therefore the proteome dataset was compared to the initial pilot HoxB8 proteome experiment that was performed using a Q-Exactive mass spectrometer (QE) (Figure 4.2F). Amalgamating median intensity values for each protein and time point common to both datasets, a PCC of 0.8 was calculated ($p < 0.001$). The high positive correlation indicates quantitation of the HoxB8 model is generally consistent between the QE and TimsTOF.

4.2.1. Ubiquitome analysis reveals dynamic changes in ubiquitination landscape

Ubiquitome analysis identified 54,297 Ub-modified precursors, corresponding to 40,974 modified peptides. After filtering for 3 out of 4 valid values within at least 1 time point and removing modified peptides in the top CV decile, 32,344 proteotypic Ub-peptides were maintained for further analysis. The number of Ub-peptide IDs generally increased during HoxB8 differentiation (Figure 4.3A), indicating increase in protein ubiquitination events as HoxB8 neutrophils mature. CV values of modified peptides were higher compared to protein group CVs calculated in the proteome analysis (Figure 4.2B, Figure 4.3B), which was expected due to the diGly-enrichment step leading to more variability. Using the conditional imputation approach, normalised peptide intensities with imputed values showed slightly different distributions at various time points, most likely due to changes in modified peptide ID numbers and how missing values were treated (Figure 4.3C, D). Nevertheless, PCCs and PCA demonstrated distinct groupings of replicates reflective of different stages of differentiation (Figure 4.3E, F).

Volcano plots of moderated t-test results, comparing each time point to day 0, showed increase in significantly upregulated modified peptides over time (Figure 4.4). A summary of the moderated t-test output and numbers of differentially abundant modified peptides (DAPeps) are provided in Table 4.1.

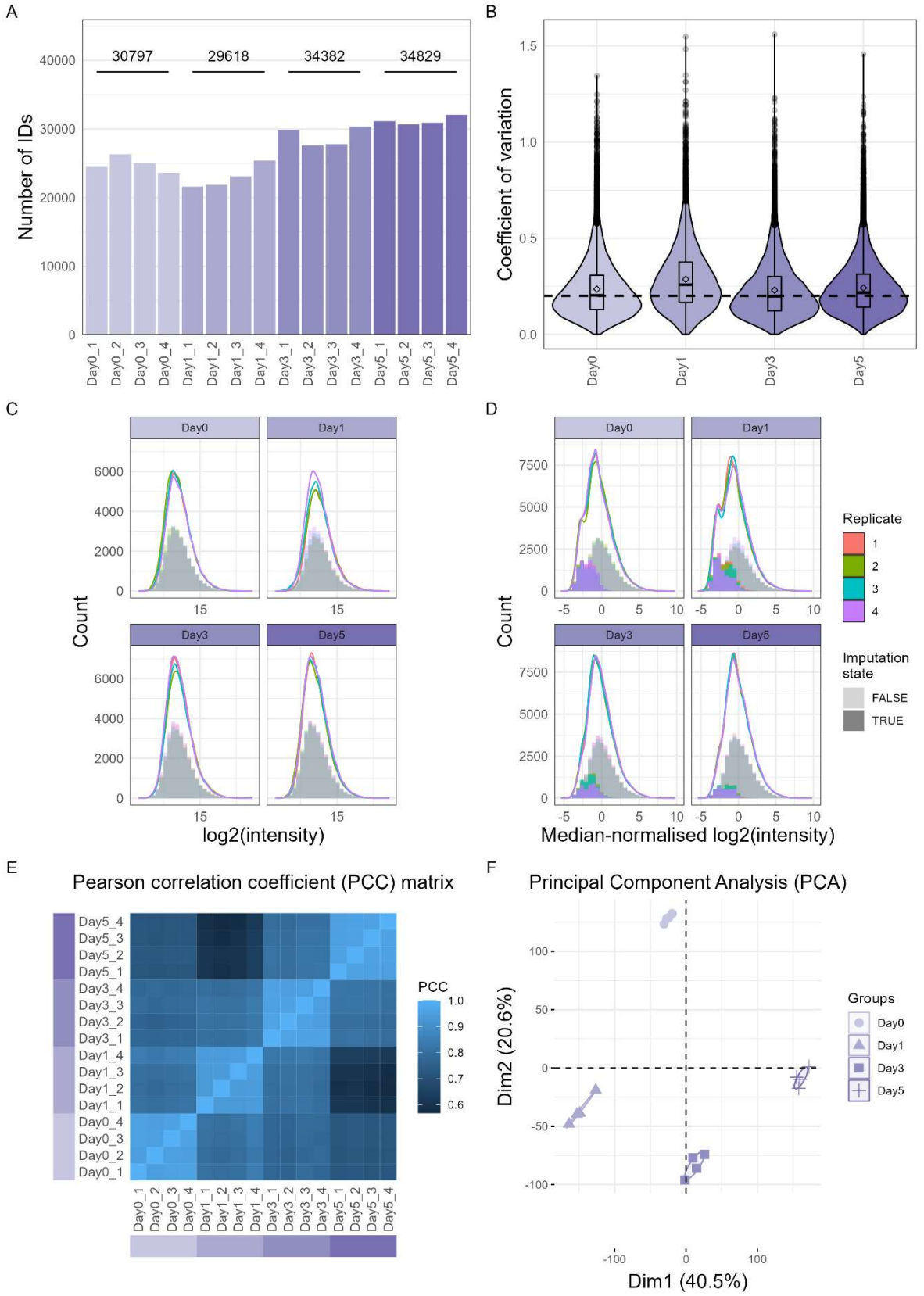


Figure 4.3 ***QC of ubiquitome data of HoxB8 neutrophils during differentiation.*** (A) Number of modified peptide IDs per biological replicate, with total IDs per time point shown above each group (n = 4 biological replicates). (B) Violin plots showing the distribution of coefficient of variation (CV) values per time point. (C) Distribution of log2(intensities) of modified peptides. (D) Distribution of median-normalised, log2-transformed protein intensities with imputed missing values (processed data). Imputed values are shown as more opaque sections of their corresponding histogram. (E) Pearson correlation coefficient between biological replicates and time points of the processed data. (F) Principal component analysis of the processed data.

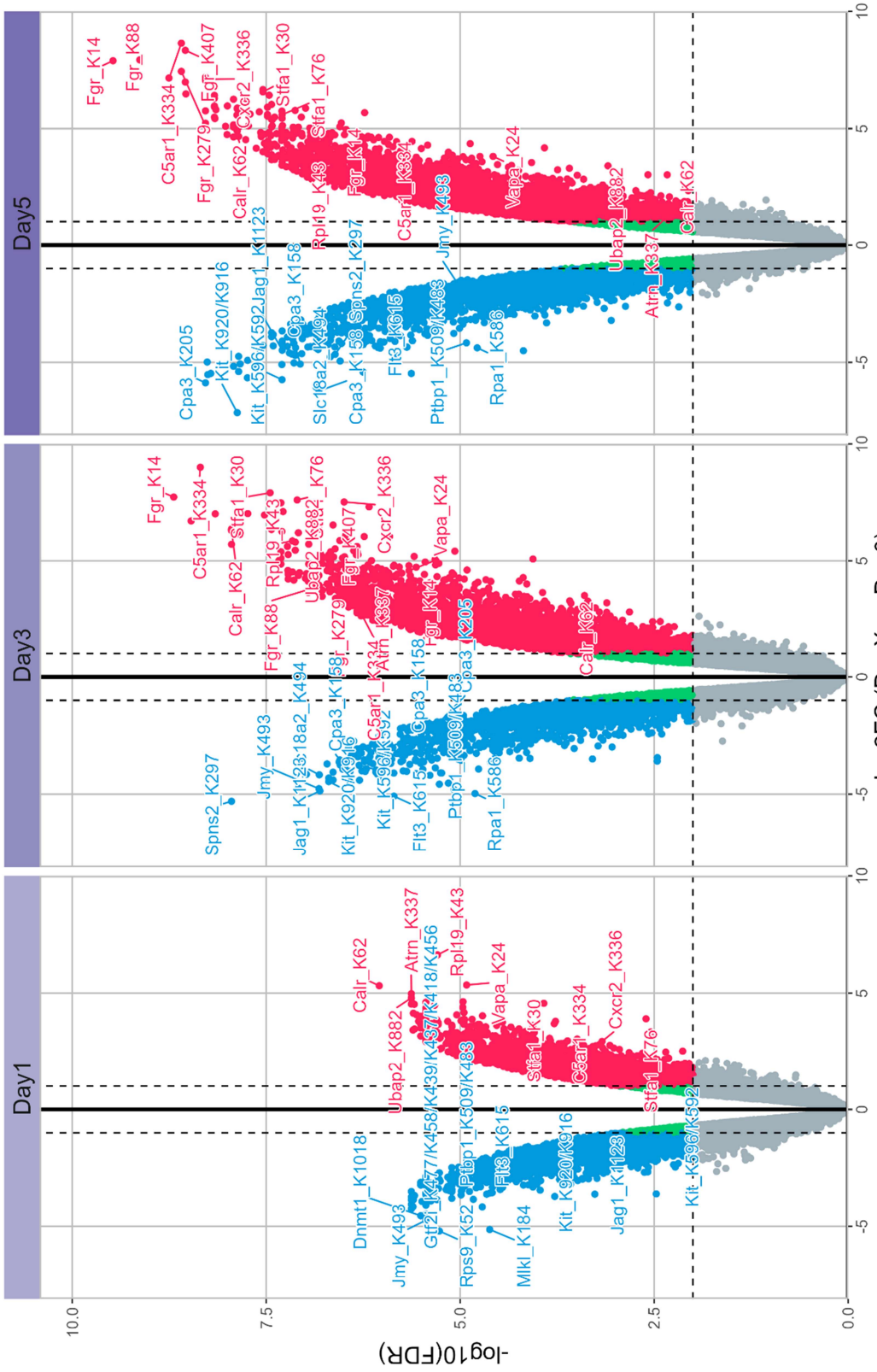


Figure 4.4 ***Volcano plots of differentially abundant modified peptides (DAPeps) identified in the ubiquitomes of different stages of HoxB8 neutrophil differentiation and day 0.*** Moderated t-test was used to compare groups (n = 4 biological replicates),, and the Benjamini-Hochberg procedure was used to control false discovery rate (FDR). Peptides with FDR < 0.01 and |log2(fold change)| > 1 were considered statistically significant and differentially abundant. Upregulated and downregulated peptides are shown in red and blue, respectively. Green represents peptides with FDR < 0.01 and |log2(fold change)| < 1.

Table 4.1 ***Summary of moderated t-test outputs for ubiquitome data of HoxB8 differentiation.*** *Each time point was compared to day 0.*

Time	Downregulated log2FC < -1, FDR < 0.01	Unchanged |log2FC| < 1	Upregulated log2FC > 1, FDR < 0.01
Day 1	4,427	24,524	3,393
Day 3	2,144	24,333	5,867
Day 5	3,330	20,850	8,164

Several Ub-peptides of the chemokine (C-X-C motif) receptor 2 (CXCR2) were differentially abundant during differentiation. Peptides corresponding to the sites K282, K336, and K341 were upregulated in day 3 and 5, compared to day 0. Little is known about ubiquitination at these sites, whereas K327 ubiquitination was shown to affect CXCR2-mediated cell signalling in human cells[144]. Only K327, K336, and K341 are conserved residues between human and mouse homologs.

Jagged1 (JAG1)-derived Ub-peptides corresponding to K1123 and K1137 showed decreased abundances during HoxB8 differentiation. Mutations in JAG1 have been observed in cases of congenital neutropoenia and associated with myeloid cell death[145]. Ubiquitination may play a role in stabilising JAG1 for appropriate neutrophil differentiation.

The non-receptor tyrosine-protein kinase FGR has been shown to be involved in neutrophil effector functions[146]. Several modified peptides were identified in the ubiquitome data, corresponding to the sites K14, K88, K91, K136, K190, K279, K282, K289, K299, K407, and K411. All sites showed increase in abundance, indicating FGR is ubiquitinated at multiple sites during HoxB8 differentiation. However, the role of ubiquitination in regulating FGR is yet to be studied.

GSEA of genes corresponding to ubiquitin-modified peptides was performed. Genes were ranked based on their associated ubiquitin-modified peptide $\log_2$FC in intensity for each time point, relative to day 0. GOBP terms showed positive enrichment for terms such as ubiquitin- dependant and/or proteasome-mediated protein catabolic processing and proteolysis, and negative enrichment for mitochondrial processes (Figure 4.5). KEGG 'Organismal System' terms that demonstrated positive enrichment included NOD-like receptor and chemokine signalling, phagocytosis, leukocyte transendothelial migration, and antigen processing (Figure 4.6). Interestingly, majority of terms showed decreasing NES as time of HoxB8 differentiation increased, suggested there is global decrease in ubiquitination of proteins associated with these pathways.

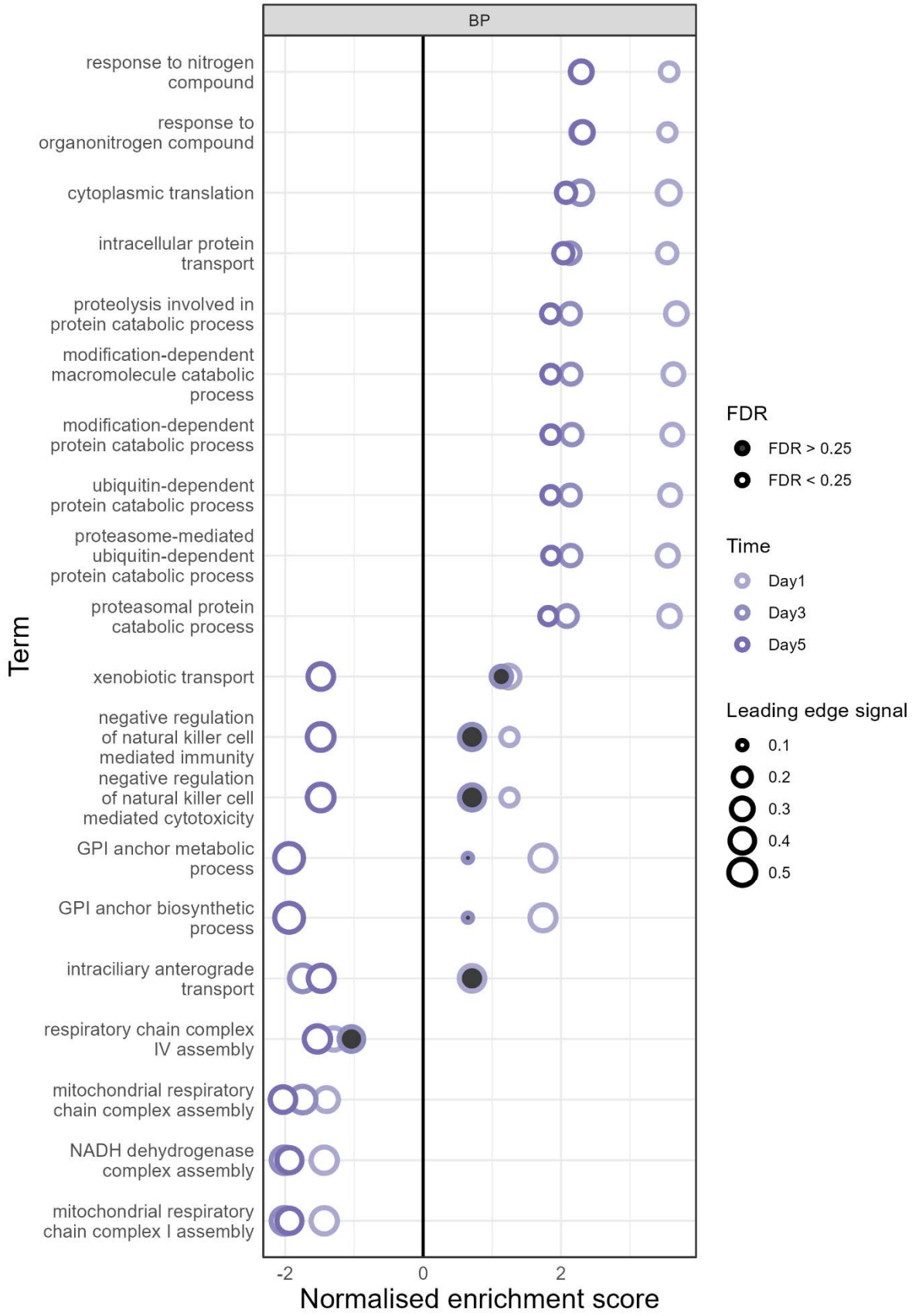


Figure 4.5 ***Gene set enrichment analysis (GSEA) of ubiquitome data from HoxB8 neutrophil differentiation (GOBP).*** Genes were pre-ranked based on corresponding modified peptide log2(fold change) compared to day 0. Stroke colour represents the time point, size indicates leading edge signal, and permutation-based FDR is indicated by fill colour. Gene ontology (GO) biological process (BP) terms are shown.

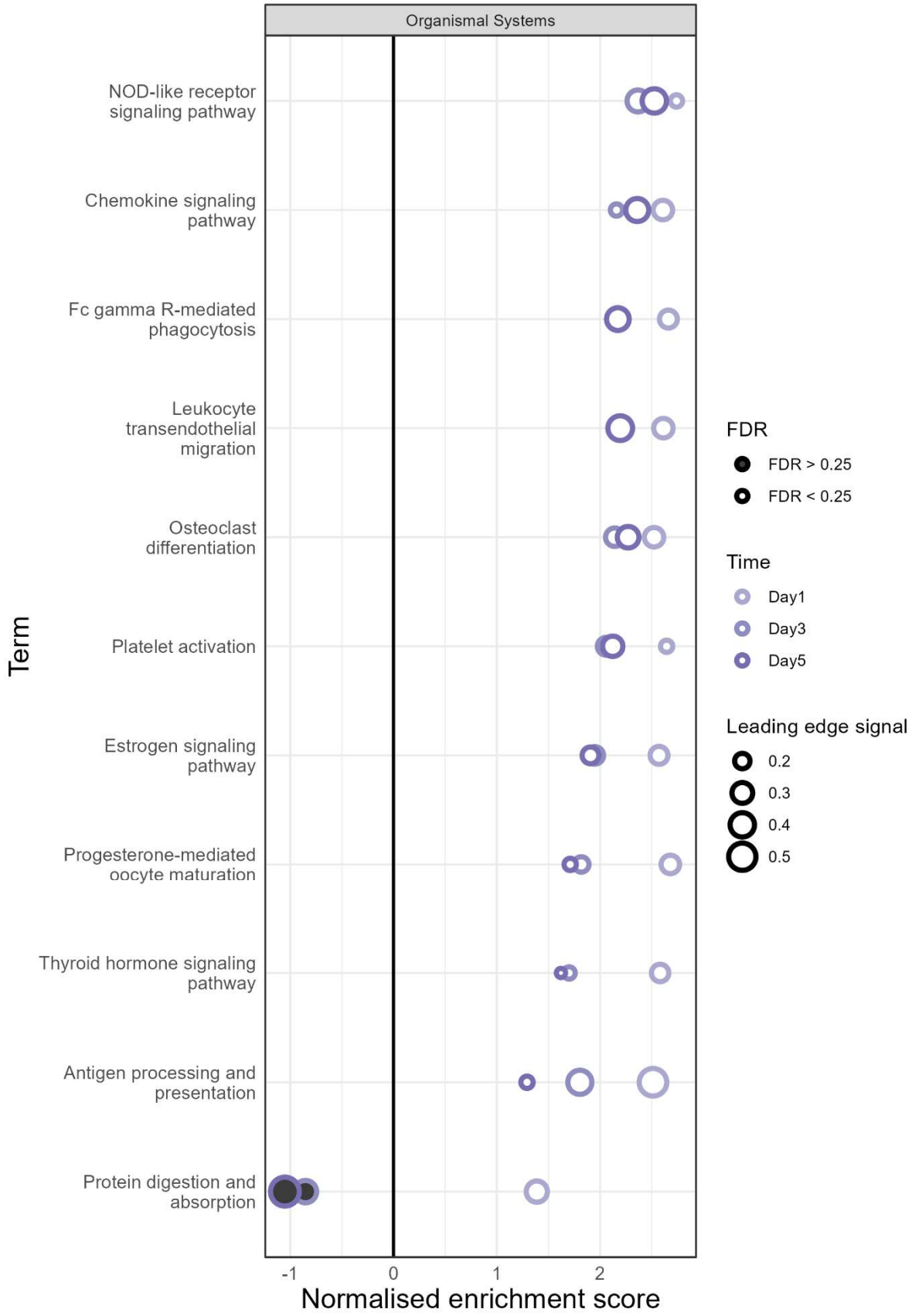


Figure 4.6 ***Gene set enrichment analysis (GSEA) of ubiquitome data from HoxB8 neutrophil differentiation (KEGG).*** Genes were pre-ranked based on corresponding modified peptide log2(fold change) compared to day 0. Stroke colour represents the time point, size indicates leading edge signal, and permutation-based FDR is indicated by fill colour. The KEGG pathway subcategory 'Organismal Systems' terms are shown.

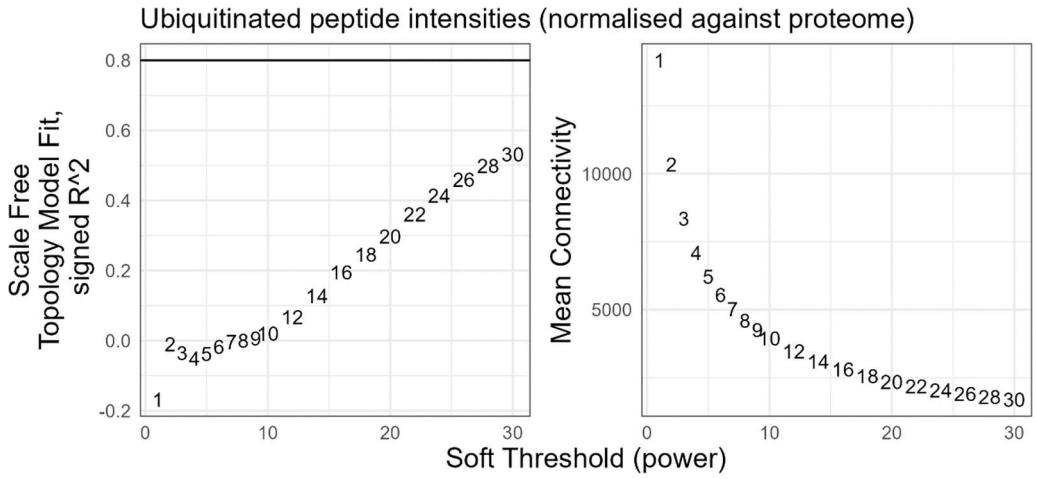


Figure 4.7 ***WCNA soft threshold 'power' parameter optimisation for ubiquitome data.*** Measures of (left) scale-free topology and (right) mean connectivity for different soft thresholding powers were calculated.

The WCNA approach to identify modules of correlated abundance profiles was explained and demonstrated in Chapter 3. Application to the ubiquitome data resulted in insufficient levels of scale-free topology and mean connectivity of the network (Figure 4.7), indicating the dataset would not be suitable for a WCNA approach.

Instead, by utilising the hierarchical clustering approach outlined in Chapter 3, differentially abundant Ub-peptide profiles were clustered into 7 groups (Figure 4.8). Functional enrichment analysis of cluster 3, corresponding to increase in ubiquitination of peptides during differentiation, corresponded to proteins associated with neutrophil effector functions, such as innate immunity, locomotion, endocytosis, and phagocytosis (Figure 4.9). Inspection of Ub protein abundance and associated ubiquitin-modified peptides, which correspond to specific types of Ub linkages, demonstrated a global overview of Ub changes that occur during HoxB8 differentiation (Figure 4.10). In particular K11 and K48 linkages, which are usually associated with protein degradation, appeared to increase in day 1 and decrease thereafter, while Ub abundance steadily increased over time. K6, K29, K33, and K63 peptides demonstrated low/no correlation with ubiquitin protein abundance. While overlapping modified peptides corresponding to each Ub linkage site generally followed similar abundance profiles, K27 peptides remarkably exhibited opposite trajectories. Although the specific type of Ub linkage and consequences of such ubiquitination for a given protein cannot be determined from this ubiquitomic approach, it is unlikely for proteins associated with neutrophil effector functions to be heavily ubiquitinated and marked for degradation, and increasing ubiquitination may be performing a functional role rather than a degradative one. Additionally, increased levels of ubiquitination may be simply a result of increased protein abundance. Incorporating information of protein abundance may capture such effects.

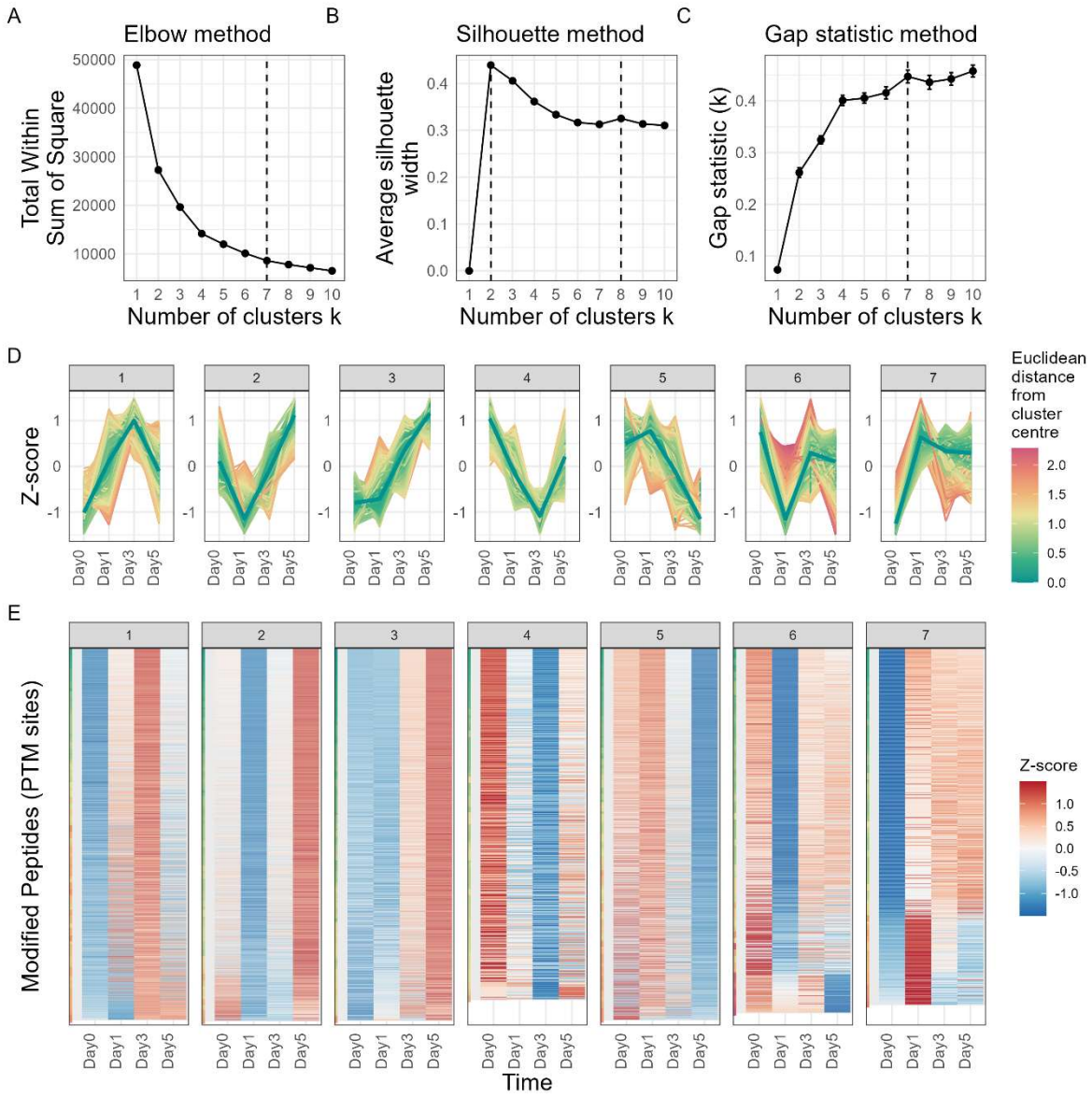


Figure 4.8 ***Hierarchical clustering approach to identify abundance profiles of significantly differentially abundant ubiquitin-modified peptides.*** Peptide values corresponding to the median peptide abundance (n = 4 biological replicates) at each time point were Z-scored across all time points. Cluster number optimisation approaches included (A) elbow, (B) silhouette, and (C) gap statistic methods. (D) Ubiquitin-modified peptide abundance profiles of clusters 1-7. Cluster centre is represented as a thick, dark-green line. (E) Heatmap of clustered modified peptides (which correspond to ubiquitination sites).

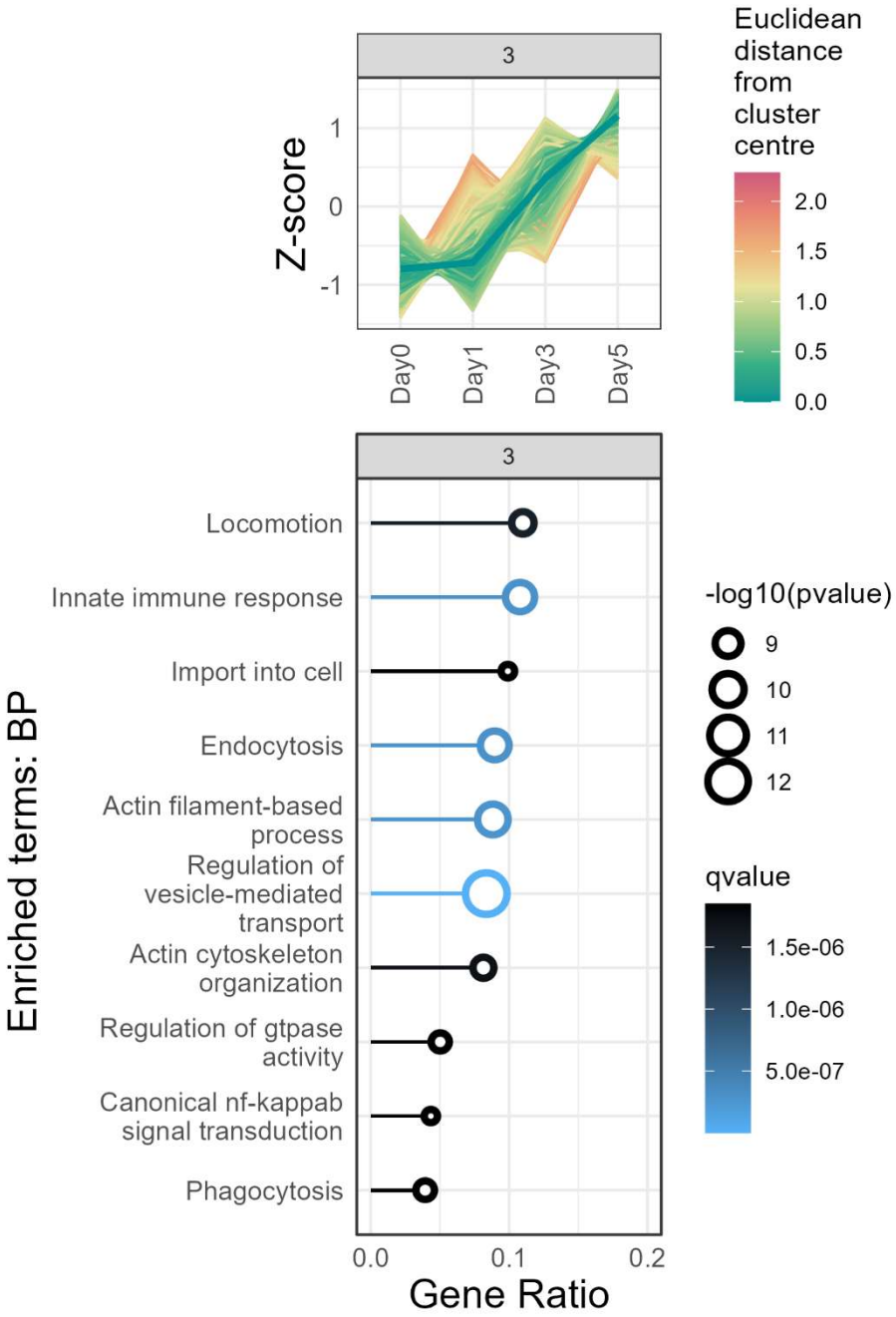


Figure 4.9 ***Functional enrichment analysis of cluster 3 ubiquitin-modified peptide profiles.*** Over-representation analysis was performed on each cluster, using the proteins identified in the ubiquitome as the background proteins. Point size and colour represent -log10(p-value) and FDR, respectively. Gene ontology (GO) biological process (BP) terms are shown for cluster 3.

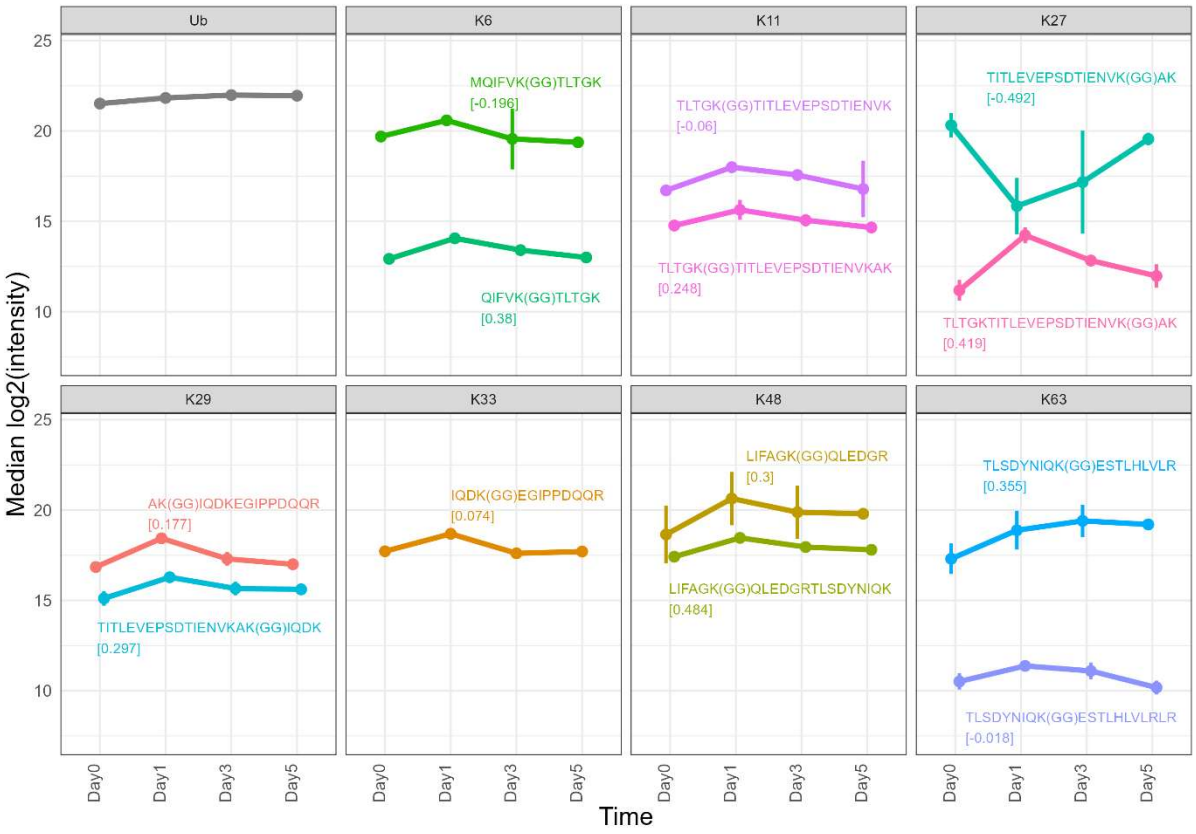


Figure 4.10 ***Abundance profiles of ubiquitin and ubiquitin-modified peptides derived from ubiquitin.*** Modified peptide sequences (GG refers to ubiquitin site) are colour coded to match corresponding abundance profile. Correlation of modified peptide abundances with ubiquitin protein abundances (Ub) are shown below the peptide sequence in square brackets. Median values are shown, and error bars represent the standard error of median (SEM).

4.2.2. Proteome-normalised ubiquitome accounts for changes in protein abundance

Due to the intrinsic link between protein abundance and levels of PTMs, ubiquitome data were normalised against corresponding total proteome data. Median intensity values for each protein group for each time point were calculated and subtracted from the modified peptide intensities (each replicate corresponding to a specific time) to achieve a 'proteome-normalised' intensity. This resulted in 26,465 modified peptides from 4,290 protein groups for further analysis.

QC of the data showed similar trends in Ub-peptide IDs with matched protein quantitation (Figure 4.11A). Minimal numbers of mutual missingness for modified peptide IDs and corresponding protein group IDs (i.e. 'absent' proteins, since median values were used for normalisation) were observed (Figure 4.11B). Data structure and groupings of biological replicates were maintained in the proteome-normalised ubiquitome data, evident from the PCCs between replicates and PCA (Figure 4.11D, E). Correlation of proteome-normalised intensities between replicates demonstrated clear distinction between biological samples, with higher correlations between replicates and lower correlations

between stages of differentiation. This is further evident from PCA, showing clear groupings of each time point. Overall PCC between modified peptide and protein quantification was low, with a coefficient of 0.29 for all data points (Figure 4.11F). Scatter plot of the intensities shows general skew towards proteins with high protein abundances and low corresponding modified peptide intensities, across all time points.

Moderated t-test with BH FDR control with the proteome-normalised ubiquitome data revealed similar trends of DAPeps, compared to analysis of standard ubiquitome data (Table 4.2). However, the changes associated with certain modified peptides were not consistent with the standard ubiquitome data analysis (Figure 4.12).

Table 4.2 ***Summary of moderated t-test outputs for proteome-normalised ubiquitome data of HoxB8 differentiation.*** *Each time point was compared to day 0.*

Time	Downregulated log2FC < -1, FDR < 0.01	Unchanged |log2FC| < 1	Upregulated log2FC > 1, FDR < 0.01
Day 1	3,827	20,095	2,543
Day 3	1,730	20,916	3,819
Day 5	2,358	18,203	5,904

While K336 and K341 of CXCR2 still demonstrated increased abundance in mature HoxB8 neutrophils, K282 showed a decrease in abundance. This may indicate K282 is associated with protein degradation, where as K336 and K341 may have functional roles other than degradation in CXCR2 activities. Many ubiquitination sites on FGR were in fact downregulated or unchanged, relative to protein abundance. K91, K136, K190, K282, K289, K299, and K411, did not show significant increase in abundance as they did in the standard ubiquitome analysis. This suggests it is important to consider changes in protein abundance to identify potential confounding factors.

Receptor tyrosine kinase KIT is a receptor that binds to SCF leading to signalling that regulates cell proliferation and survival[147]. HSCs are known to express KIT, with eventual loss of expression upon differentiation[148]. However, KIT abundance did not change during HoxB8 differentiation. Interestingly, ubiquitination of several ubiquitination sites of KIT showed statistically significant decreases in abundance, with K596/K592 and K920/K916 demonstrating the most substantial changes, during HoxB8 differentiation. Ligand-induced polyubiquitination of KIT has been observed to regulate KIT stability and function in a human myeloid cell line[149], as well as monoubiquitination through interaction with the E3 ubiquitin-protein ligase CBLI[150]. This may be indicative of KIT maintenance during neutrophil differentiation where specific ubiquitination sites are more crucial in stabilisation of the protein.

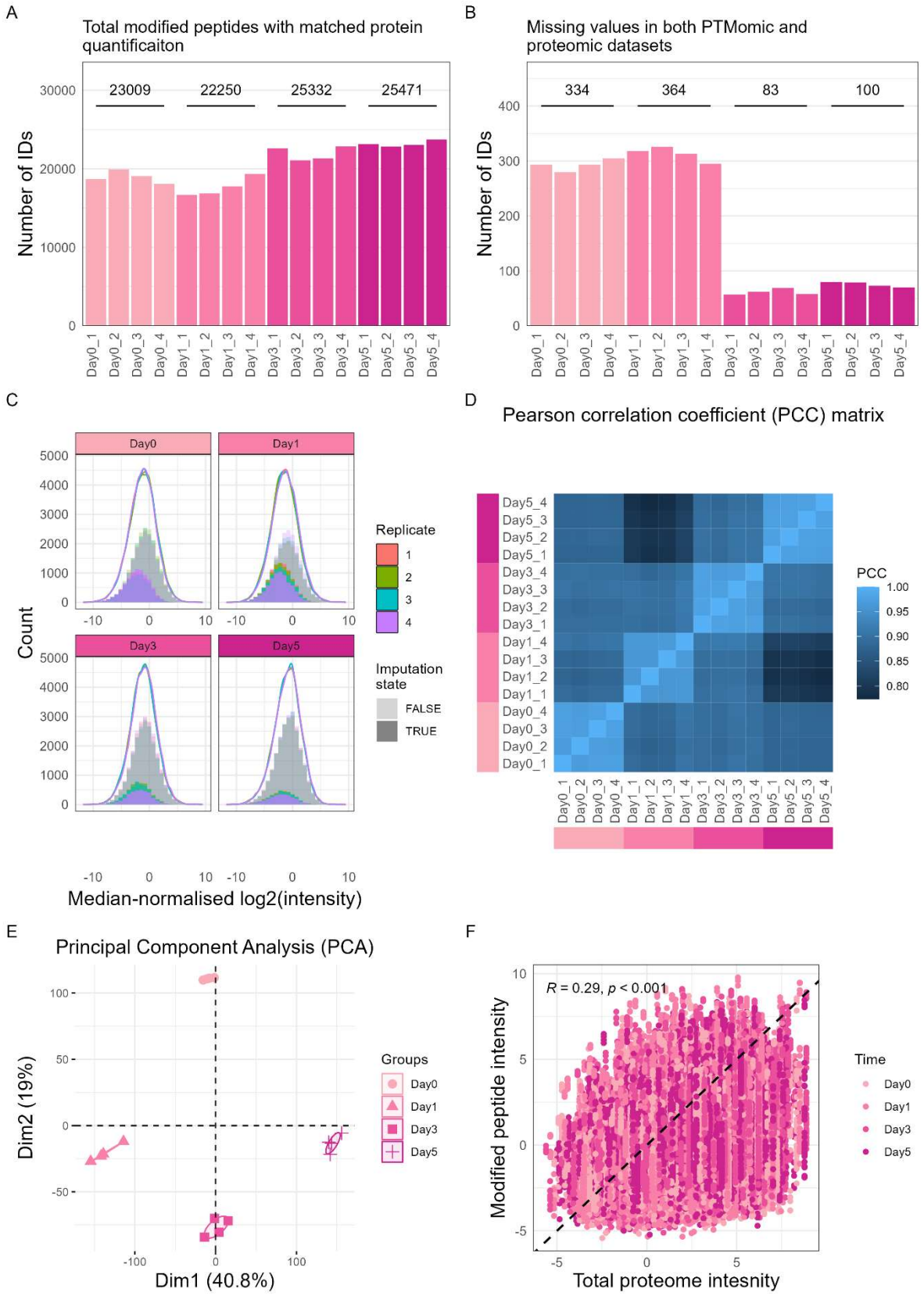


Figure 4.11 ***QC of proteome-normalised ubiquitome data of HoxB8 neutrophils during differentiation.*** (A) Number of ubiquitin-modified peptide IDs with corresponding protein identification from proteome dataset, per biological replicate. Total IDs per time point are shown above each group (n = 4 biological replicates).. (B) Number of missing values common to peptide and corresponding protein IDs, per biological replicate. (C) Distribution of median-normalised, log2-transformed peptide intensities with imputed missing values (processed data). Imputed values are shown as more opaque sections of their corresponding histogram. (D) Pearson correlation coefficient between biological replicates and time points of the processed data. (E) Principal component analysis of the processed data. (F) Comparison of ubiquitin-modified peptide intensities and corresponding protein intensities. Each point represents the median intensities of a specific modified peptide-protein group pair, for a specific time point.

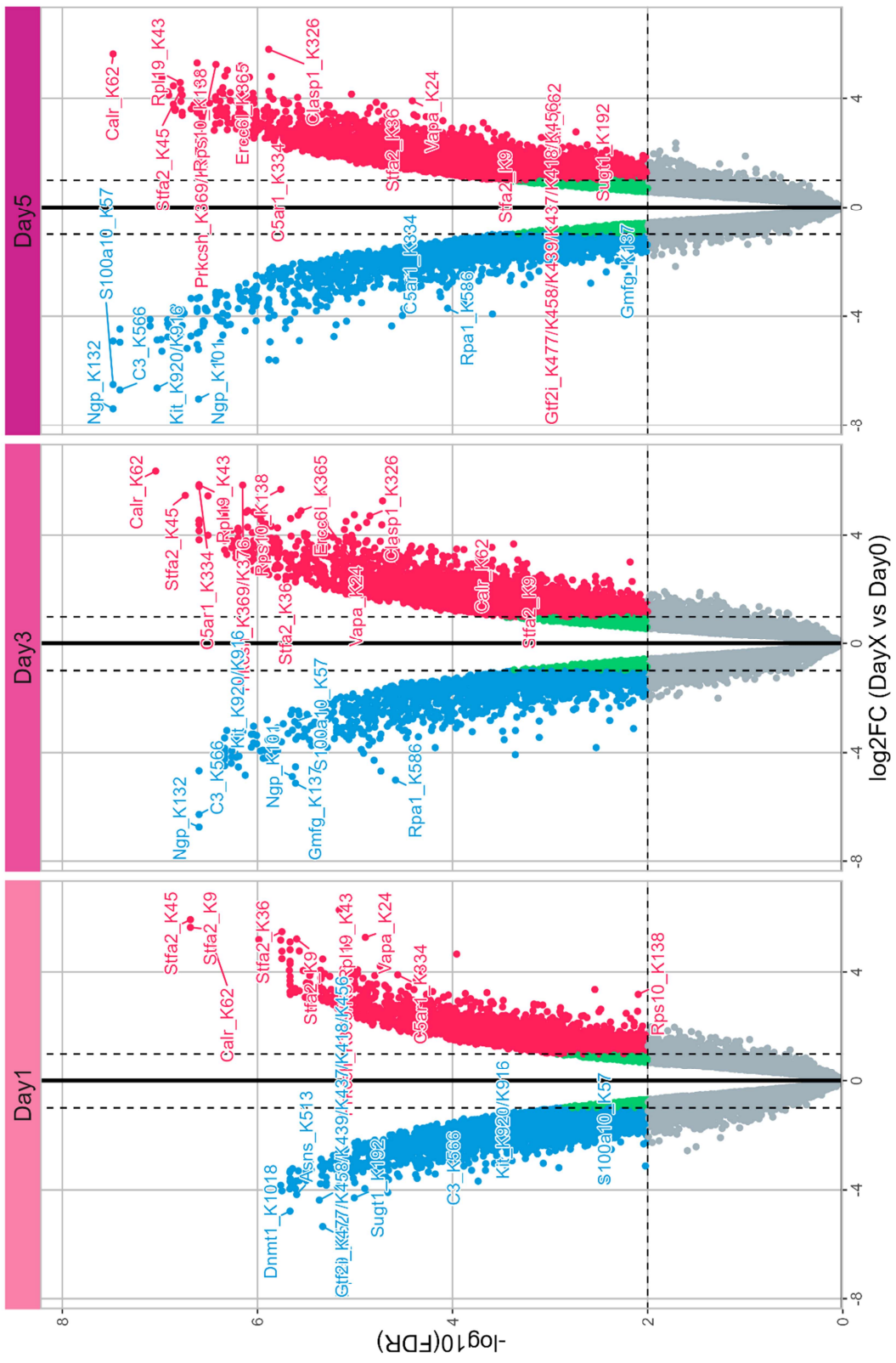


Figure 4.12 ***Volcano plots of differentially abundant modified peptides (DAMPs) in the proteome-normalised ubiquitome of different stages of Hox8 neutrophil differentiation and day 0.*** Ubiquitin-modified peptide intensity values were normalised against median proteome values of corresponding proteins. Moderated t-test was used to compare groups (n = 4 biological replicates), and the Benjamini-Hochberg procedure was used to control false discovery rate (FDR). Peptides with FDR < 0.01 and |log2(fold change)| > 1 were considered statistically significant and differentially abundant. Upregulated and downregulated peptides are shown in red and blue, respectively. Green represents peptides with FDR < 0.01 and |log2(fold change)| < 1.

Complement proteins and receptors are important in innate immune defence. Neutrophils store complement proteins intracellularly and express complement receptors on their CSM as a way to not only activate other immune cell types, but also other neutrophils via positive feedback amplification[151,152]. Complement factor 3 (C3) showed increased protein intensity during differentiation and decreased intensities of K398, K566, and K733 peptides during differentiation, potentially indicative of protein stabilisation. The complement component 5a (C5a) receptor 1 (C5aR1) showed increase protein expression during HoxB8 differentiation. This was accompanied by an increase in modified peptide intensities corresponding to K334, which can be found on the intracellular side of the cell membrane. This may signify ubiquitination does not contribute to degradation and may instead participate in other biological processes such as signalling. It is important to note that 2 peptides corresponding to this site were detected, with one peptide showing much lower intensity values, and so, upon proteome-normalisation, surprisingly showed downregulation in mature HoxB8 cells.

GSEA of genes corresponding to the Ub-peptides, ranked based on their log_2FC, revealed negative NES of GOBP terms associated with neutrophil effector functions such as leukocyte migration, inflammatory response, and phagocytosis, in mature HoxB8 neutrophils (Figure 4.13). This suggests proteins that are pivotal for neutrophil activities may be stabilised through decreased protein ubiquitination, relative to their protein abundance, preventing their degradation and maintaining their specialised functions for use in mature HoxB8 neutrophils. In contrast, GSEA of KEGG 'Organismal Systems' still showed positive NES (with slight decrease in score during differentiation) for terms such as chemokine signalling pathway and leukocyte transendothelial migration (Figure 4.14), suggesting overall pathways consist of proteins that are, in general, ubiquitinated. This potentially acts as a regulatory mechanism to prevent inappropriate egress from the differentiation niche or plays non-degradative roles in protein regulation.

Figure 4.13 ***Gene set enrichment analysis (GSEA) of proteome-normalised ubiquitome data from HoxB8 neutrophil differentiation (GOBP).*** Genes were pre-ranked based on corresponding ubiquitin-modified peptide log2(fold change) compared to day 0. Stroke colour represents the time point, size indicates leading edge signal, and permutation-based FDR is indicated by fill colour. Gene ontology (GO) biological process (BP) terms are shown.

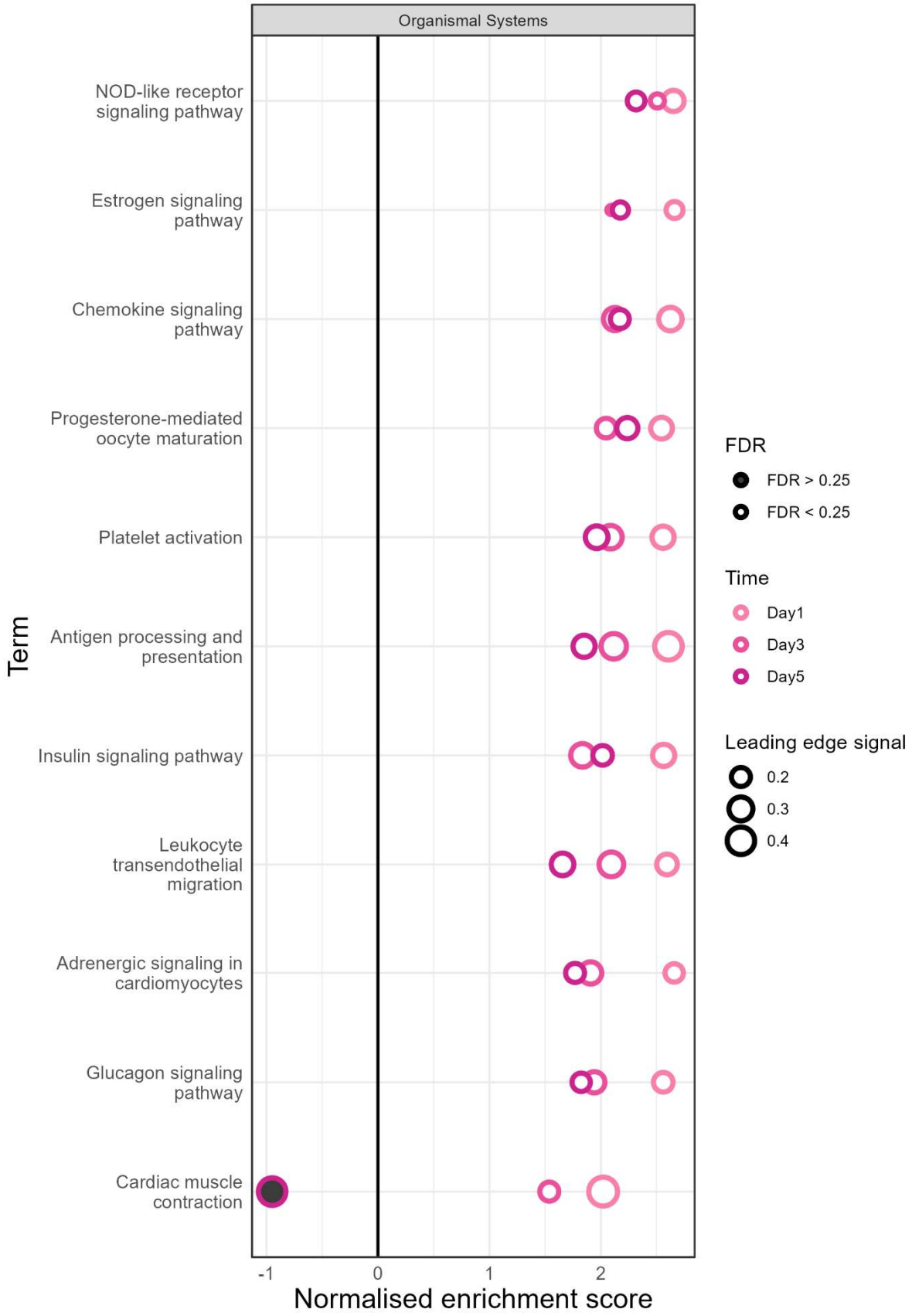


Figure 4.14 ***Gene set enrichment analysis (GSEA) of proteome-normalised ubiquitome data from HoxB8 neutrophil differentiation (KEGG, Organismal Systems).*** Genes were pre-ranked based on corresponding ubiquitin-modified peptide log2(fold change) compared to day 0. Stroke colour represents the time point, size indicates leading edge signal, and permutation-based FDR is indicated by fill colour. The KEGG pathway subcategory 'Organismal Systems' terms are shown.

Employing the hierarchical clustering approach outlined in chapter 3 on the proteome-normalised ubiquitome data, including removal of modified peptides with no significant changes in abundance during differentiation, identified 9 unique clusters representing temporally-resolved abundance profiles (Figure 4.15). Functional enrichment analysis identified gene sets that were overrepresented in each cluster (Figure 4.16, Figure 4.17, Figure 4.18). Profiles associated with overall decrease in relative protein ubiquitination (clusters 2 and 7) were associated with cell migration, and increasing profiles (clusters 1 and 5) were associated with DNA metabolism, RNA processing, and ribosomal processes. Profiles with intermediate decreases (during day 1 and 3) were associated with cell cycle regulation, DNA repair, and protein phosphorylation (clusters 5 and 6). Conversely, profiles with intermediate increases (clusters 3 and 4) showed enrichment for cytoskeleton organisation, protein transport, membrane organisation, and endoplasmic reticulum (ER)- and Golgi-associated terms. GOMF terms for clusters 4, 5, 6, and 8 suggest an interplay between ubiquitination and phosphorylation machinery that may contribute to regulation of neutrophil differentiation, potentially modulating signalling pathways in a time-specific manner to elicit appropriate differentiation programmes for neutrophil maturation. Additional ORA of clustered proteome-normalised ubiquitome profiles can be found in Supplementary Figure 23.

ESI analysis of clustered profiles identified E3 ligases and DUBs that may regulate protein ubiquitination. Majority of enzyme-substrates failed to reach low FDR, again, due to limited data availability of ESI. However, cluster 4 showed enrichment of reported substrates of the deubiquitinase YOD1 that are validated in the literature. Ub-peptides corresponding to the E3 ubiquitin-protein ligase NEDD4, the Valosin-containing protein (VCP), and the HECT-type E3 ubiquitin ligase ITCH demonstrated abundance profiles that peak on day 3 of HoxB8 differentiation. NEDD4 and ITCH are both E3 ligases and so control cellular ubiquitination events, and VCP is involved in protein quality control and response to protein misfolding. Additionally, known ubiquitin specific peptidase 8 (USP8) substrates were identified in the majority of clusters, potentially indicating a more global role in regulating cellular ubiquitination status. YOD1 protein was not detected in the proteome data, potentially due to lack of proteotypic peptides allowing for unique quantitation. Instead, ubiquitome data identified YOD1-derived Ub-peptides corresponding to K252. These peptides showed slight increase in abundance during day 1 of HoxB8 differentiation with subsequently slight decreases in intensities. The profile of these YOD1 ubiquitinated profiles appears to follow a similar, leading trend to those of the substrate Ub-peptides. Additionally, transcriptomic data of the HoxB8 model[110] shows YOD1 transcript levels increasing during neutrophil differentiation.

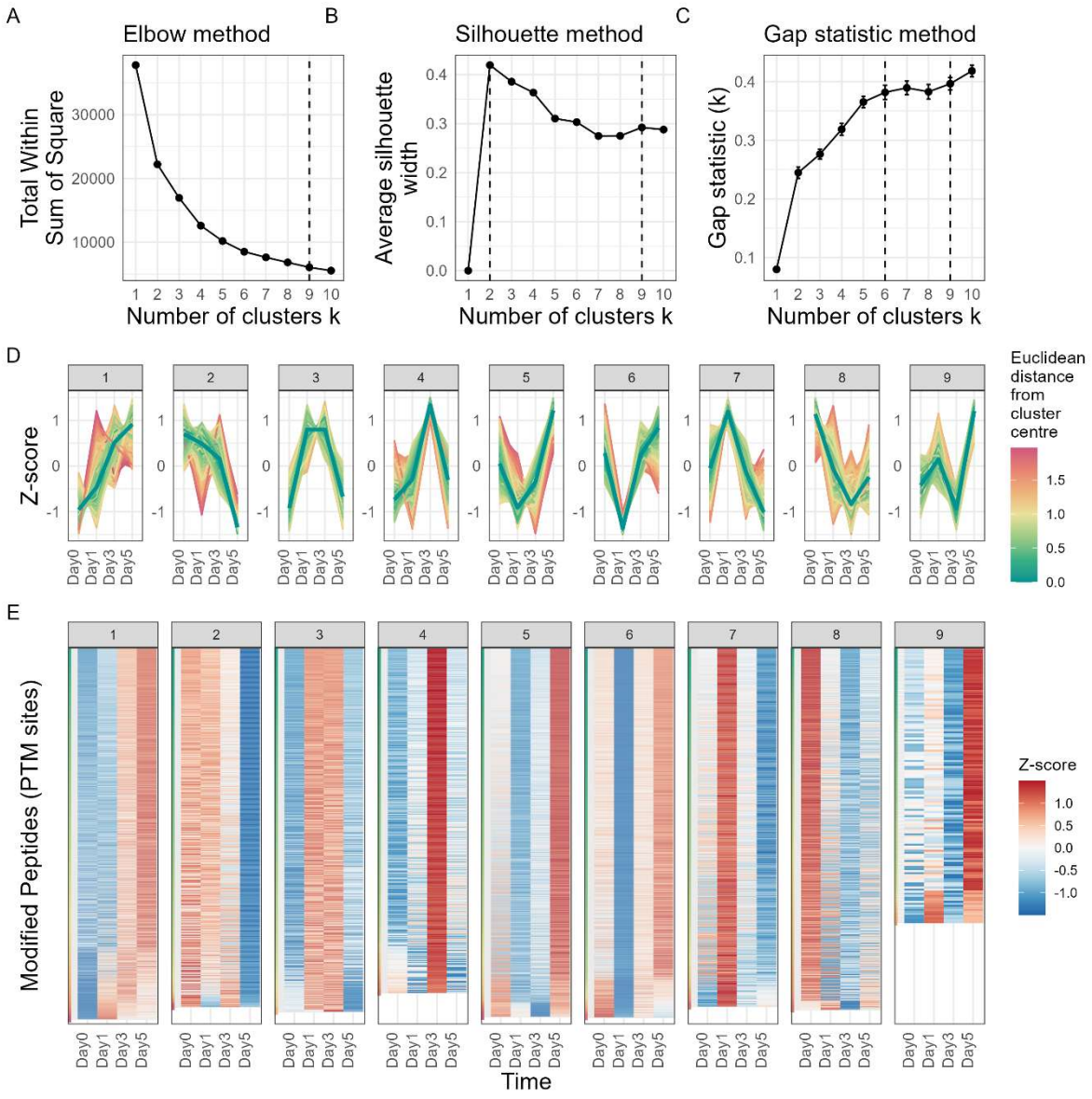


Figure 4.15 ***Hierarchical clustering approach to identify abundance profiles of significantly differentially abundant, proteome-normalised ubiquitin-modified peptides.*** Peptide values corresponding to the median proteome-normalised peptide abundance (n = 4 biological replicates) at each time point were Z-scored across all time points. Cluster number optimisation approaches included (A) elbow, (B) silhouette, and (C) gap statistic methods. (D) Proteome-normalised ubiquitin-modified peptide abundance profiles of clusters 1-9. Cluster centre is represented as a thick, dark-green line. (E) Heatmap of clustered modified peptides (which correspond to ubiquitination sites).

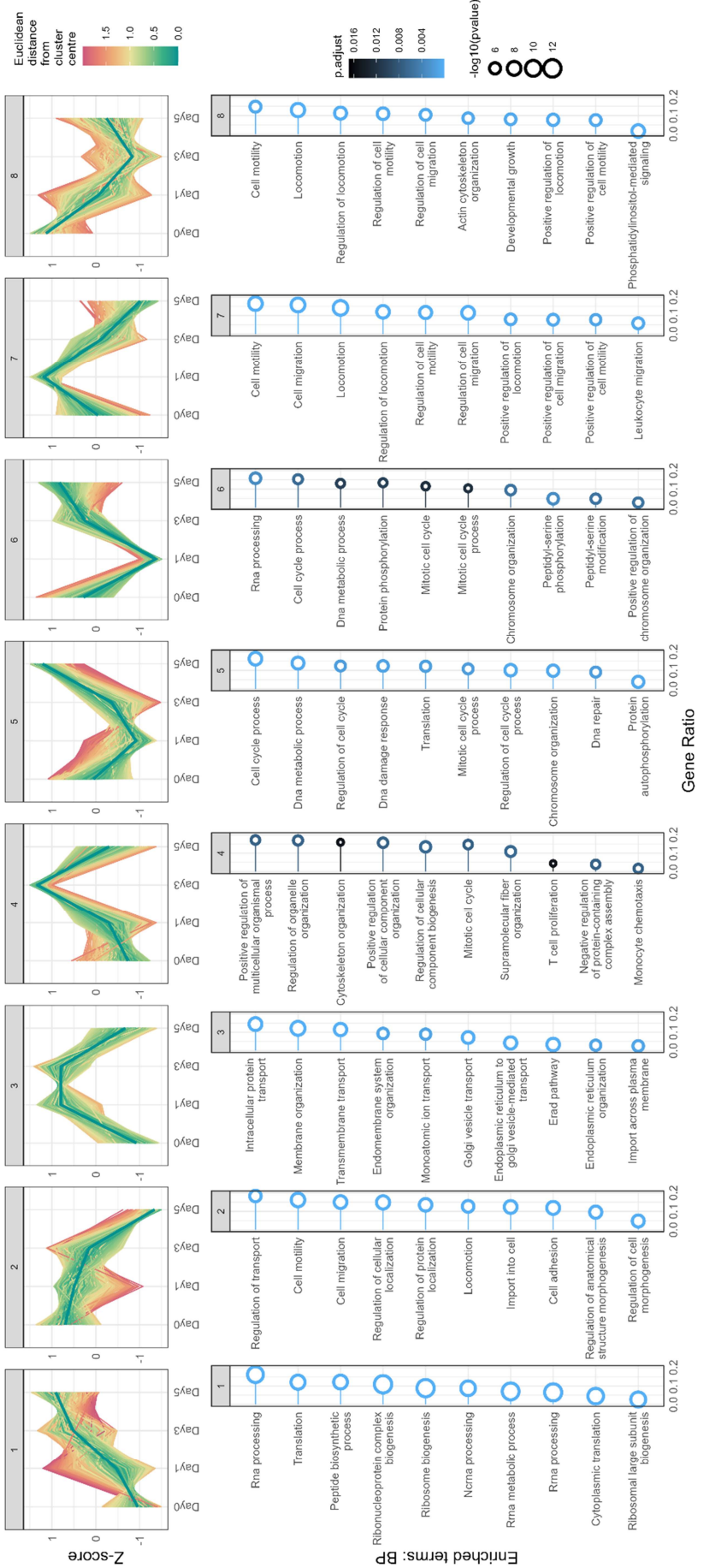


Figure 4.16 ***Functional enrichment analysis of clustered proteome-normalised ubiquitin-modified peptide profiles that are differentially abundant during HoxB8 neutrophil differentiation (GOBP).*** Over-representation analysis was performed on each cluster, using the proteins identified in the ubiquitome and total proteome as the background proteins. Point size and colour represent -log10(p-value) and FDR, respectively. Gene ontology (GO) biological process (BP) terms are shown.

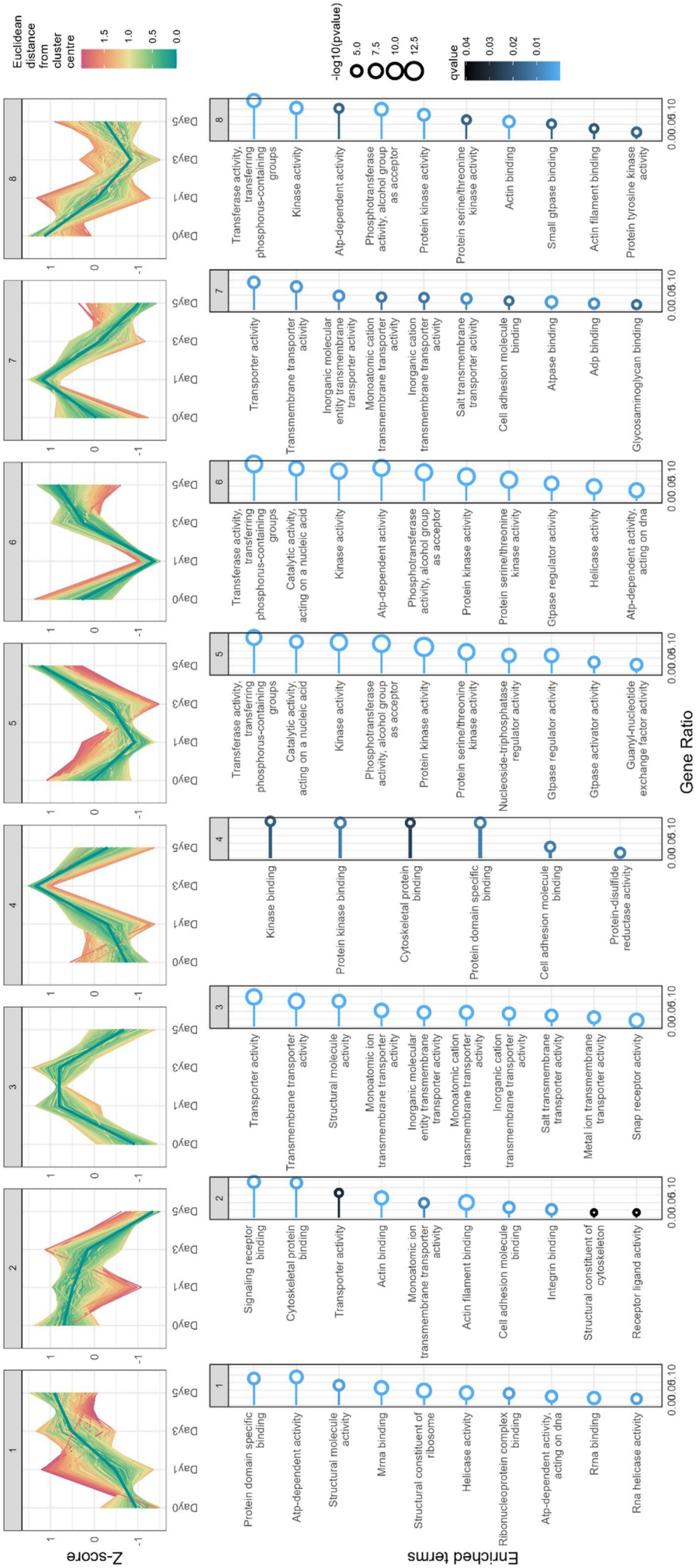


Figure 4.17 ***Functional enrichment analysis of clustered proteome-normalised ubiquitin-modified peptide profiles that are differentially abundant during HoxB8 neutrophil differentiation (GOMF).*** Over-representation analysis was performed on each cluster, using the proteins identified in the ubiquitome and total proteome as the background proteins. Point size and colour represent -log10(p-value) and FDR, respectively. Gene ontology (GO) molecular functions (MF) terms are shown.

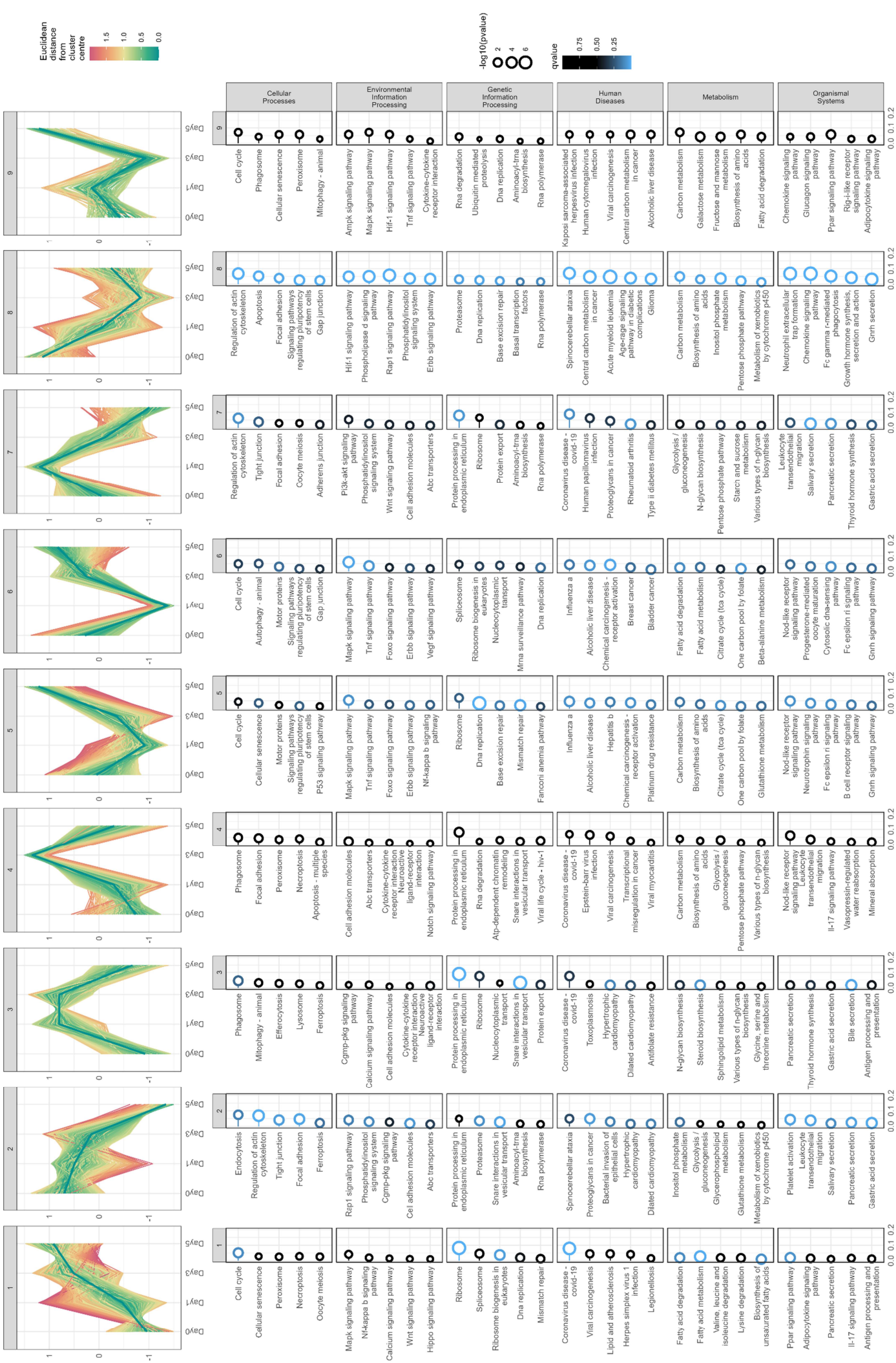


Figure 4.18 ***Functional enrichment analysis of clustered proteome-normalised ubiquitin-modified peptide profiles that are differentially abundant during HoxB8 neutrophil differentiation (KEGG).*** Over-representation analysis was performed on each cluster, using the proteins identified in the ubiquitome and total proteome as the background proteins. Point size and colour represent -log10(p-value) and FDR, respectively. KEGG pathways and subcategories are shown.

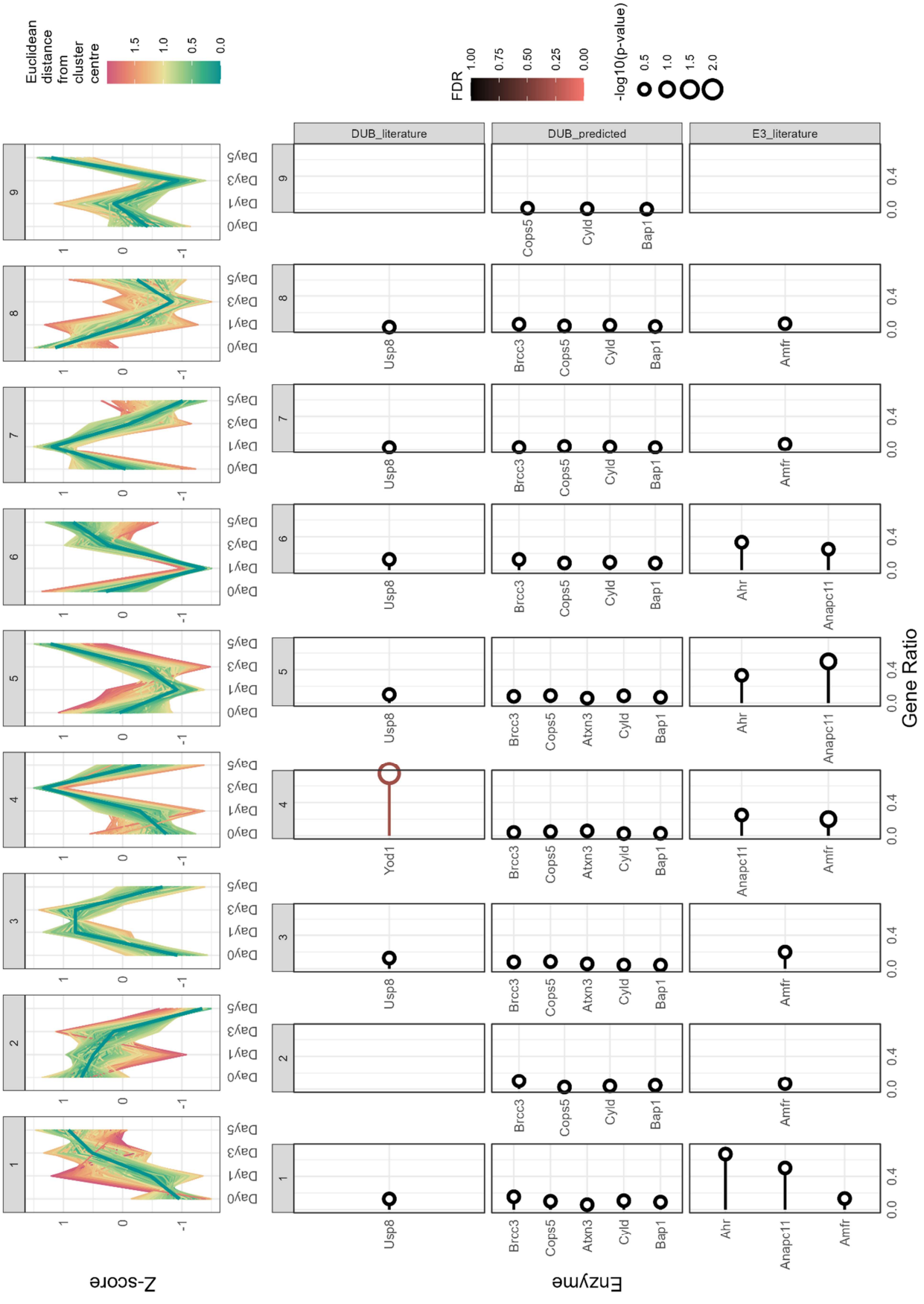


Figure 4.19 ***Enzyme-substrate interaction analysis identifies potential DUB and E3 ligase enzymes that may regulate proteins within ubiquitin-modified peptide profile clusters.*** Clusters of proteins identified using the hierarchical approach were analysed. Literature/predicted DUB- and literature E3 ligase-substrate interactions are shown for clusters 1-9. Hypergeometric testing with BH procedure for FDR control were used to determine statistical significance.

Predictions regarding the E3 ligases included aryl hydrocarbon receptor (AHR), a ligand-dependent transcription factor and E3 ligase, and autocrine motility factor receptor (AMFR), an ER-resident E3 ligase. These findings pose new hypotheses regarding the role of the ubiquitin system in neutrophil differentiation.

4.2.3. Phosphoproteome data reveals differing phosphopeptide identifications and quantitation depending on method

Phosphoproteomic analysis of HoxB8 differentiation involved enrichment of phosphopeptides from trypsin-digested whole cell lysates and analysis using MS in both DDA and DIA modes. Raw DDA data was processed in FragPipe v23.1 using the default workflow (phospho-DDA), and raw DIA data was processed using either DIA-NN v2.1 or FragPipe v23.1, the latter of which used the diaTracer implementation (to assess boost in IDs) and DIA-NN for quantitation.

Comparing outputs from the raw MS data processing, DDA and diaTracer data showed comparable phosphopeptide IDs, with 16,660 and 18,377 phosphopeptides, respectively, far exceeding DIA-NN, which identified 10,980 (Figure 4.20A). Looking at the overlapping IDs, around 16% of all IDs were common to all three datasets. DDA and diaTracer shared the most IDs (roughly 31% of IDs). However, examining the quantitation of phosphopeptides from each replicate common to each comparison, DDA and DIA-NN quantitation showed higher correlation for each replicate (Figure 4.20C), comparable to the correlations in quantitation between DIA-NN and diaTracer (which both use DIA-NN for quantitation) (Figure 4.20B). DDA and diaTracer showed the lowest PCCs across the board (Figure 4.20D). Due to established MS protocols within the lab, the DDA phosphoproteomic data was taken forward for analysis.

4.2.4. Proteome-normalised phosphoproteome data reveals altered phosphorylation patterns during neutrophil differentiation

DDA data was processed similarly to the ubiquitome data, using intensity values for phosphopeptides, filtered for localisation probabilities greater than 0.75. Intensities were $\log_2$-transformed, median-normalised, and conditional imputation was performed for absent and missing data. Phosphoproteome data were filtered to include phosphopeptides with valid intensity values in at least 3 out of 4 replicates, and CV adjustment removed phosphopeptides in the highest CV decile. This resulted in 9,742 phosphopeptides, 7,862 of which were matched to 2,092 proteins, with a generally consistent number of phosphopeptide IDs across all time points (Figure 4.21A). Less than 150 phosphopeptides containing missing values corresponded to an absent protein value for any given time point. However, due to the smaller overall number of phosphopeptide IDs, missing values common to both phosphoproteome and proteome data represented a larger proportion in days 0 and

1 (Figure 4.21B), compared to that of the proteome-normalised ubiquitome data (Figure 4.11B). Nevertheless, QC of proteome-normalised phosphopeptide data demonstrated a clear distinction between the differentiation stages, evident from PCA and PCCs between biological replicates (Figure 4.21D, E). Similar to ubiquitome data, phosphopeptide intensities showed low correlation with protein intensity across all data points (PCC ≈ 0.25) (Figure 4.21F).

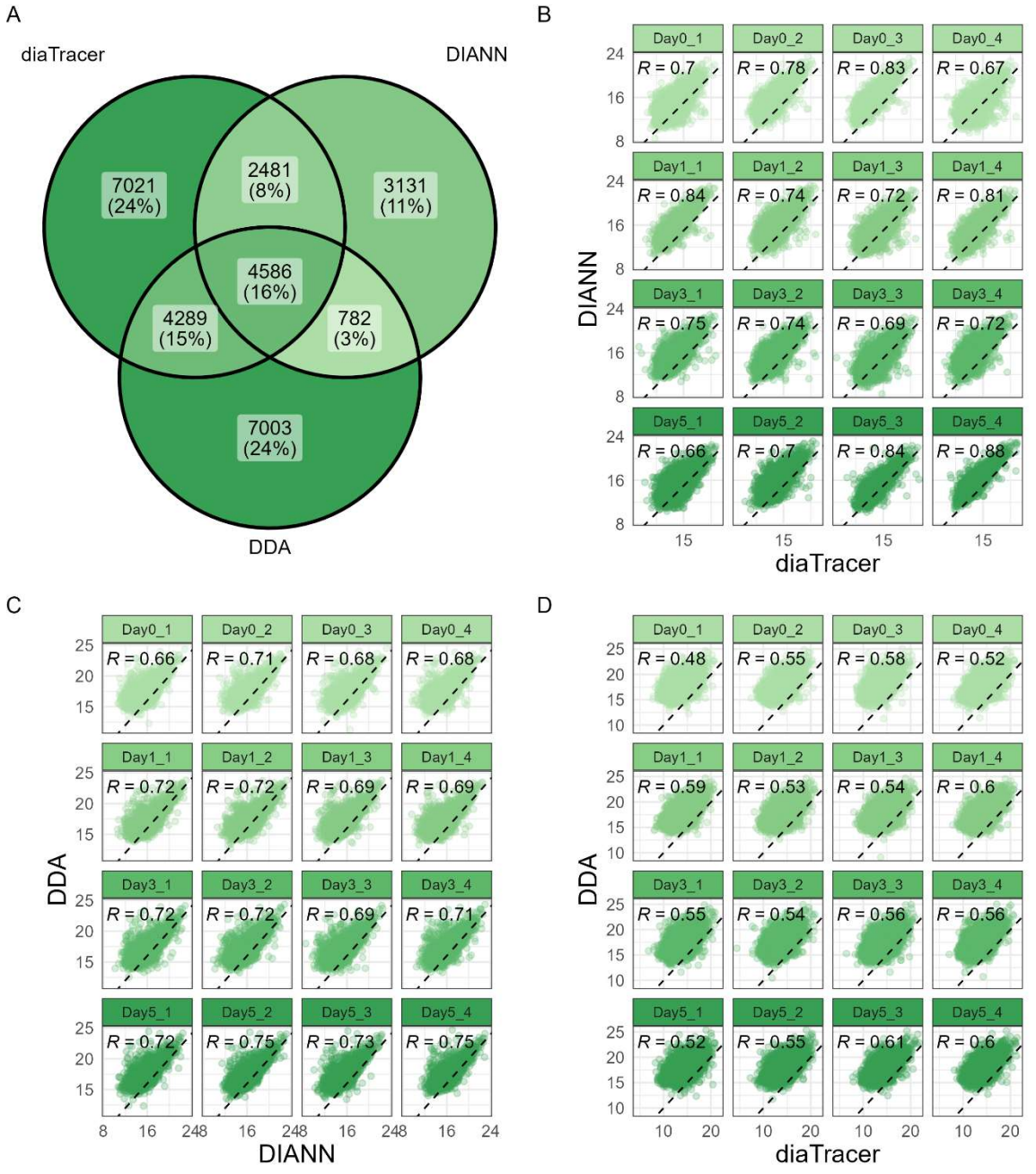


Figure 4.20 ***Comparison of raw MS data processing approach for DDA and DIA phosphoproteome data.*** Raw DDA and DIA data were processed using FragPipe 23.1, with DIA data using the diaTracer implementation. DIA data was also processed using DIA-NN v2.1. (A) Phosphopeptide IDs from each approach were compared for overlap. Phosphopeptide quantitation was compared between (B) DIA-NN and diaTracer, (C) DIA-NN and DDA, and (D) diaTracer and DDA datasets. Pearson correlation coefficients (PCCs) for each biological replicate are shown in corresponding panel (R = PCC). Dotted line represents 1:1 quantitation.

Volcano plotting of moderated t-test output with BH FDR control for each phosphopeptide, comparing each day to day 0, revealed increasing numbers of DAPeps as HoxB8 neutrophils differentiated (Figure 4.22, Table 4.3).

Table 4.3 ***Summary of moderated t-test outputs for proteome-normalised phosphoproteome data of HoxB8 differentiation.*** *Each time point was compared to day 0.*

Time	Downregulated log2FC < -1, FDR < 0.01	Unchanged |log2FC| < 1	Upregulated log2FC > 1, FDR < 0.01
Day 1	458	6,867	585
Day 3	744	6,394	772
Day 5	921	6,174	815

Y-box binding protein 1 (YBX1) is a transcription factor that participates in transcription, translation, DNA damage repair, and cell signalling, through its protein-protein interactions and ability to bind both DNA and RNA. Accordingly, YBX1 is implicated in various disease contexts, such as cancer[153]. Several phosphosites were identified for YBX1: T23, S100, S163, S165, S174, S311, S312, showed upregulated phosphorylation; S2, T7, S19, S27, and S172 showed no change; and S3 was downregulated. YBX1 phosphorylation modulates YBX1 activity, with S30, S34, S102, S165, S174, and S176 being reported in the literature[154-158]. YBX1 has been shown to regulate expression of C-X-C motif chemokine ligand 1 (CXCL1)[159] (known to be secreted by neutrophils) and C-C motif ligand 5 (CCL5)[160], the latter of which involves protein phosphorylation to activate YBX1 function.

Negative elongation factor complex member B (NELFB) is a subunit of the NELF complex, which negatively regulates transcription elongation of RNA polymerase II, affecting cell proliferation and differentiation. NELFA is known to be phosphorylated by MAPK and dephosphorylated by PP2A, which affects NELF complex association with paused polymerases, modulating transcript elongation[161]. The role of NELFB phosphorylation remains unclear. The HoxB8 model showed downregulation of NELFB phosphorylation at sites S608 and T610, which may play a role in transcriptional regulation.

The SON protein is involved in regulating pre-mRNA processing and splicing. Its depletion by RNAi resulted in cell cycle arrest in metaphase, indicating its regulatory function is vital for cellular homeostasis and proliferation[162]. The phosphoproteome data revealed SON is a highly phosphorylated protein, with 22 phosphosites identified. However, only 2 of these sites, S1998 and T2000, were upregulated in mature HoxB8 neutrophils, relative to SON protein abundance.

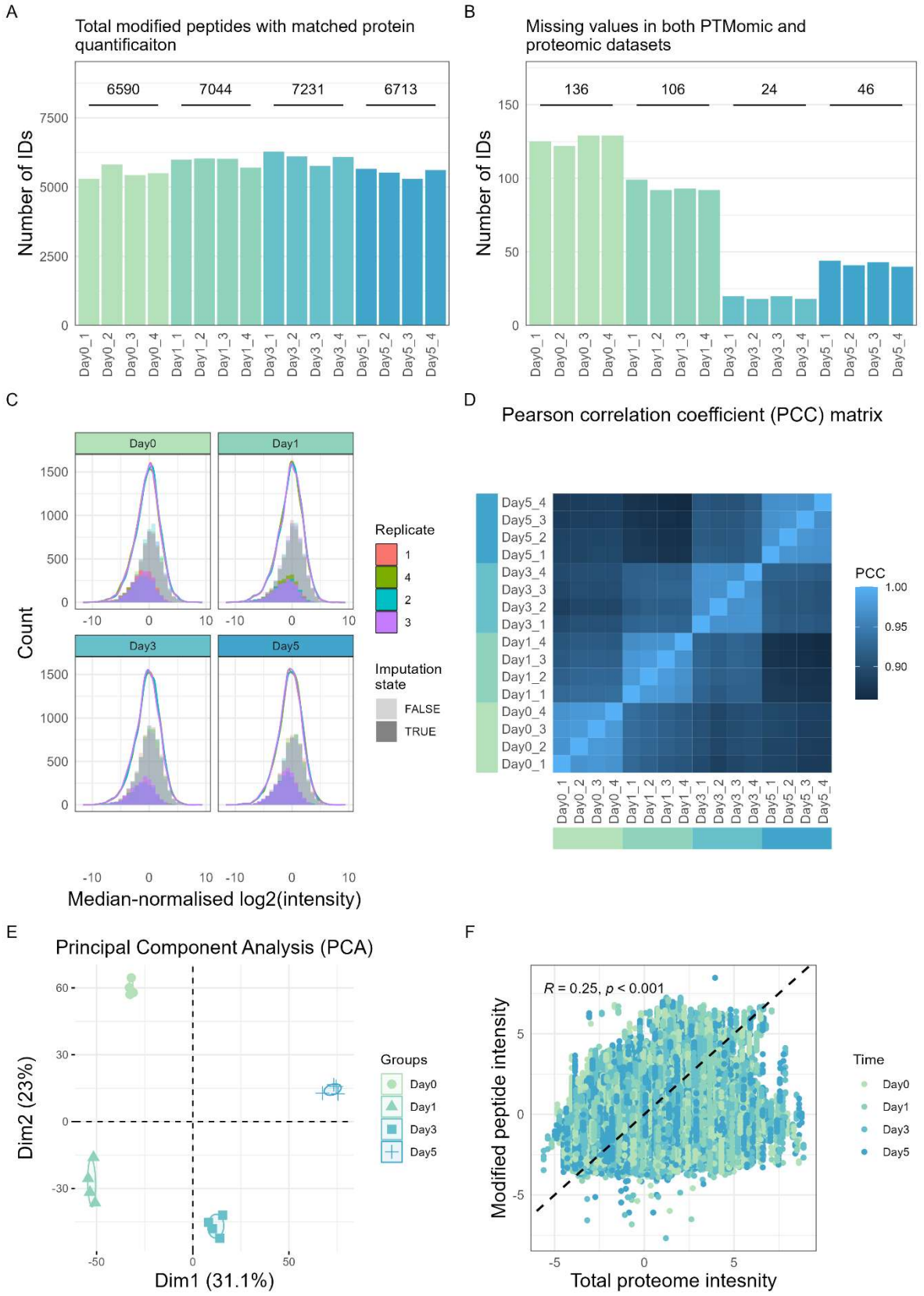


Figure 4.21 ***QC of proteome-normalised phosphoproteome data of HoxB8 neutrophils during differentiation.*** (A) Number of phosphopeptides with corresponding protein identification from proteome dataset, per biological replicate. Total IDs per time point are shown above each group (n = 4 biological replicates). (B) Number of missing values common to phosphopeptide and corresponding protein IDs, per biological replicate. (C) Distribution of median-normalised, log2-transformed phosphopeptide intensities with imputed missing values (processed data). Imputed values are shown as more opaque sections of their corresponding histogram. (D) Pearson correlation coefficient between biological replicates and time points of the processed data. (E) Principal component analysis of the processed data. (F) Comparison of phosphopeptide intensities and corresponding protein intensities. Each point represents the median intensity of a specific modified peptide-protein group pair, for a specific time point.

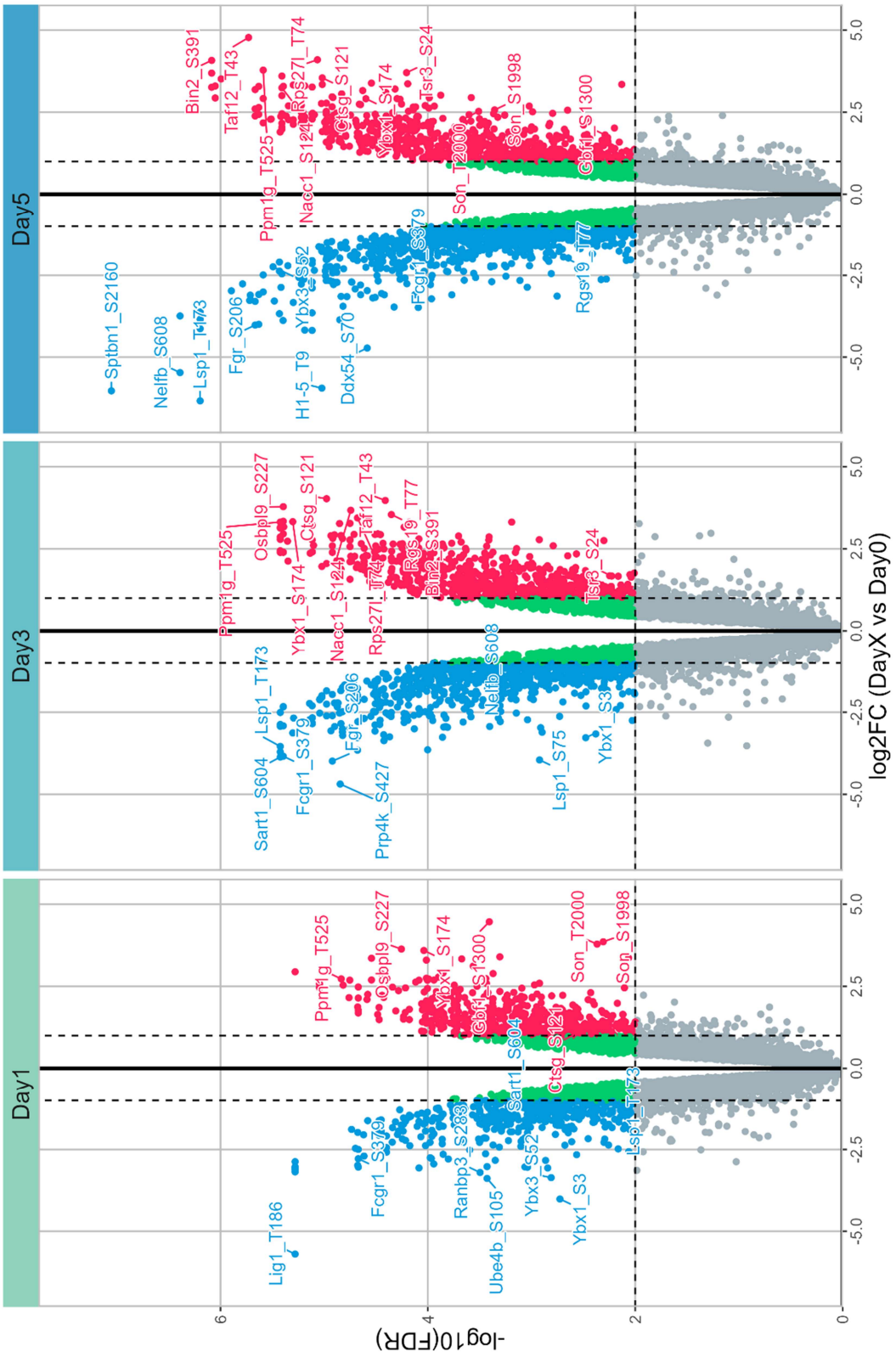


Figure 4.22 ***Volcano plots of differentially abundant phosphopeptides identified in the proteome-normalised phosphoproteome of different stages of HoxB8 neutrophil differentiation and day 0.*** Phosphopeptide intensity values were normalised against median proteome values of corresponding proteins. Moderated t-test was used to compare groups (n = 4 biological replicates), and the Benjamini-Hochberg procedure was used to control false discovery rate (FDR). Sites with FDR < 0.01 and |log2(fold change)| > 1 were considered statistically significant and differentially abundant. Upregulated and downregulated peptides are shown in red and blue, respectively. Green represents peptides with FDR < 0.01 and |log2(fold change)| < 1.

Lymphocyte-specific protein 1 (LSP1) is an actin-binding protein that participates in cytoskeletal reorganisation, with several phosphorylation sites that have been previously reported[163]. Phosphorylation of LSP1 has been shown to be induced upon neutrophil activation[164], which may indicate why the majority of LSP1 phosphosites showed no change or downregulation, with only S177 showing upregulation, during HoxB8 differentiation. LSP1 has also been demonstrated to participate in Fcγ receptor-mediated phagocytosis in macrophages[165].

Bridging integrator 2 (BIN2) is an adapter protein that is highly expressed in leukocytes, affecting podosome formation, motility, and phagocytosis in leukocytes through interactions with membranes and cytoskeletal components[166]. Little is known about its role in neutrophils. BIN2 and phosphorylated BIN2 have been shown to contribute to ovarian cancer progression and affect the phosphorylation of other proteins[167]. S391 phosphorylation is highly upregulated in day-5 HoxB8 neutrophils, along with S420 phosphorylation showing increased abundance, while other BIN2 phosphopeptides were downregulated or unchanged.

GSEA of genes, ranked based on corresponding phosphopeptide log2FC, showed positive NES for GOBP terms associated with RNA metabolism and processing, and negative NES for terms related to phagocytosis and endocytosis (vesicle fusion and endocytic recycling), in mature HoxB8 neutrophils (Figure 4.23). Additionally, KEGG 'Genetic Information Processing' terms showed positive enrichment for protein processing, Ub-mediated proteolysis, as well as DNA replication, chromatin remodelling, and spliceosome (Figure 4.24). These gene sets suggest phosphorylation plays a role in regulating pathways associated with gene expression and protein translation, potentially activating enzymes within such pathways to bring about mature neutrophil phenotypes as professional phagocytes. Many signalling pathways are known to involve phosphorylation for signal transduction, and GSEA highlighted the MAPK, mTOR, Ras, and Notch signalling pathways, among others, as significantly enriched in the positive direction (Figure 4.25). Notch[168], MAPK[103,169], and Ras[170] signalling pathways have been implicated in neutrophil differentiation, while mTOR is associated with neutrophil effector functions such as NETosis[96]. Additional GSEA of proteome-normalised phosphoproteome data can be found in Supplementary Figures 24-28.

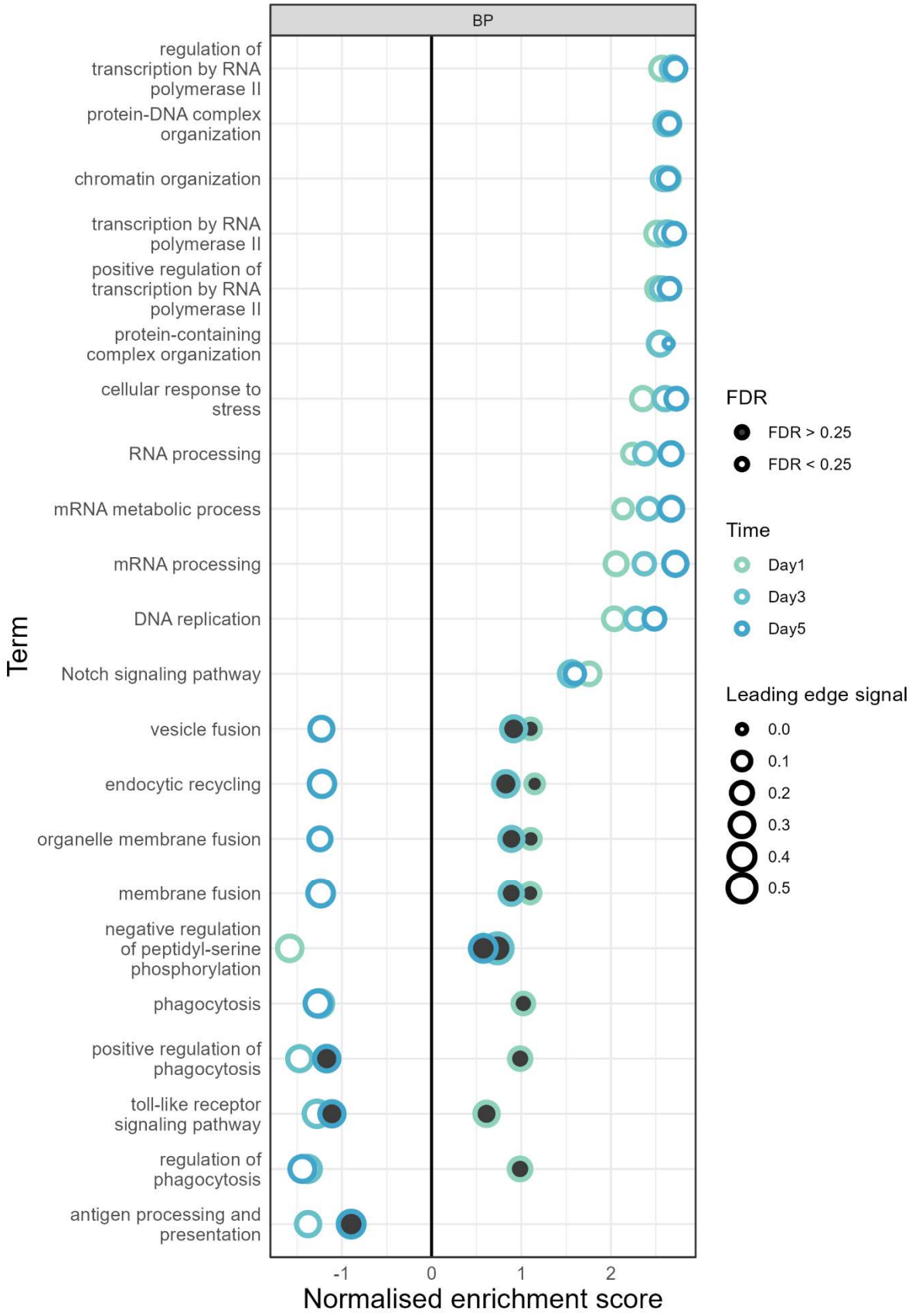


Figure 4.23 ***Gene set enrichment analysis (GSEA) of proteome-normalised phosphoproteome data from HoxB8 neutrophil differentiation (GOBP).*** Genes were pre-ranked based on corresponding phosphopeptide log2(fold change) compared to day 0. Stroke colour represents the time point, size indicates leading edge signal, and permutation-based FDR is indicated by fill colour. Gene ontology (GO) biological process (BP) terms are shown.

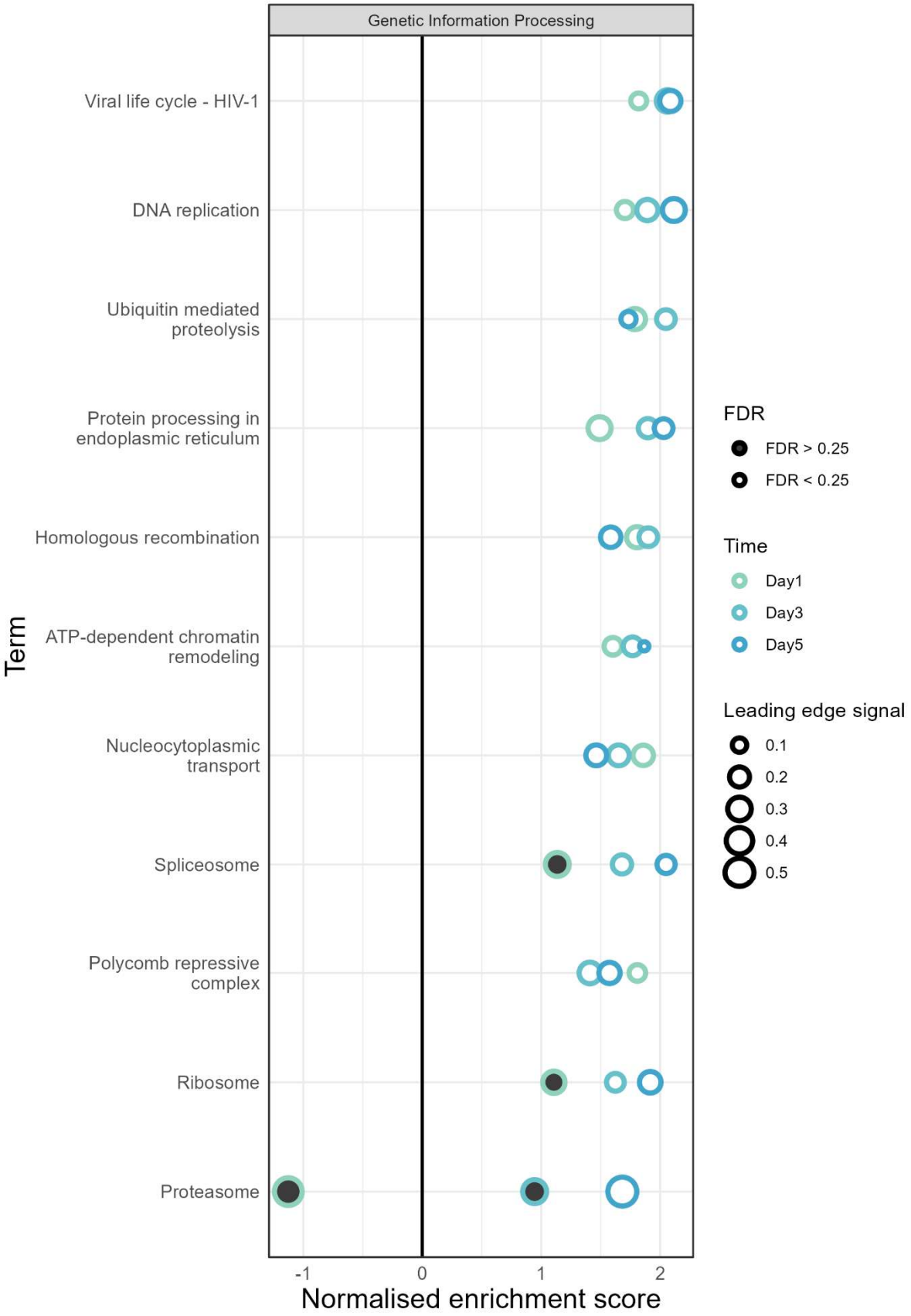


Figure 4.24 ***Gene set enrichment analysis (GSEA) of proteome-normalised phosphoproteome data from HoxB8 neutrophil differentiation (KEGG, Genetic Information Processing).*** Genes were pre-ranked based on corresponding phosphopeptide log2(fold change) compared to day 0. Stroke colour represents the time point, size indicates leading edge signal, and permutation-based FDR is indicated by fill colour. KEGG subcategory 'Genetic Information Processing' terms are shown.

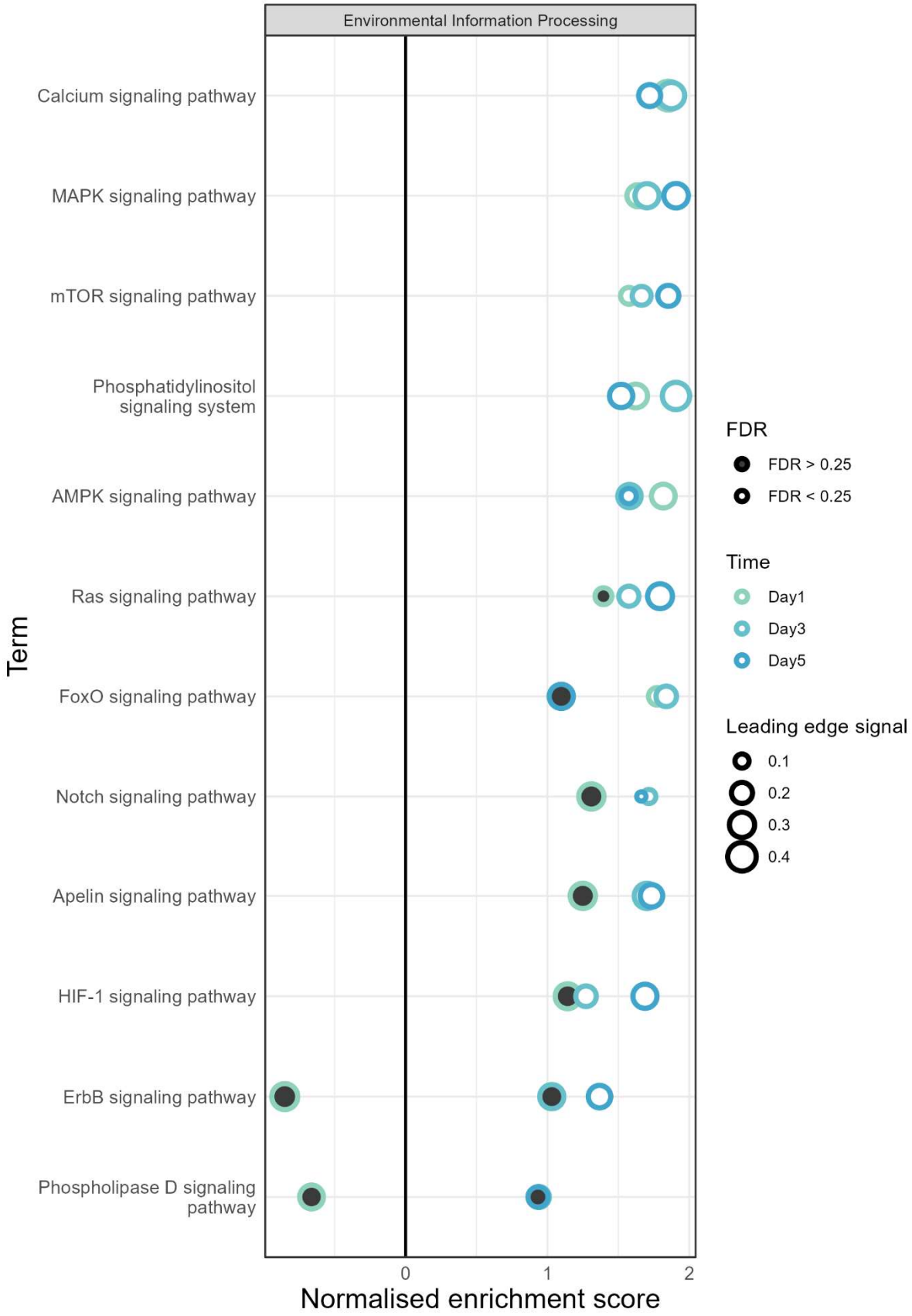


Figure 4.25 ***Gene set enrichment analysis (GSEA) of proteome-normalised phosphoproteome data from HoxB8 neutrophil differentiation (KEGG, Environmental Information Processing).*** Genes were pre-ranked based on corresponding phosphopeptide log2(fold change) compared to day 0. Stroke colour represents the time point, size indicates leading edge signal, and permutation-based FDR is indicated by fill colour. KEGG subcategory 'Environmental Information Processing' terms are shown.

Hierarchical clustering of proteome-normalised phosphopeptide abundance profiles focused on phosphopeptides that were differentially abundant during at least 1 stage of HoxB8 differentiation. 6 clusters were identified (Figure 4.26) and subjected to ORA and ESI analysis.

Many functional enrichment terms failed to reach significant thresholds, likely due to the small number of genes within each cluster, with only GOBP and GOCC terms achieving reasonable FDR levels (< 0.2). Cluster 3 represented a decreasing phosphorylation profile and showed enrichment for terms associated with cytoskeletal reorganisation (Figure 4.27). Cluster 1 showed a peaked phosphopeptide abundance during day 1, and corresponded to terms associated with blood vessels and, notably, Ras protein signalling and regulation, potentially indicating that several proteins within this signalling pathway are more active during earlier stages of HoxB8 differentiation. GOCC terms showed enrichment in cluster 3 (decreasing profile) for terms associated with the cytoskeleton, plasma membrane, and cell periphery (Figure 4.28). This may signify proteins involved in cell migration are dephosphorylated in mature HoxB8 neutrophils, potentially as a way to modulate appropriate cellular morphology and chemotaxis.

ESI analysis (Figure 4.29) revealed substrates of the serine/threonine-protein kinase VRK2 showing increasing phosphorylation profiles (cluster 2 and 5). VRK2 is involved in stress response, MAPK signalling, and p53 activity[171,172]. Moreover, the TTK protein kinase is involved in the spindle assembly checkpoint during mitosis, which is essential for the correct segregation of chromosomes[173]. TTK substrates showed decreasing phosphorylation profiles, suggesting decreased TTK activity is a feature of neutrophil differentiation. The serine/threonine-protein kinase SMG1 is involved in nonsense-mediated decay (NMD), and several cancer cell lines, including multiple myeloma, were shown to be sensitive to SMG1 inhibition[174]. Substrates of SMG1 showed decreasing phosphorylation profiles after day 1 of HoxB8 differentiation, suggesting SMG1 and NMD may be important in cells that closer resemble progenitor cells.

Regarding phosphatases, substrates of protein tyrosine phosphatase non-receptor type 2 (PTPN2) were identified in clusters 1, 2, and 3, which represented distinct phosphorylation profiles. PTPN2 plays regulatory roles in immune signalling, immune cell functions, and immune cell development[175]. PTPN2 has been shown to regulate inflammasome activation and IL1β production in macrophage cells[176]. Calcineurin (PPP3C) participates in calcium signalling and cell cycle progression[177]. The substrates of PPP3C isoforms were identified in clusters 2 and 3, suggesting stage-specific phosphatase activities result in distinct temporal phosphorylation profiles.

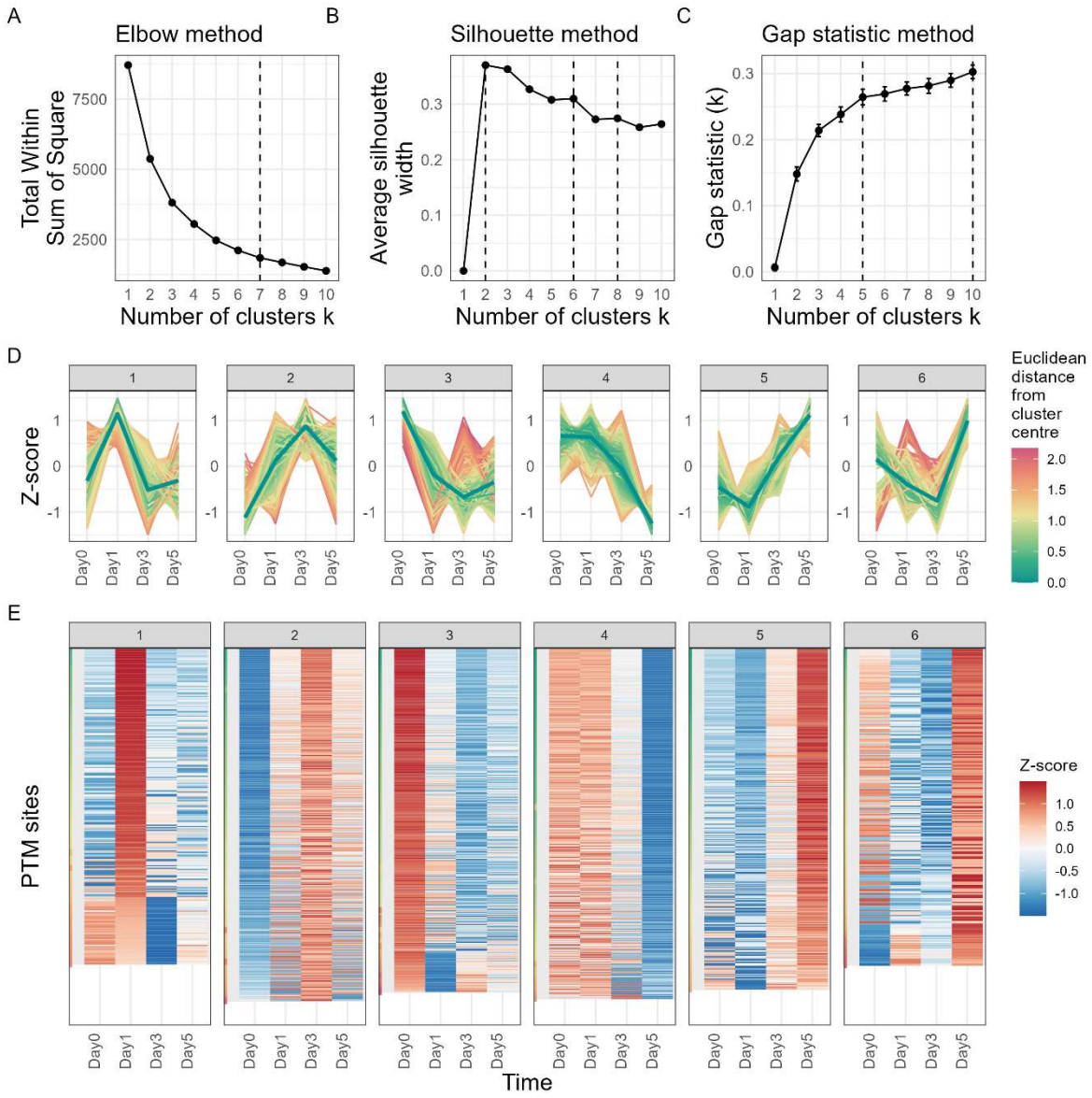


Figure 4.26 ***Hierarchical clustering approach to identify abundance profiles of significantly differentially abundant, proteome-normalised phosphopeptides.*** Peptide values corresponding to the median proteome-normalised peptide abundance (n = 4 biological replicates) at each time point were Z-scored across all time points. Cluster number optimisation approaches included (A) elbow, (B) silhouette, and (C) gap statistic methods. (D) Proteome-normalised phosphopeptide abundance profiles of clusters 1-6. Cluster centre is represented as a thick, dark-green line. (E) Heatmap of clustered phosphopeptides.

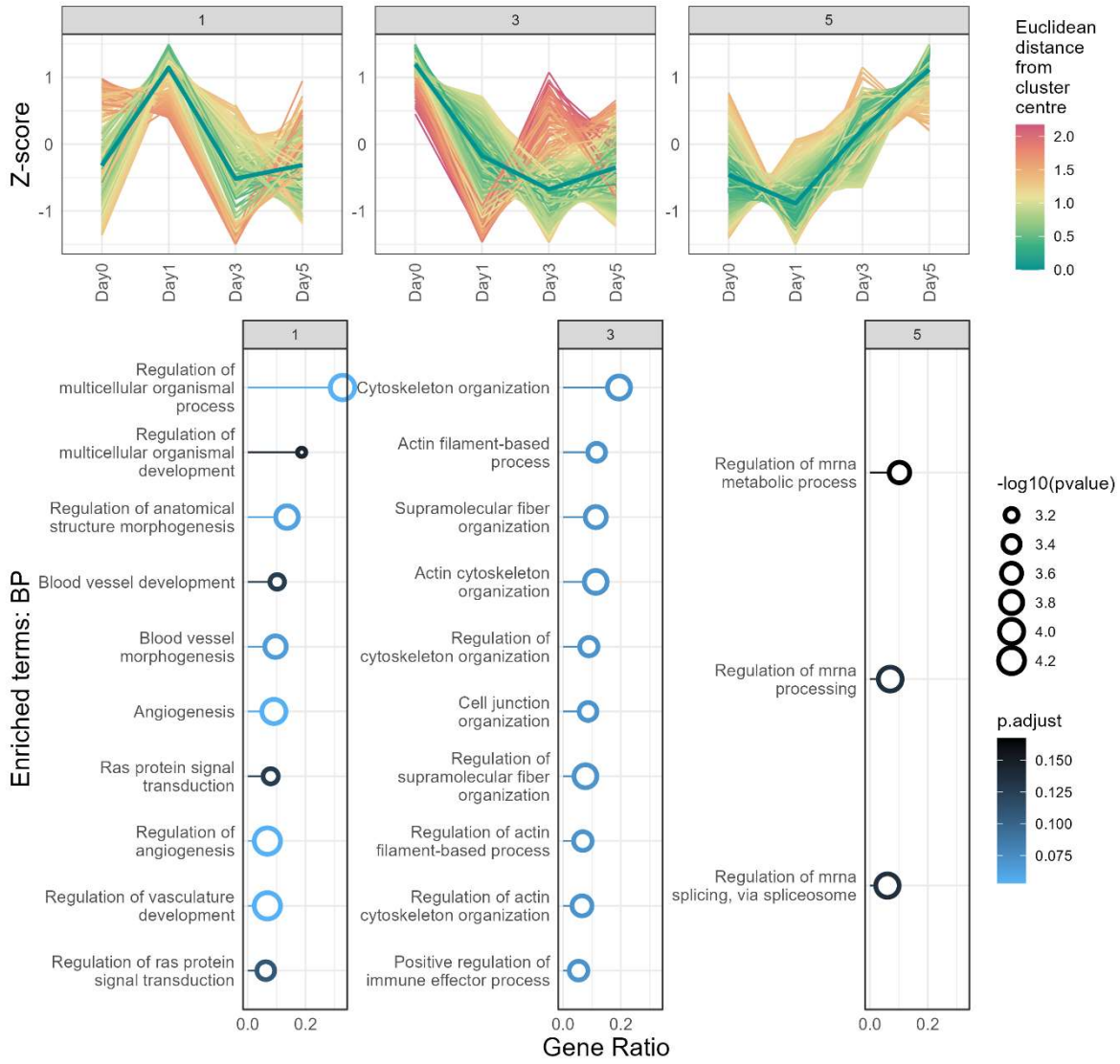


Figure 4.27 ***Functional enrichment analysis of clustered proteome-normalised phosphopeptide profiles that are differentially abundant during HoxB8 neutrophil differentiation (GOBP).*** Over-representation analysis was performed on each cluster, using the proteins identified in the phosphoproteome and total proteome as the background proteins. Point size and colour represent -log10(p-value) and FDR, respectively. Gene ontology (GO) biological process (BP) terms are shown.

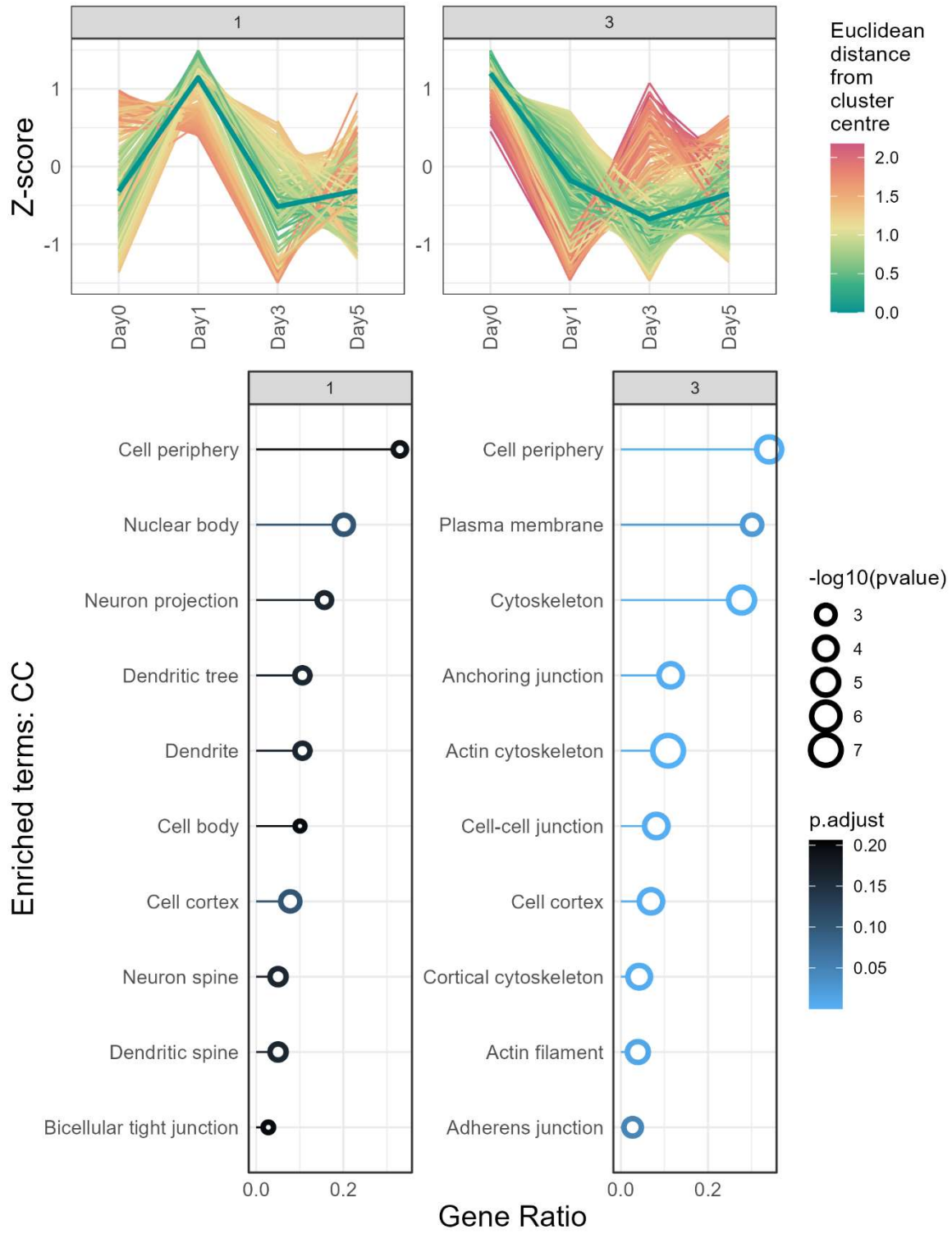


Figure 4.28 ***Functional enrichment analysis of clustered proteome-normalised ubiquitin-modified peptide profiles that are differentially abundant during HoxB8 neutrophil differentiation.*** Over-representation analysis was performed on each cluster, using the proteins identified in the phosphoproteome and total proteome as the background proteins. Point size and colour represent -log10(p-value) and FDR, respectively. Gene ontology (GO) cellular component (CC) terms are shown.

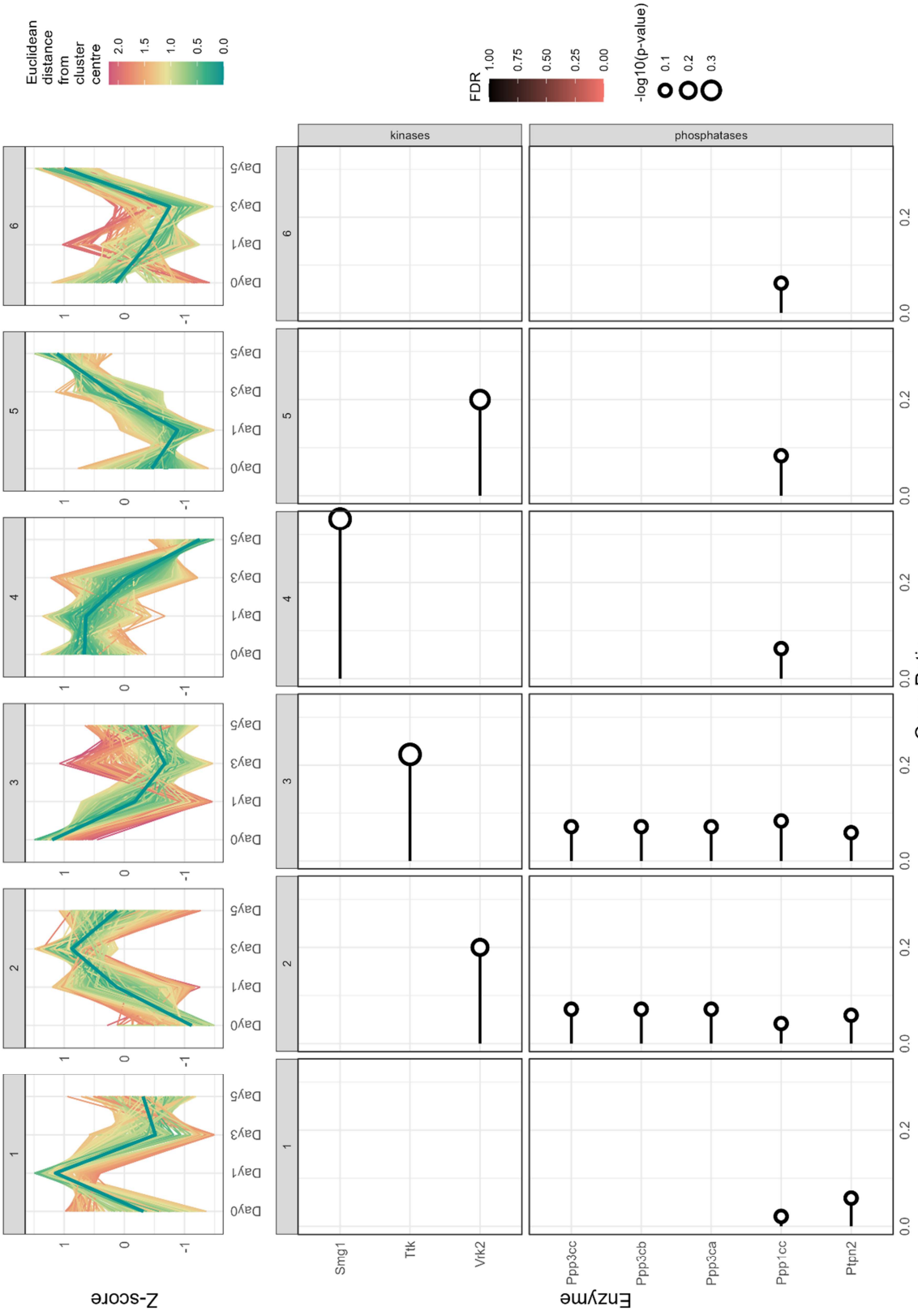


Figure 4.29 ***Enzyme-substrate interaction analysis identifies potential kinases and phosphatases that may regulate proteins within phosphopeptide profile clusters.*** Clusters of proteins identified using the hierarchical approach were analysed. Kinase- and phosphatase-substrate interactions are shown for clusters 1-6. Hypergeometric testing with BH procedure for FDR control were used to determine statistical significance.

4.3. Discussion

Characterisation of the proteome, ubiquitome, and phosphoproteome of the HoxB8 model provided detailed molecular insights into protein and PTM changes that occur during neutrophil differentiation.

The depth of the ubiquitome highlights the pivotal role Ub plays in many cellular processes. Some Ub-peptides with different peptide sequences corresponded to the same ubiquitination site within a protein. However, in some instances, their quantitation was contradictory, showing different abundance profiles over time. For example, two Ub-peptides that correspond to the site K27 on ubiquitin demonstrated contrary abundance profiles, with the two different K27 Ub-modified peptide showing low positive and negative correlations with Ub protein abundance. This may have been caused by sample preparation effects, such as trypsin cleavage efficiency or sequence preference, or bias in MS analysis, where different peptides showed differing ionisation efficiencies, etc. Methods to incorporate such information to provide a representative value for each PTM site, rather than each peptide, would improve the interpretability of the data. Additionally, information regarding the types of Ub linkages and Ub chain architecture is lost upon tryptic digestion of proteins. These details are essential to consider when inferring the biological consequences of ubiquitination, as different Ub linkage types and Ub chain structures can result in substrate- and context-specific behaviours. For example, K27 ubiquitination is typically associated with DNA damage and anti-viral immunity. The ubiquitomic data provides a global overview of this linkage type rather than protein specific Ub linkages, and the opposite abundance profiles of Ub-peptides of Ub K27 creates ambiguity regarding the real changes in the levels of this Ub linkage type. Moreover, the enrichment of Ub-modified peptides was limited to ubiquitination on lysine side chains. It is known that ubiquitination can also occur on other amino acid side chains such as those of cysteine and serine. This information regarding non-canonical ubiquitination is missing from the data.

Phosphoproteome analysis of both DDA- and DIA-MS data enabled a comparison between different data acquisition modes and subsequent database searching methods for precursor identification and quantitation. Since DDA approaches involve the selection of precursors and subsequent fragmentation, as opposed to DIA, which fragments all precursors within an m/z window, assignment of MS2 spectra to MS1 provides a direct route to sequencing peptides, potentially mitigating the effects of neutral loss in phosphopeptides affecting precursor sequence determination and assignment of phosphorylation localisation[178]. However, precursor selection can be random, especially in complex mixtures of peptides, therefore DDA data can contain more missing values and potentially introduce bias. Using ‘pseudo-tandem mass spectra,’ diaTracer improved the number of confident IDs for DIA

data, compared to DIA-NN. However, it is difficult to gauge the impact without a thorough experimental and analytical design with appropriate benchmarking.

Normalisation of PTM levels to protein abundance provides a relative quantification of the 'amount' of protein that is ubiquitinated or phosphorylated. By subtracting the median protein abundance (providing a consensus protein level within the proteome data) from modified peptide abundances, information regarding the relative intensities of PTM and protein can be compared over the different stages of differentiation. More sophisticated approaches include those employed by the MSstatsPTM R package, which uses linear mixed effect model fitting and model-based inferences to determine (near-)stoichiometric quantities of PTMs relative to protein abundance[179].

GSEA used ranking based on log_2FC values (from moderated t-tests) to infer changes in biological process and pathways during each stage of HoxB8 differentiation. Many ubiquitinated and phosphorylated peptides were identified, with several modified peptides derived from the same protein. As such, different levels of modifications could be observed at distinct sites within the same protein. GSEA is performed at the gene level; therefore, it would be unable to capture information about individual protein modification sites that exhibit diverse abundance profiles during differentiation. This may influence resultant functional enrichment results when generating a ranked list. ORA can mitigate these effects since subsets of genes are compared, without the need for ranking metrics.

Several Ub-peptides that showed statistically significant differential abundance during HoxB8 differentiation were derived from kinases and kinase receptors. Conversely, many phosphopeptides were derived from proteins of the ubiquitin system. This may indicate cross-talk between ubiquitination and phosphorylation pathways to bring about appropriate protein regulation and signal transduction, enabling neutrophil differentiation programmes. Receptor proteins, in general, were identified as containing differentially abundant Ub-peptides, suggesting that extracellular cues are essential for neutrophil differentiation, and signal transduction is mediated by both phosphorylation and ubiquitination events. Targeting receptors and receptor-ligand interaction may be a way to modulate granulopoiesis, which could have therapeutic opportunities. For example, G-protein coupled receptors (GPCRs) present one of the most prominent family of proteins that are targeted by drugs[180]. Neutrophils express many GPCRs on their cell surface[181], and (non-)receptor tyrosine kinases[182]; these signalling pathways provide potential avenues to intervene in neutrophil differentiation[183] and effector functions[184,185] through pharmacological agents.

ESI analysis was limited due to the availability of reference databases and data pertaining to validated enzyme-substrate relationships. Expanding these relevant sources of information would improve

statistical testing capabilities and enhance the ability to make predictions on similar datasets. Nevertheless, several DUBs and E3 ligases were identified, with substrate-derived peptides showing interesting and unique abundance profiles during HoxB8 differentiation.

YOD1 substrates showed ubiquitin-modified abundance profiles, relative to their corresponding protein abundances, with increasing levels until day 3 and a subsequent decrease on day 5 of HoxB8 neutrophil differentiation. This unique profile may suggest proteins need to be stabilised for differentiation progression but then degraded in mature HoxB8 neutrophils. Transcriptome and ubiquitome data indicate YOD1 is indeed expressed in HoxB8 neutrophils, suggesting it may be a master regulator of HoxB8 proliferation and subsequent differentiation. YOD1 has been shown to bind to and deubiquitinate NEDD4, supressing NEDD4-induced cell proliferation via the Hippo signalling pathway[186]. Additionally, YOD1 has been demonstrated to stabilise a different E3 ligase, ITCH, enhancing its stability and also modulating Hippo signalling[187]. YOD1 and USP21 exert a synergistic effect on the Hippo signalling pathway, with YOD1 affecting USP21 stability, potentially implicating USP21 in regulatory mechanisms of cell proliferation and survival[188]. The final substrate of YOD1 identified in the ubiquitome data was VCP, a protein involved in lysosomal homeostasis and protein quality control[189,190]. Interactions between YOD1 and VCP play essential roles in the ER-associated degradation (ERAD) pathway[191], which may be a feature of neutrophil differentiation. YOD1 has been shown to stabilise the E3 ligase TRIM33, suppressing tumour progression in head and neck squamous cell carcinoma[192]. Moreover, Yod1^{tm1bNarl}/ Yod1^{tm1bNarl} mice showed an increase in neutrophil cell numbers[193]. Its role in various signalling pathways and biological processes has highlighted YOD1 as a potential therapeutic target in various disease contexts[194], and proteomic analysis of HoxB8 neutrophil differentiation suggests YOD1 may contribute to neutrophil development and maturation.

Phosphorylation appeared to play broader roles in cell cycle progression, DNA and chromatin processing, and mRNA regulation. This was further supported by the types of kinases and phosphatases that were identified, which control cellular processes involved in transcriptional regulation and cell proliferation.

Although the PTMomic data was acquired in the HoxB8 model of mouse neutrophil differentiation, the broader role of ubiquitination and phosphorylation in regulating protein stability and function can be assessed. Proteins may play a general role in cellular homeostasis, and changes in PTM levels could be indicative of a protein's regulatory mechanism. Many ubiquitination and phosphorylation sites for numerous proteins have not been reported in the literature. These datasets provide information on the possible PTMs that can be detected on various proteins, irrespective of the biological context.

Chapter 5: Integrative -omics analysis of HoxB8 mouse neutrophil differentiation reveals novel protein markers and insights into signalling pathway regulation

5.1. Introduction

A model of neutrophil differentiation based on immortalised murine GMP cells – referred to as the HoxB8 model[32] – has been characterised using various modalities to assess the genome, transcriptome, and now proteome and PTMome. Previous work from the Udalova group performed ATAC-seq and RNA-seq to characterise the epigenome and transcriptome of HoxB8 neutrophil differentiation, respectively[110]. Leveraging these diverse datasets using a systems biology approach has the potential to reveal regulatory networks and potential mechanisms that control neutrophil differentiation. For example, key signalling pathways involved in cell differentiation and homeostasis may be regulated at different levels, from the genome through to the proteome.

In the context of a collaborative effort, transcriptome and epigenome data for the HoxB8 neutrophil differentiation model[110] were pre-processed and variance-stabilising transformed (VST) data was shared by Dr. Abhinandan Devaprasad from the Udalova Group in the Kennedy Institute of Rheumatology, University of Oxford. Furthermore, Western blot analysis of proteins identified in the HoxB8 differentiation proteomics was performed to validate the protein changes that were identified from MS-based proteomic workflows, using an orthogonal approach.

5.2. Results

5.2.1. Transcriptome profiles reflect distinct stages of neutrophil differentiation

Transcripts with low VST counts in all stages of differentiation were removed from the dataset, resulting in 18695 transcripts corresponding to 15174 genes (Figure 5.1A). QC of the transcriptome data revealed that VST counts exhibited Poisson distributions (Figure 5.1B) and demonstrated expected groupings and correlations between biological replicates and time points, as illustrated in the PCA plot and PCC matrix (Figure 5.1C,D). Day 0 appeared to have a more distinct transcriptome compared to the other days, suggesting changes in transcriptional programs occur at the early stage of differentiation, and subsequently followed by distinct proteomic changes between day 1 and 3 (see Chapter 4).

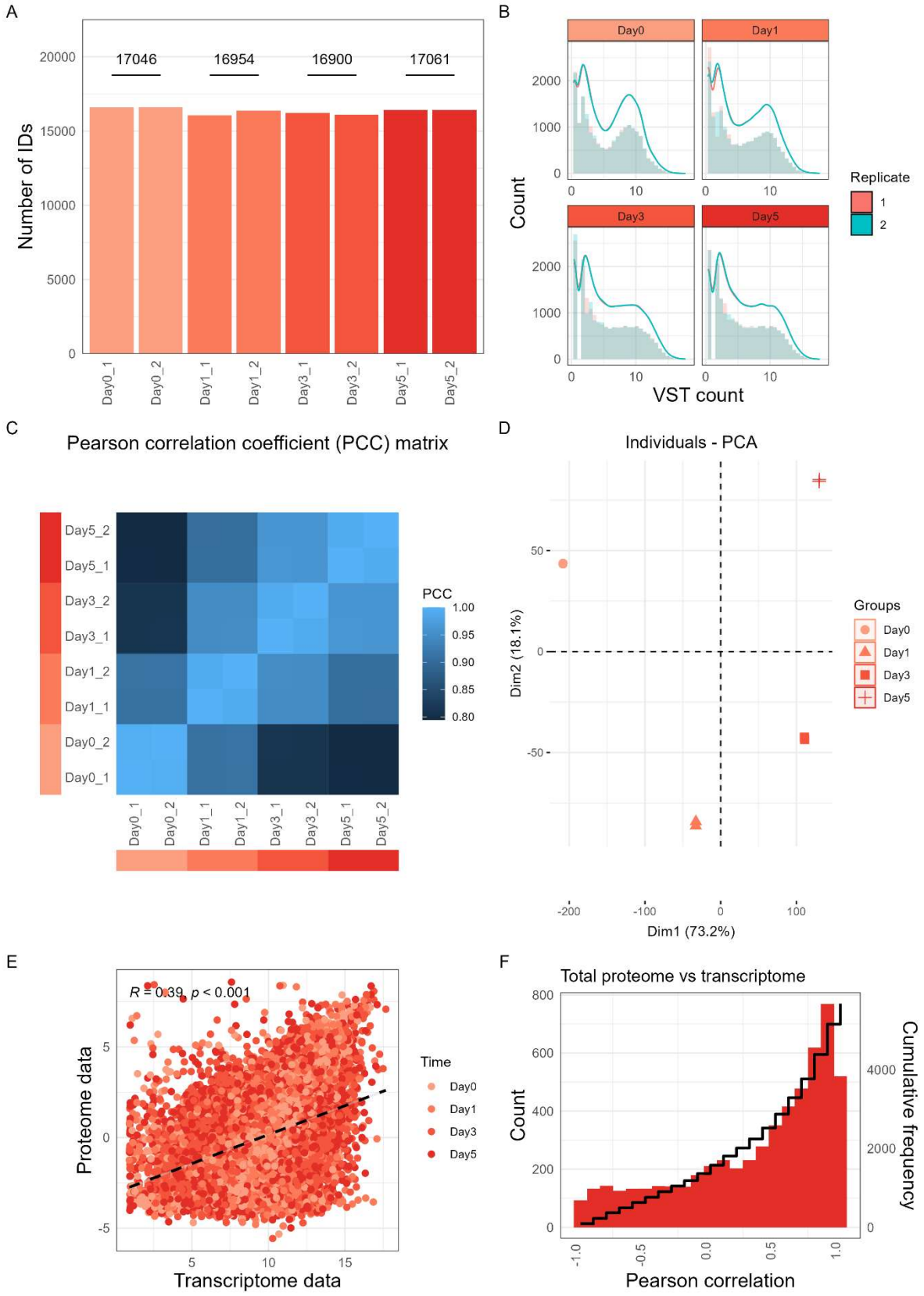


Figure 5.1 ***QC of transcriptome data of HoxB8 neutrophils during differentiation.*** (**A**) Number of transcript IDs per biological replicate, with total IDs per time point shown above each group (n = 2 biological replicates). (**B**) Distribution of variance stabilisation transformed (VST) count data. (**C**) Pearson correlation coefficient between biological replicates and time points of the processed data. (**D**) Principal component analysis of the processed data. (**E**) Comparison of transcriptome and proteome data for all transcripts and corresponding protein levels, for all time points. (**F**) Distribution of Pearson correlation coefficients (PCCs) between mRNA and protein levels, for each transcript-protein pair, over time.

Transcripts were matched to corresponding proteins and correlation between transcriptome counts and proteome intensities, for all time points, were calculated. Overall, the proteome and transcriptome demonstrated low correlation (PCC ≈ 0.39) for all time points (Figure 5.1E).

Correlation per transcript-protein pair over time was then calculated (Figure 5.1F). The distribution of PCCs showed that the majority of proteins correlate with their mRNA transcript levels. Corresponding genes were ranked based on the correlation between protein and transcript levels, and GSEA was performed. GOBP terms showing positive enrichment included mitochondrial gene expression and translation, actin-associated processes, and positive regulation of the ERK1/2 cascade, indicating the proteins connected to these processes have abundance profiles that follow mRNA levels, suggesting transcriptional regulation (Figure 5.2). Interestingly, protein K11-linked ubiquitination also showed positive enrichment, potentially indicating that the ubiquitination machinery may be transcriptionally regulated, which could affect degradation and/or function. Negative enrichment was observed for terms associated with cytokine (IL-7) response, p53-mediated DNA damage response, and cell fate specification. The KEGG subcategory 'Genetic Information Processing' showed positive enrichment for ribosome- and DNA-related terms, and negative enrichment for RNA-related terms such as mRNA surveillance pathway, spliceosome, and RNA degradation, in addition to terms linked to protein processing in the ER and the proteasome (Figure 5.3). These findings suggest that there may be associations between transcriptional regulation and associated machinery and post-translational regulatory mechanisms. Additional GSEA of transcriptome-proteome data can be found in Supplementary Figures 29-34.

5.2.2. Predicting protein abundance from the transcriptome, proteome, and ubiquitome

To better understand the relationship between transcriptome, proteome, and ubiquitome datasets, linear regression was used to assess whether protein abundance could be predicted from corresponding mRNA transcripts and ubiquitinated peptide levels. Median values for each time point were calculated for respective datasets and combined, resulting in 4173 transcripts, 4174 proteins, and 23757 ubiquitinated peptides for analysis. The lm function in the 'stats' R package was used to predict estimates for coefficients of each transcript-protein-ubiquitinated peptide combination, using the following model:

$$Y_p = b_0 + b_1 X_t + b_2 (X_u - X_p)$$

Y_p is protein abundance, X_t is mRNA counts, X_u is ubiquitinated peptide intensity, X_p is protein intensity, and b_0, b_1, and b_2 are coefficients to be estimated for each variable. Protein intensity values were

subtracted from ubiquitinated peptide intensities to account for protein abundance influencing the levels of ubiquitination, as outlined in Chapter 4.

Estimates for each variable coefficient were calculated and filtered based on statistical significance (FDR < 0.05 for b_1 or b_2), resulting in 2322 proteins with corresponding mRNA levels and 8166 ubiquitin-modified peptide profiles. Genes were then ranked by either b_1 or b_2, corresponding to effects from mRNA transcript levels or ubiquitinated peptide intensities, respectively. GSEA was performed on each ranked list separately.

Different KEGG terms for 'Environmental Information Processing' for various signalling pathways were identified for each gene list. Rap1, Ras, Wnt, and sphingolipid signalling pathways were positively enriched when genes were ranked based on the effect of mRNA levels (Figure 5.4), whereas the same pathways showed negative enrichment when genes were ranked based on impact of ubiquitination levels (Figure 5.5). This suggests that these pathways are likely to be transcriptionally upregulated due to larger positive effects of mRNA and post-translationally maintained due to more negative effects of ubiquitination on corresponding protein abundance. MAPK, TGF-beta, and FoxO signalling pathways were positively enriched for genes ranked by ubiquitination levels, potentially indicating that ubiquitination may play a functional role in these pathways rather than a regulatory or degradative one. Signalling pathways that showed positive enrichment common to both gene lists were NF-κB, mTOR, VEGF, and HIF-1, suggesting transcriptional regulation and post-translational modulation contribute to protein activities within these pathways. Further GSEA of transcriptome and ubiquitome contributions can be found In Supplementary Figures 35-48.

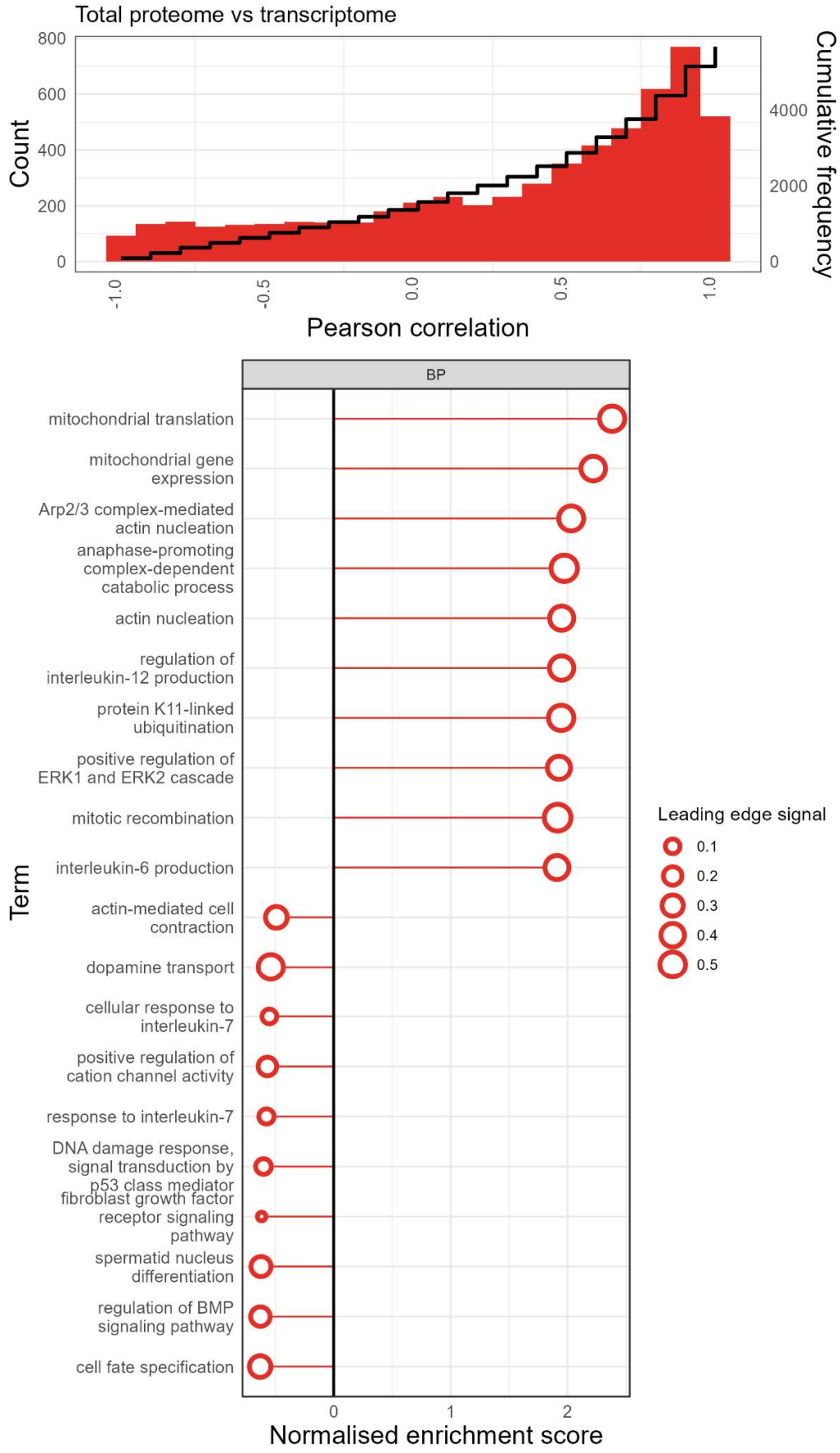


Figure 5.2 ***Gene set enrichment analysis (GSEA) of transcriptome-proteome data during HoxB8 neutrophil differentiation (GOBP).*** Genes were pre-ranked based on correlation between mRNA levels and corresponding protein levels, across all time points of HoxB8 differentiation. Gene ontology (GO) biological process (BP) terms are shown.

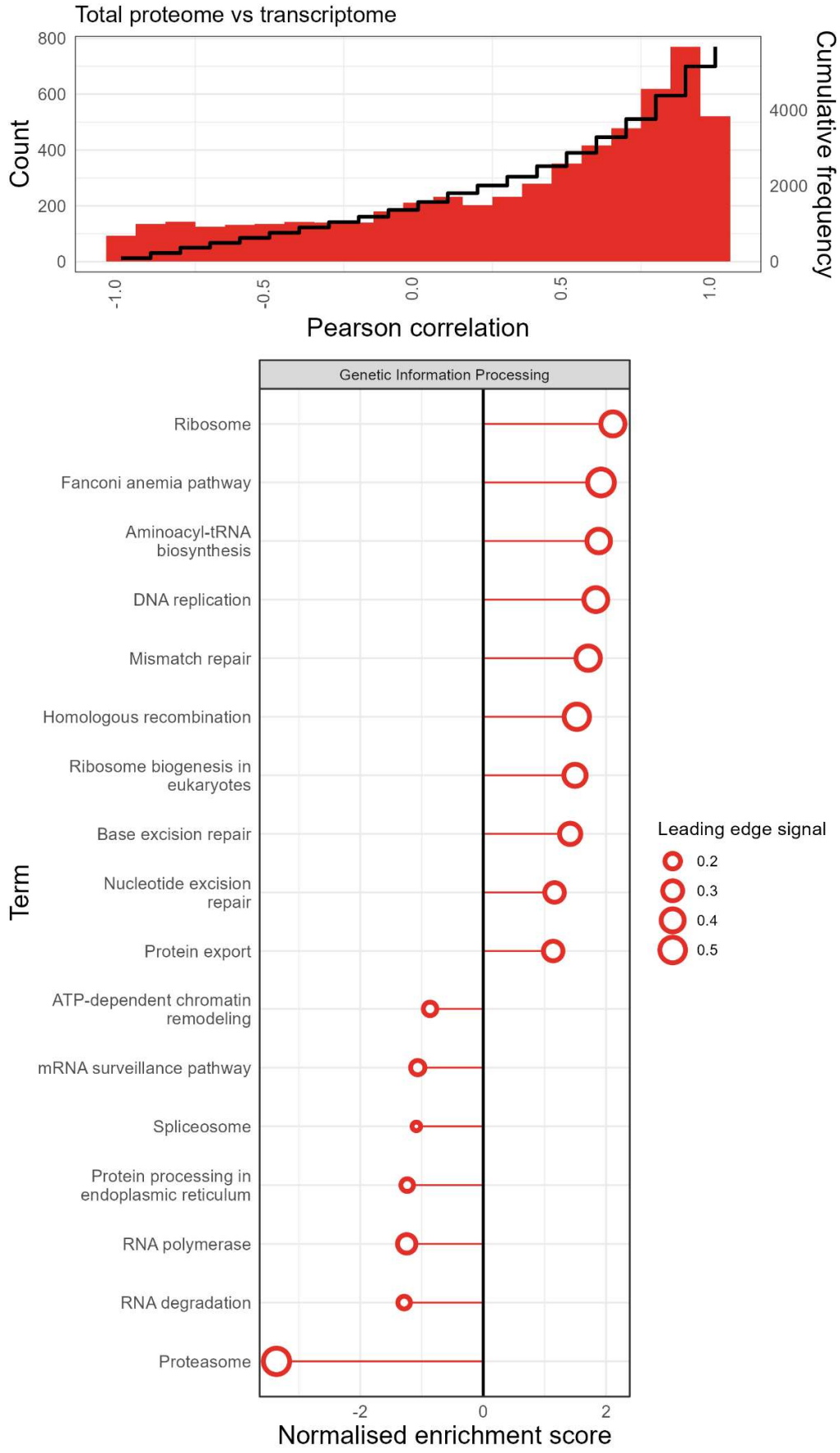


Figure 5.3 ***Gene set enrichment analysis (GSEA) of transcriptome-proteome data during HoxB8 neutrophil differentiation (KEGG, Genetic Information Processing).*** Genes were pre-ranked based on correlation between mRNA levels and corresponding protein levels, across all time points of HoxB8 differentiation. The KEGG pathway subcategory 'Genetic Information Processing' terms are shown.

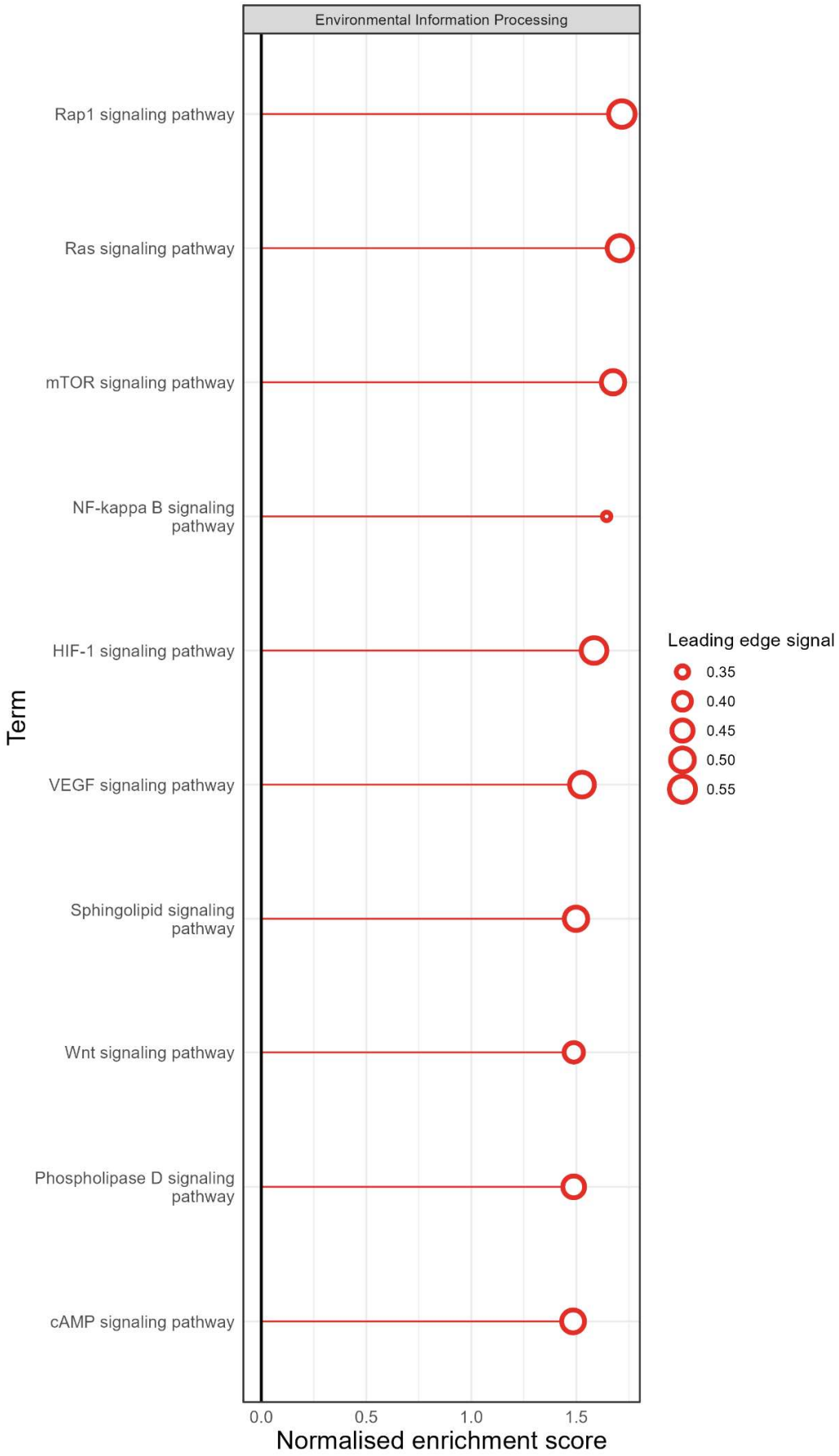


Figure 5.4 ***Gene set enrichment analysis (GSEA) of transcriptome contributions to proteome levels during HoxB8 neutrophil differentiation, using a linear regression approach.*** Genes were pre-ranked based on statistically significant coefficients corresponding to transcriptome variables used to predict protein abundance. The KEGG pathway subcategory 'Environmental Information Processing' terms are shown.

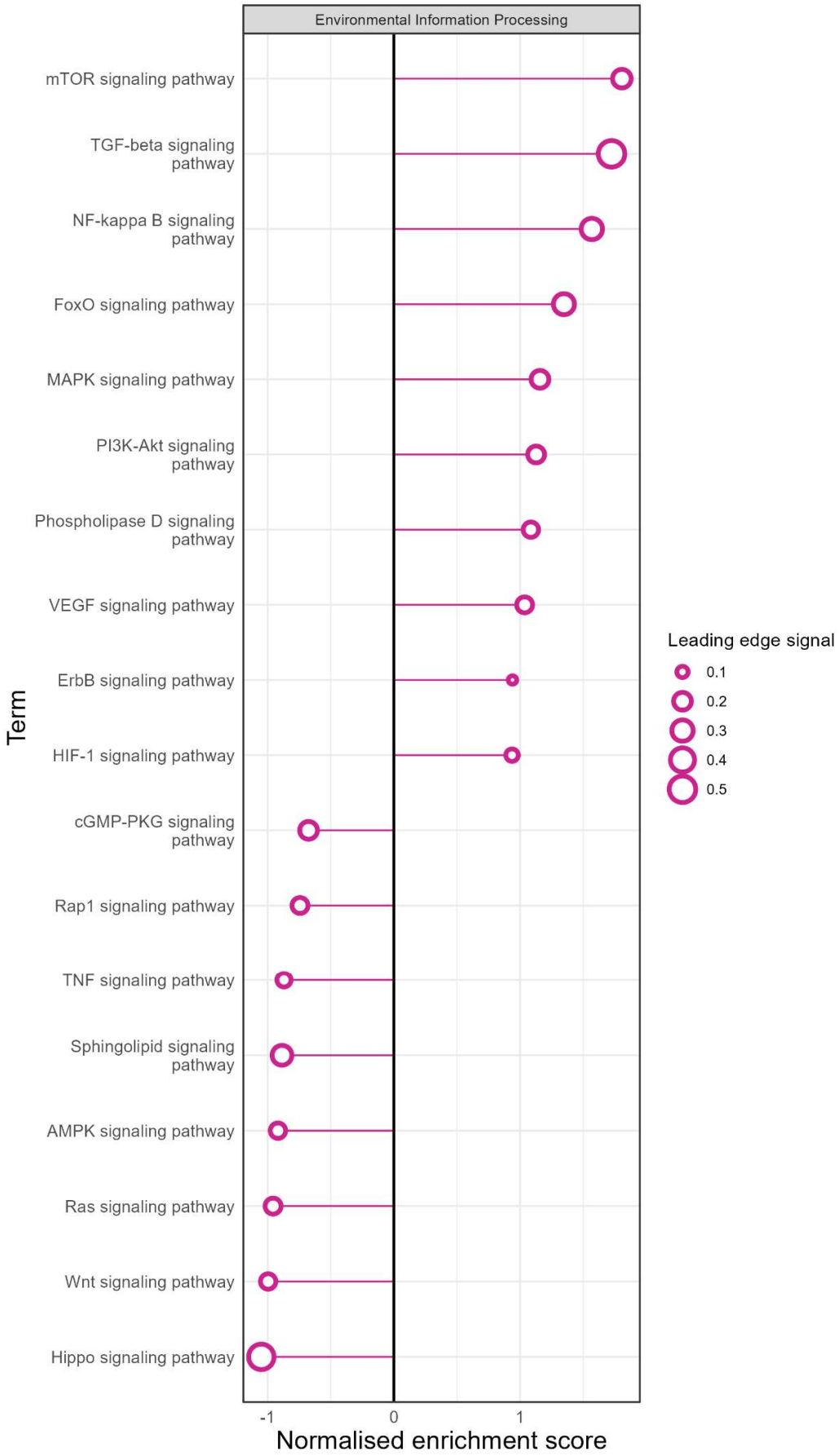


Figure 5.5 ***Gene set enrichment analysis (GSEA) of ubiquitome contributions to proteome levels during HoxB8 neutrophil differentiation, using a linear regression approach.*** Genes were pre-ranked based on statistically significant coefficients corresponding to ubiquitome variables used to predict protein abundance. The KEGG pathway subcategory 'Environmental Information Processing' terms are shown.

5.2.3. Epigenome profiles altered during early stages of neutrophil differentiation

ATAC-seq data was acquired to analyse changes in chromatin accessibility during HoxB8 neutrophil differentiation. Distributions of VST counts suggest chromatin becomes less accessible as HoxB8 neutrophils mature, as indicated by the shift of Poisson distributions towards lower values (Figure 5.6A). PCA and PCCs of the data demonstrated that the epigenome is more similar during day 3 and 5, suggesting that epigenomic changes mostly occur before day 3 (Figure 5.6B,C). Epigenome and transcriptome data, overall, showed little to no correlation (PCC ≈ 0.12) (Figure 5.6D). However, by correlating the epigenome to transcriptome levels for regions proximal to genes and their corresponding mRNA transcript levels, a bimodal distribution of PCCs were observed (Figure 5.6E), with the majority of ATAC-seq peaks showing strong positive (PCC > 0.7) and negative (PCC < -0.7) correlations with transcript levels.

GSEA, on gene lists ranked using epigenome-transcriptome correlation, demonstrated positive enrichment of GOBP terms associated with cell migration, adhesion, and development, as well as negative enrichment for mitochondrial and ribosomal terms (Figure 5.7). Additionally, gene sets associated with protein processing, ubiquitin-mediated proteolysis, mRNA surveillance, RNA degradation, and chromatin remodelling were positively enriched (Figure 5.8). These enriched terms suggest that the relationship between chromatin accessibility and gene expression may primarily affect and regulate homeostatic activities and associated cellular processes. Additional GSEA of epigenome data can be found in Supplementary Figures 49-54.

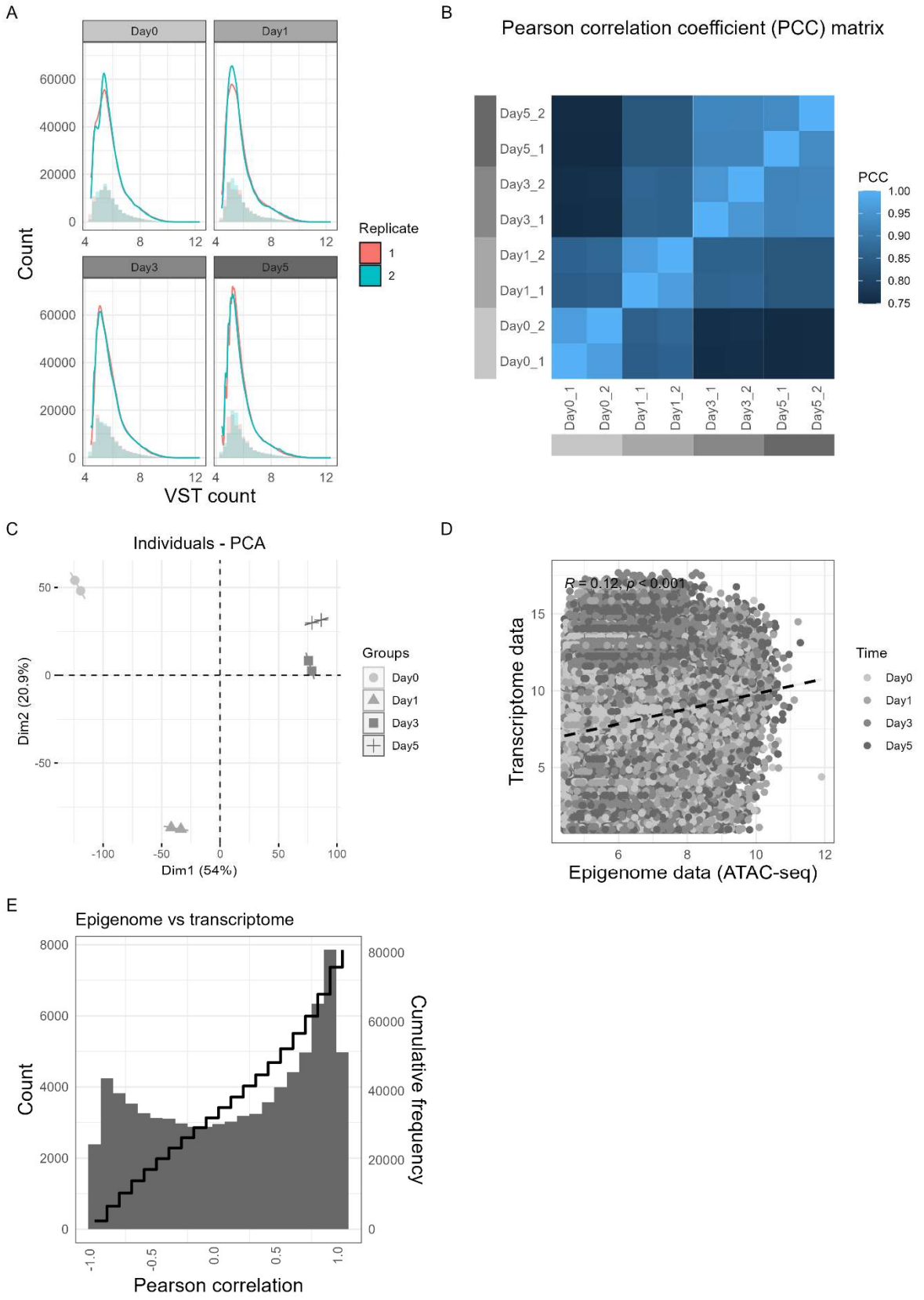


Figure 5.6 ***QC of epigenome data of HoxB8 neutrophils during differentiation.*** (**A**) Distribution of variance stabilisation transformed (VST) count data (n = 2 biological replicates). (**B**) Pearson correlation coefficient between biological replicates and time points of the processed data. (**C**) Principal component analysis of the processed data. (**D**) Comparison of epigenome and transcriptome data for all ATAC-seq peaks corresponding mRNA transcript levels, for all time points. (**E**) Distribution of Pearson correlation coefficients (PCCs) between epigenome and mRNA levels, for each ATAC-seq peak-transcript pair, over time.

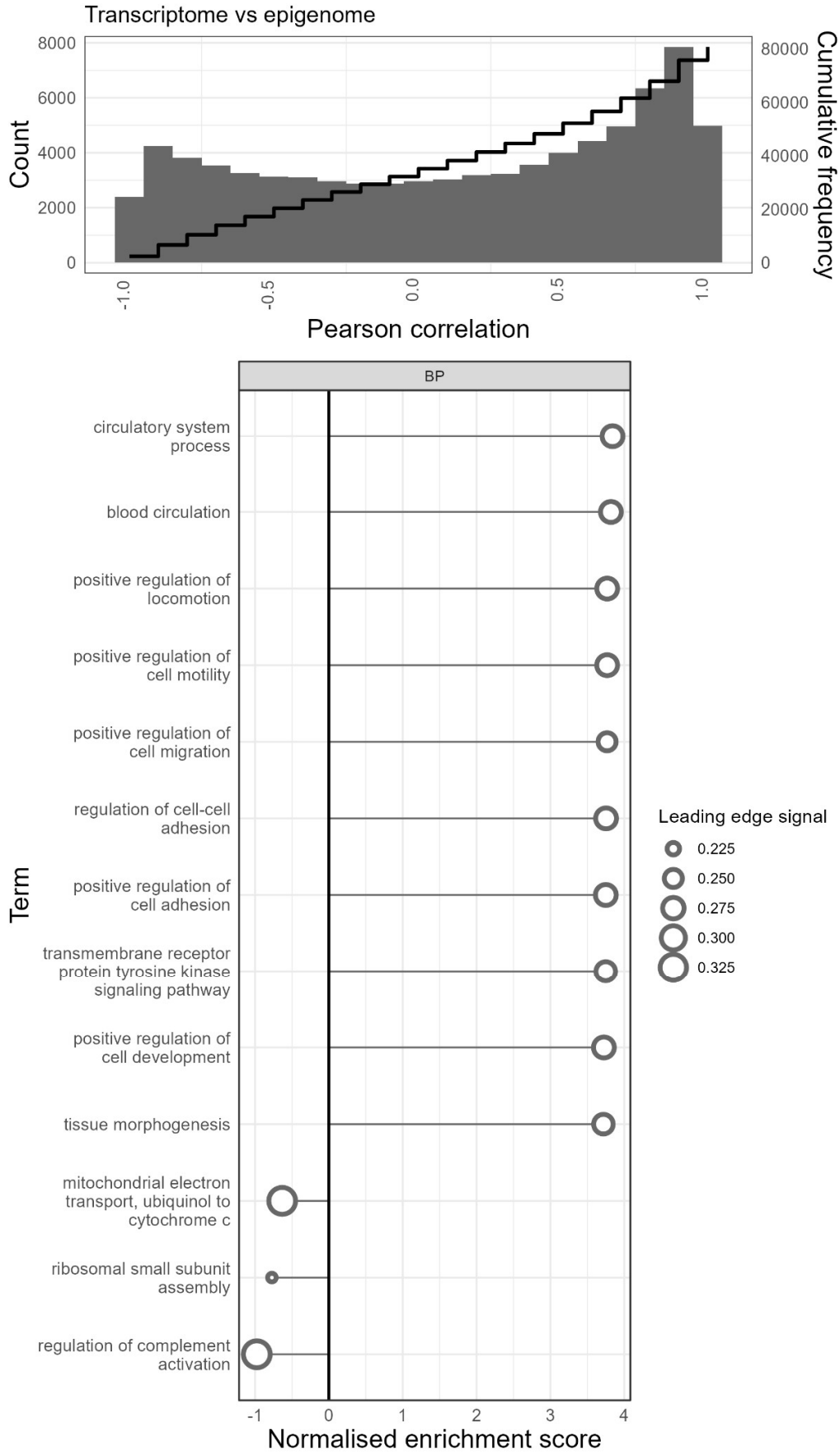


Figure 5.7 ***Gene set enrichment analysis (GSEA) of epigenome-transcriptome data during HoxB8 neutrophil differentiation (GOBP).*** Genes were pre-ranked based on correlation between epigenome levels and corresponding mRNA levels, across all time points of HoxB8 differentiation. Gene ontology (GO) biological process (BP) terms are shown.

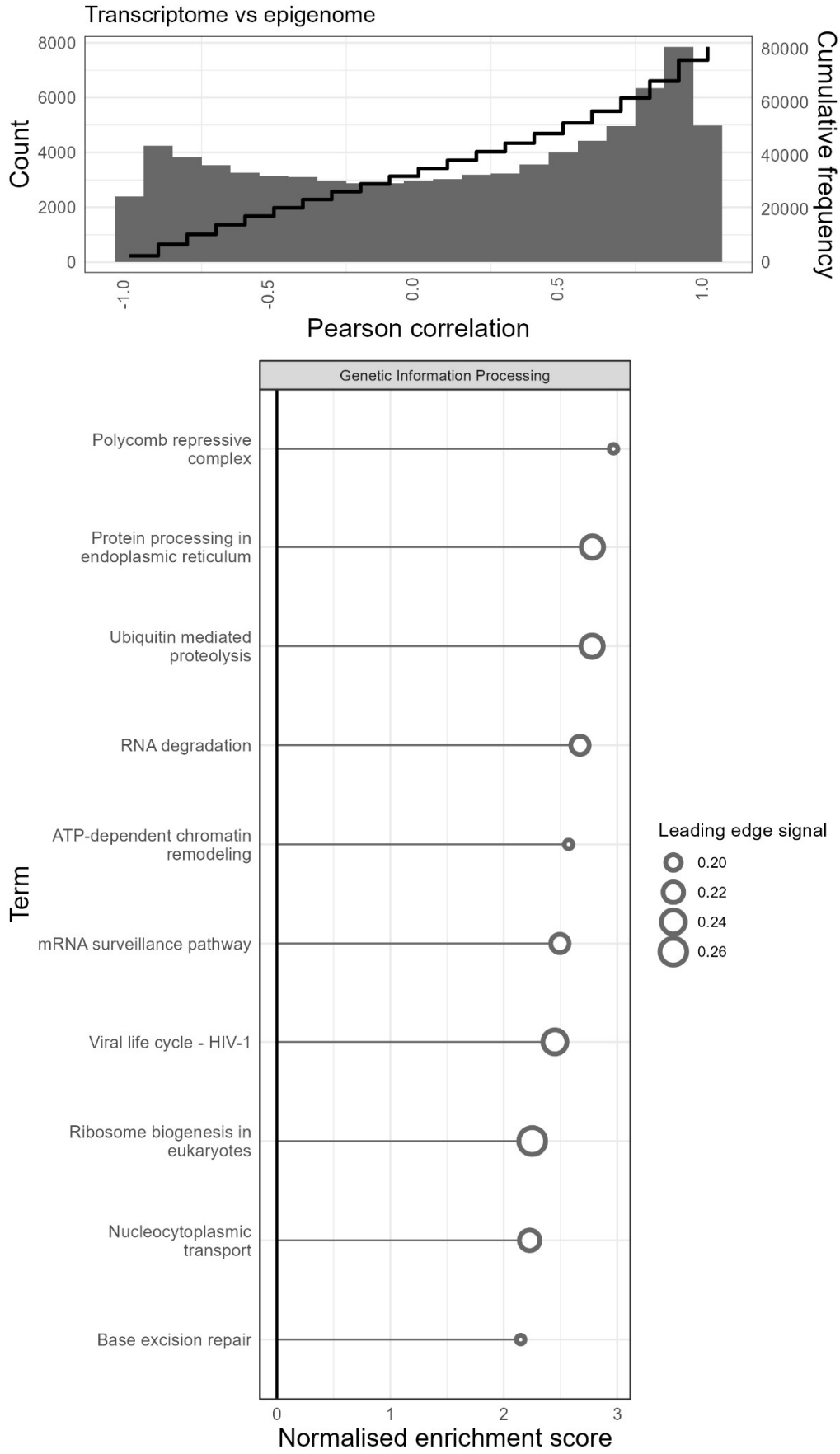


Figure 5.8 ***Gene set enrichment analysis (GSEA) of epigenome-transcriptome data during HoxB8 neutrophil differentiation (KEGG, Genetic Information Processing).*** Genes were pre-ranked based on correlation between epigenome levels and corresponding mRNA levels, across all time points of HoxB8 differentiation. The KEGG pathway subcategory 'Genetic Information Processing' terms are shown.

5.2.4. Predicting regulatory roles of transcription factors in neutrophil differentiation

TFs play a crucial role in epigenetic regulation. Incorporating information from ATAC-seq, RNA-seq, and MS-based proteomic analysis, and leveraging data from TF Link[127] (a database consisting of comprehensive transcription factor-target gene interactions), TFs identified on the proteome level can be linked to their target genes on the epigenome and transcriptome levels. By doing so, predictions can be made regarding the effect of transcription factor presence, or abundance, and epigenome changes on mRNA transcript levels during HoxB8 differentiation. Using a linear regression approach, the coefficients for the following equation were predicted:

$$Y_t = b_0 + b_1 X_e + b_2 X_{TF}$$

$$Y_t = b_0 + b_1 X_e + b_2 D_{TF}$$

Where Y_t corresponds to mRNA transcript level, X_e is chromatin accessibility (epigenome) levels, X_{TF} is TF abundance, D_{TF} is a binary variable representing TF presence or absence, and b_0, b_1, and b_2 are coefficients to be predicted for each variable. Since TF activity may not necessarily depend on its abundance, both approaches were considered.

Using the above models to estimate the effects of TF and epigenome levels on mRNA levels, the predictor variable coefficients failed to reach an appropriate FDR threshold, potentially due to the large number of comparisons made and the complex interplay between TF abundance/presence, chromatin accessibility, and gene expression. Instead, all data were maintained without p-value or FDR filtering. Accordingly, these coefficient estimates were used to rank corresponding genes for GSEA. GSEA did not provide any meaningful enrichment when the TF presence/absence model was used (data not shown). Therefore, the TF abundance model was used, and the analysis space covered 104 TFs, 1028 target gene transcripts, and 8592 ATAC-seq peaks proximal to those genes.

Looking at KEGG ‘Environmental Information Processing’ terms, ranking based on epigenome and TF abundance effects both showed positive enrichment for TNF, PI3K/AKT, FoxO, Ras, and Wnt signalling pathways, suggesting increased chromatin accessibility and changes in TF abundance lead to increased mRNA levels of associated genes (Figure 5.9, Figure 5.10). The TFs whose target genes contributed to the core enrichment of each signalling pathway are listed in Table 5.1. Further GSEA of epigenome and TF abundance contributions can be found in Supplementary Figures 55-68.

Table 5.1 ***Transcription factors whose genes contributed to the core enrichment of signalling pathways in the context of neutrophil differentiation.***

Signalling pathway	Transcription factors
FoxO signalling pathway	CREBBP, E2F4, EP300, GABPA, NFKB1, SP1, SPI1, STAT3, TP53
Hippo signalling pathway	CREB1, EHMT2, NFKB1, RFX1, RUNX2, SP1, SPI1, TP53
PI3K-Akt signalling pathway	CREB1, EZH2, FMR1, HDAC2, HDAC3, NFKB1, NR3C1, POU2F1, RELA, RFX1, SP1, SPI1, STAT3, TP53
Ras signalling pathway	FMR1, POU2F1, RBPJ, RELA, RFX1, SMARCA4, SP1, TP53
TNF signalling pathway	CEBPA, CEBPB, ESR1, EZH2, NFKB1, NR3C1, RELA, SP1, STAT4, TP53
Wnt signalling pathway	EHMT2, SP1, TP53

Ras and Wnt signalling pathways also demonstrated large positive effects of mRNA levels and negative effects of ubiquitination levels on protein abundance, as shown in the previous analysis (Figure 5.4, Figure 5.5). In contrast, TNF, PI3K/AKT, and FoxO signalling pathway proteins did not show enrichment for the contributions of mRNA levels to their protein abundance. However, the contributions of ubiquitination to protein abundance were positive for PI3K/AKT and FoxO, and negative for TNF. Moreover, the Hippo signalling pathway was only positively enriched based on TF abundance effects, and associated proteins showed negative effects from ubiquitination. Table 5.2 provides a summary of the contributions of the epigenome and TF abundance to mRNA levels, as well as the relationship between mRNA levels, ubiquitination, and protein abundance. These observations suggest that different signalling pathways are likely regulated at different levels across gene expression, influencing overall protein abundance, with potential combinatorial contributions, during HoxB8 neutrophil differentiation.

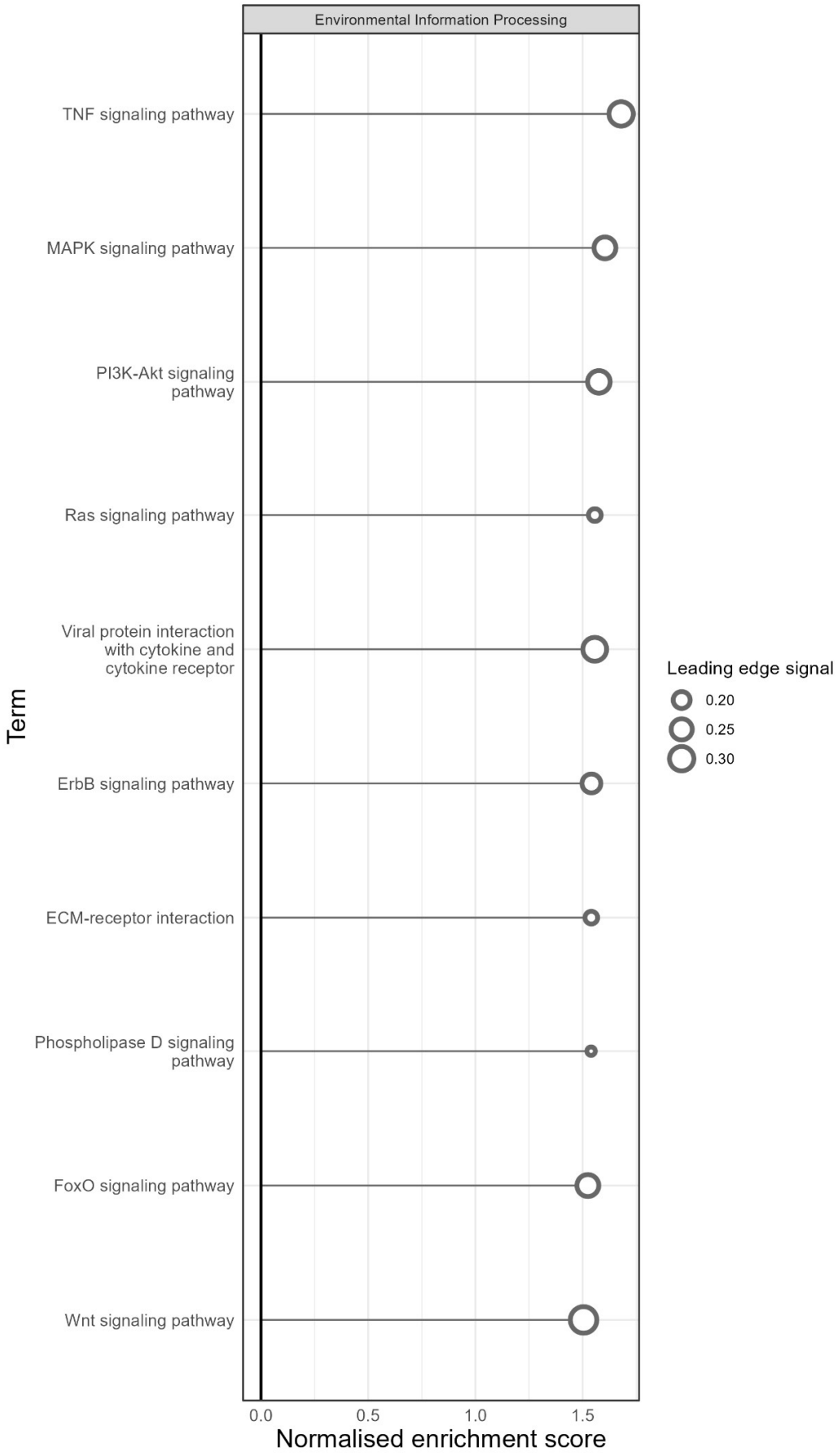


Figure 5.9 ***Gene set enrichment analysis (GSEA) of epigenome contributions to mRNA levels during HoxB8 neutrophil differentiation, using a linear regression approach.*** Genes were pre-ranked based on statistically significant coefficients corresponding to epigenome variables used to predict mRNA levels. The KEGG pathway subcategory 'Environmental Information Processing' terms are shown.

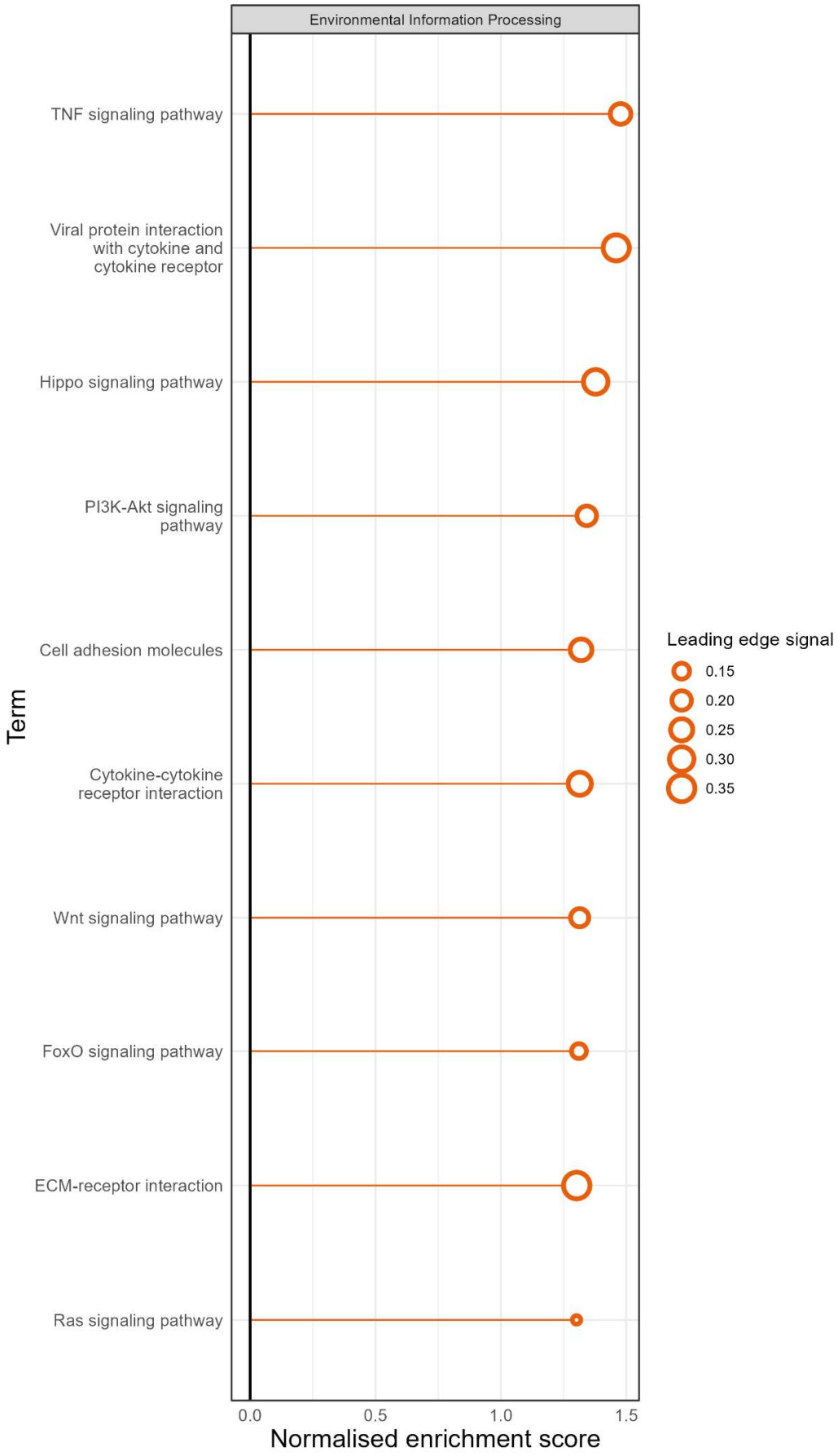


Figure 5.10 ***Gene set enrichment analysis (GSEA) of TF abundance contributions to mRNA levels during HoxB8 neutrophil differentiation, using a linear regression approach.*** Genes were pre-ranked based on statistically significant coefficients corresponding to TF abundance variables used to predict mRNA levels. The KEGG pathway subcategory 'Environmental Information Processing' terms are shown.

Table 5.2 ***Summary of contributions of epigenome and TF abundance to mRNA levels, and mRNA and ubiquitination levels to protein abundance in enriched signalling pathways.*** *'+' indicates a positive effect, '–' indicates a negative effect, and a blank indicates no enriched effect.*

Signalling pathway	mRNA levels		Protein abundance	
	Epigenome (ATAC-seq) contributions	TF abundance contributions	mRNA level contributions	Ubiquitination contributions
TNF	+	+		–
MAPK	+			+
PI3K/AKT	+	+		+
Ras	+	+	+	–
ErbB	+			+
Phospholipase D	+		+	+
FoxO	+	+		+
Wnt	+	+	+	–
Hippo		+		–
Rap1			+	–
mTOR			+	+
NF-κB			+	+
HIF-1			+	+
VEGF			+	+
Sphingolipid			+	–
cAMP			+	
TGFβ				+
cGMP-PKG				–
AMPK				–

5.2.5. HoxB8 neutrophil proteomics reveals novel protein markers associated with differentiation

Due to large proteomic changes that occur during HoxB8 differentiation, whole cell lysates from day 0, 1, 3, and 5 were quantified for protein amounts and samples were normalised for SDS-PAGE loading. Coomassie staining confirmed that similar amounts of protein were loaded onto the gel across all samples (Figure 5.11). A prominent band between 37 and 50 kDa exhibited an elevated signal in later stages of differentiation (days 3 and 5). This band likely corresponds to β-Actin (ACTB), a key component of the actin cytoskeletal that also demonstrated increasing abundance in the HoxB8 proteomic analysis. The increased abundance of cytoskeletal proteins is consistent with neutrophil biology, as neutrophils are highly motile cells that exhibit a dynamic and flexible morphology. ACTB is typically used as a loading control in Western blot analysis. However, the proteomic data indicates that ACTB would not be suitable for this purpose.

Western blot analysis detected protein bands, across all time points, for Integrin alpha M (ITGAM) or CD11b, CD34, high affinity immunoglobulin epsilon receptor subunit gamma (FCER1G), RHOA, NAD-dependent protein deacetylase sirtuin-3, mitochondrial (SIRT3), and transglutaminase 2 (TGM2) (Figure 5.12A). The observed protein band signals recapitulated protein abundance trends determined in the HoxB8 differentiation proteomics data (Figure 5.12B,C).

5.2.6. NLRP3 inflammasome components are upregulated during neutrophil differentiation

Data from Chapter 3 and Chapter 4 indicated a potential role of NOD-like receptor signalling in HoxB8 differentiation. Accordingly, components of the NLRP3 inflammasome showed increased protein abundance over time. Integration with other -omics datasets suggest this may be transcriptionally regulated (Figure 5.13A). Transcriptome and ubiquitome changes in NLRP3, apoptosis-associated speck-like protein containing a CARD (PYCARD or ASC), caspase 1 (CASP1), and Gasdermin D (GSDMD) showed similar trends to corresponding protein abundance changes. Examining the epigenome data, genomic regions near these NLRP3 inflammasome genes exhibited increased chromatin accessibility during HoxB8 differentiation. This was accompanied by an increase in mRNA levels. Despite increasing chromatin accessibility and mRNA levels of Interleukin-1 beta (IL1β) and IL18, IL1β and IL18 proteins were not detected. Only one IL1β-derived ubiquitinated peptide was identified on day 5, suggesting IL1β may be readily ubiquitinated and degraded if and when translated. Alternatively, IL1β and IL18 could be secreted. These findings suggest NLRP3 inflammasome priming naturally occurs during HoxB8 neutrophil differentiation. Since IL1β and IL18 are effector molecule of the NLRP3 inflammasome, HoxB8 neutrophils may need activation in order to detect these cytokines. GSDMD appears to be the only main effector molecule of the NLRP3 inflammasome that was detectable in the total proteome.

Western blot validation confirmed increased protein levels of NLRP3 and CASP1 in mature HoxB8 neutrophils (Figure 5.13B,C). IL1β was not detected in mature HoxB8 cell lysates during Western blot analysis (data not shown).

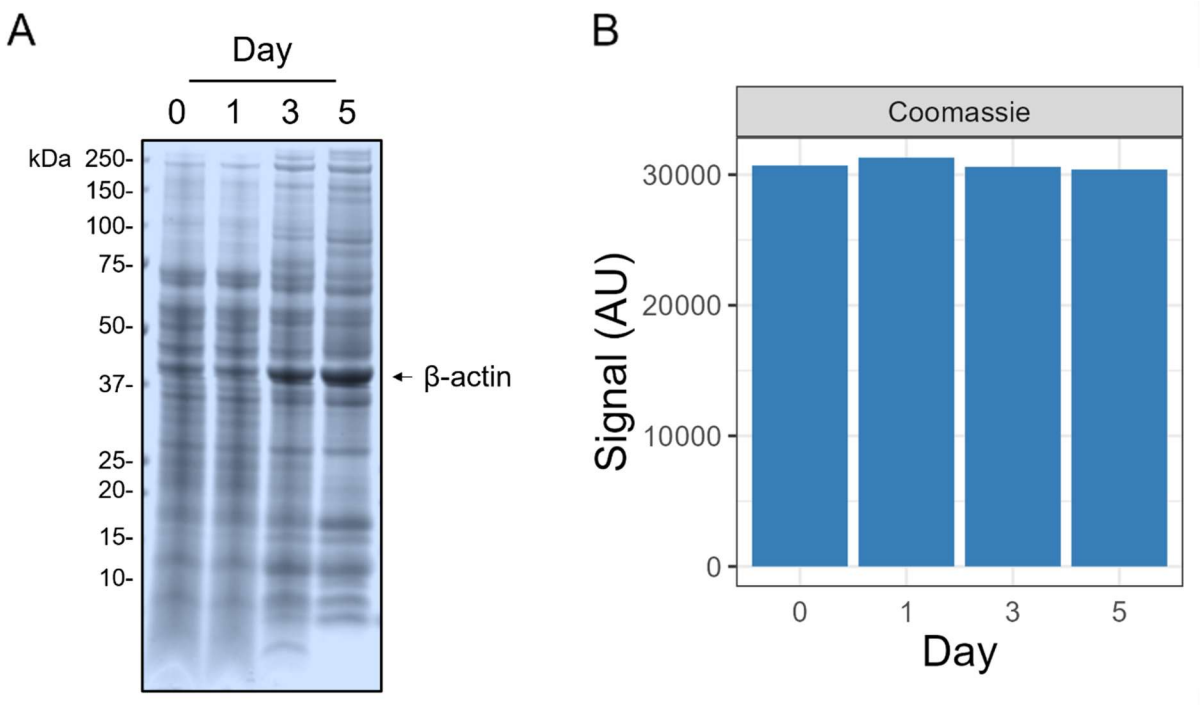


Figure 5.11 ***SDS-PAGE analysis of HoxB8 neutrophils during differentiation.*** (**A**) Coomassie staining of SDS-PAGE gel. Cells were collected on days 0, 1, 3, and 5 post-induction of differentiation. The arrow indicates approximate molecular weight of β-actin (ACTB) (**B**) Coomassie stain signals (n = 1).

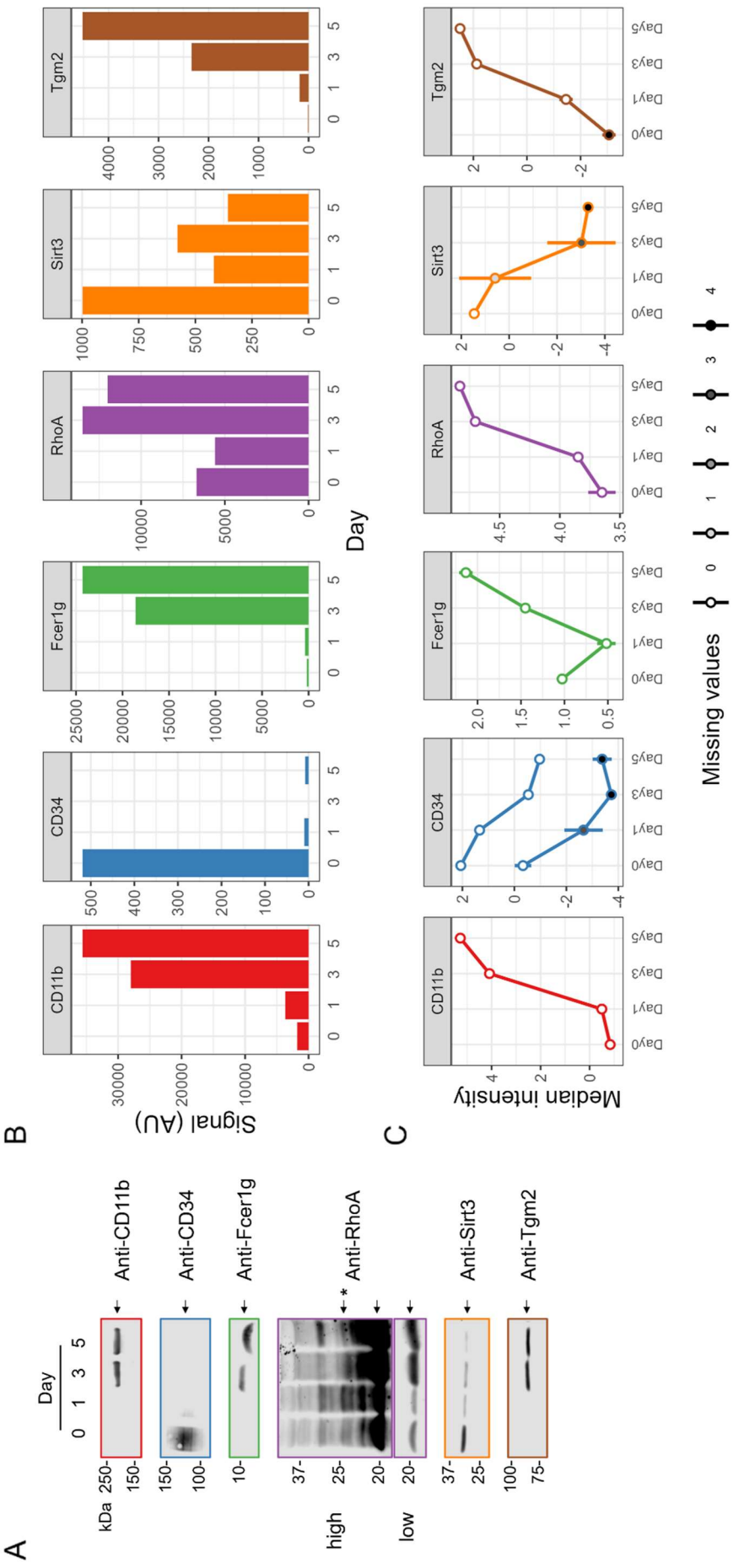


Figure 5.12 ***Western blot validation of select proteins during HoxB8 neutrophil differentiation.*** (**A**) Western blot analysis. Antibodies are shown on the right. Arrows point to corresponding protein bands. High and low correspond to exposure levels during imaging. (**B**) Western blot signal quantitation. Antibody signals for each protein are shown (n = 1). (**C**) Protein abundance profiles. Median intensity values are on a log2 scale. Error bars represent standard error of median. Fill colour of points represents number of missing values for the corresponding protein and time point. Two protein isoforms of CD34 are shown. Error bars represent the standard error of median (SEM).

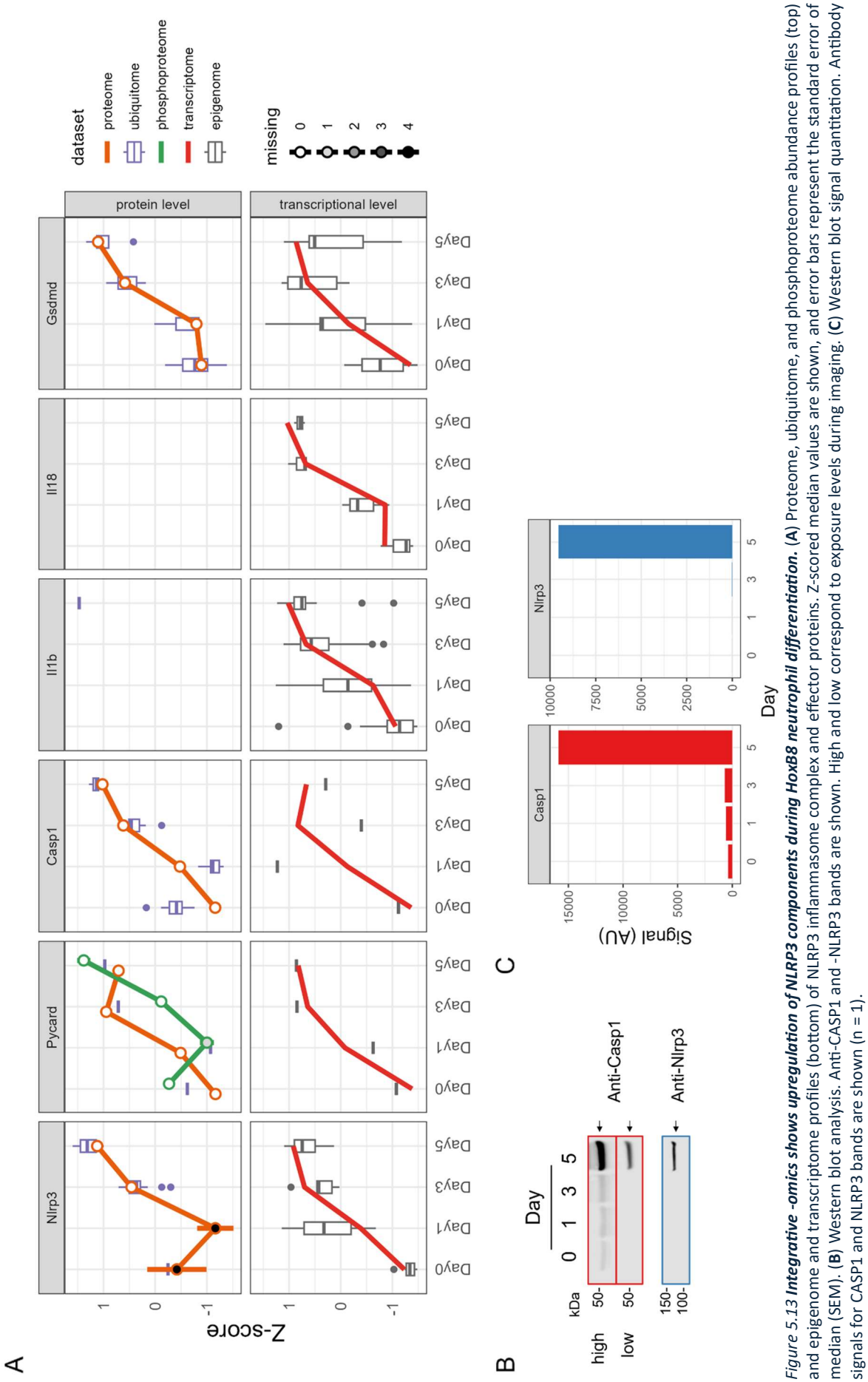


Figure 5.13 ***Integrative -omics shows upregulation of NLRP3 components during HoxB8 neutrophil differentiation.*** **(A)** Proteome, ubiquitome, and phosphoproteome abundance profiles (top) and epigenome and transcriptome profiles (bottom) of NLRP3 inflammasome complex and effector proteins. Z-scored median values are shown, and error bars represent the standard error of median (SEM). **(B)** Western blot analysis. Anti-CASP1 and -NLRP3 bands are shown. High and low correspond to exposure levels during imaging. **(C)** Western blot signal quantitation. Antibody signals for CASP1 and NLRP3 bands are shown (n = 1).

5.3. Discussion

5.3.1. Integrative -omics offers unique molecular insights into neutrophil differentiation

Genetic information processing is fundamental to cellular function. Genes, encoded in the genome of a cell (the DNA), are expressed through transcription, resulting in mRNA. Gene expression depends on the interplay between chromatin dynamics and TFs. The resultant mRNA is then processed and translated at ribosomes, giving rise to protein production. Proteins can then be post-translationally modified, which affects their stability and/or function. These correlative analyses attempted to ascertain potential links between different biological macromolecules associated with specific genes within the context of the central dogma of molecular biology. TF abundance and chromatin accessibility surrounding target genes may affect mRNA expression levels. Similarly, mRNA levels may affect protein expression, and these proteins could be modified, consequently affecting overall protein abundance.

Gene expression is a complex process involving chromatin dynamics, TF interactions, and regulatory elements proximal and distal to gene loci. ATAC-seq data may not fully capture the regulatory aspects for all genes and their expression, and may not be indicative of transcriptionally active regions or TF binding hotspots. Additionally, the activities of a TF may depend on factors such as TF binding sites, transcriptional activation versus repression, and combinatorial effects with other TFs, protein complexes, chromatin architecture, chromatin-associated proteins, and TF PTM status. CHIP-seq approaches would allow for the direct characterisation of TF binding, but only granting a snapshot of a specific TF, limited by the availability of robust antibodies that target these proteins.

TFLink contains a diversity of TF-target gene interactions. However, some TFs identified in the database are not conventional TF, and their target genes were obtained in large-scale experiments. As such, analysis focused on data derived from small scale experiments but this limits the analysis to validated TF-target gene interactions. Nevertheless, several TFs, as identified in the proteomic data, and their abundances were predicted to influence mRNA levels of target genes involved in several neutrophil-associated pathways. C/EBPα, C/EBPβ, and SPI1 (PU.1) are already known to participate in neutrophil differentiation[1]. Signal transducer and activator of transcription 3 (STAT3) has also been demonstrated to facilitate neutrophil differentiation in response to G-CSF exposure[195,196].

NF-κB signalling pathway participates in neutrophil inflammatory response and can even induce C/EBPα transcription, increasing neutrophil production upon cytokine exposure[197,198]. TFs involved in this pathway include NFKB1 and RELA. The cAMP responsive element binding protein 1 (CREB1) is involved in neutrophil activation and driving pro-inflammatory phenotypes[199]. The CREB binding

protein (CREBBP) is a tumour suppressor gene involved in gene expression, cell growth, and differentiation. CREBBP/EP300 activity was shown to affect neutrophil differentiation, where inhibition led to a decrease in abnormal differentiation and proliferation of neutrophil progenitors[200], in the context of triple-negative breast cancer (TNBC). TNBC has been demonstrated to promote the accumulation of neutrophils in the blood and immature neutrophils in the bone marrow[201,202]. These TFs demonstrate potential links between immune response regulation and granulopoiesis.

Euchromatic histone lysine methyltransferase 2 (EHMT2) is involved in gene silencing and repressive chromatin. EHMT1 and EHMT2-associated gene silencing of myeloperoxidase (MPO) and proteinase 3 (PRTN3), two proteins found in azurophilic granules in neutrophils, was lost in patients with anti-neutrophil cytoplasmic autoantibody (ANCA)-associated vasculitis (AAV)[203]. EHMT2 deletion in pancreatic epithelial cells affected the recruitment of immune cells, which included neutrophils[204]. The exact role of EHMT2 in neutrophil biology and differentiation remains undetermined. Similarly, histone deacetylase (HDAC) proteins regulate chromatin structure through modification of histones, consequently affecting gene expression[205]. Enhancer of zeste homolog 2 (EZH2) is a subunit of the polycomb repressive complex 2 (PRC2) that acts as a methyltransferase, again, modifying histones and chromatin organisation[206]. Interestingly, EZH2 was also found to regulate integrin signalling and adhesion dynamics in neutrophils, demonstrating a non-histone methylation role in controlling neutrophil behaviour[207]. These proteins may play general or specific roles in chromatin remodelling to enable neutrophil transcription programs.

E2F transcription factor 4 (E2F4) regulates the cell cycle. Its expression was shown to negatively correlate with neutrophil infiltration in human gastric carcinoma tumour tissue[208]. P53 participates in the DNA-damage response and cell survival[209]. SP1 influences cell proliferation, apoptosis, and differentiation[210]. These TFs may perform more general roles in transcriptional regulation, housekeeping gene expression, and control of cellular state. The potential role of other identified TFs in neutrophil biology and differentiation remains unclear and warrants further investigation.

Protein abundance can be partially linked to mRNA levels and PTM regulation. The data indicates several signalling pathways may be regulated at different stages, from gene expression to protein translation and beyond. However, other information needs to be considered. Importantly, the effects of post-transcriptional regulation and protein translation (synthesis) are other stages that can influence protein abundance. This information is currently missing in from the datasets. Still, it would provide valuable insights into how genes are expressed, how proteins are synthesised, and how proteins are subsequently regulated, leading to the identification of key regulatory mechanisms as well as potential targets that may be crucial for neutrophil differentiation. Analyses in this work have

revealed indications that ribosomal pathways and mRNA processing (e.g. splicing) proteins may indeed participate in bringing about appropriate molecular changes during neutrophil development.

Hoogendijk *et al.* performed correlation network analysis on transcriptome and MS-based proteome data from cells corresponding to the different stages of human, primary (pro)myelocyte cell differentiation towards the neutrophil lineage[211]. Consistent with these data, transcriptome and proteome changes from HoxB8 -omics showed decreasing levels of mitochondria- and ribosome-associated proteins, which also demonstrated high correlation with mRNA levels. Similarly, neutrophil effector proteins, such as those involved in cell migration, adhesion, and cytoskeleton-related functions, showed marked increase in expression, on both the proteome and transcriptome levels. Hoogendijk *et al.* raise valid speculations regarding discrepancies between protein and mRNA levels. Protein localisation, such as compartmental storage or complex formation, could potentially prevent the degradation of certain proteins, which exhibit stable abundances despite declining mRNA levels, thereby rendering them inaccessible to degradation. Nevertheless, these data recapitulate the observed changes in human neutrophil differentiation, reinforcing the robustness of the HoxB8 model. The complementarity of these datasets will enable the identification of commonalities between human and mouse neutrophils, as well as distinct aspects that may be unique to each organism.

Using linear regression to predict effects of various -omics variables on molecular outcomes only integrates specific levels of data. There are biological links between the genome, transcriptome, proteome, and PTMome. Therefore, using approaches that effectively integrate all sources of biological information may provide a more wholistic view and reveal more interesting phenomena[212]. In addition, supplementary information from external databases, such as protein-protein interaction data, may augment bioinformatic analysis. The development of models, such as gene regulatory networks and similar approaches, could help identify valuable links between the different -omics modalities. For example, phosphorylation data could be used to possibly indicate the activation status of proteins, such as TFs, which may affect associations between regulatory and regulated genes. Nevertheless, understanding the temporal dynamics of complex biological processes presents different challenges. It may be the case that the integration of different, longitudinal -omics datasets requires allowances for asynchronous changes and effects.

5.3.2. NLRP3 inflammasome

The NLRP3 inflammasome is a protein complex that regulates IL1β, IL18, and Gasdermin D (GSDMD) in response to diverse stimuli. It is typically associated with macrophages and monocytes[213]. However, emerging evidence demonstrates a new appreciation for the role of the NLRP3 inflammasome in neutrophil biology. In addition to cytokine production and secretion, the NLRP3 inflammasome has

also been shown to affect key neutrophil effector functions such as migration and NETosis[214-216]. NLRP3 inflammasome assembly and activation typically occur in two stages. Signal 1 leads to priming, causing increased expression of NLRP3 inflammasome components, and signal 2 determines activation, inducing cleavage of pro-IL1β and/or pro-IL18 and secretion of mature cytokine proteins[217]. Additionally, PTMs such as ubiquitination contribute to the assembly of the NLRP3 inflammasome complex[217].

Intriguingly, components of the NLRP3 inflammasome appear to be upregulated as HoxB8 neutrophils mature. The data suggests that HoxB8 neutrophils may undergo priming as part of their differentiation process. It is unlikely that priming signals were present in day 5 HoxB8 neutrophils since no PAMPs (e.g. LPS) or pro-inflammatory cytokines were introduced. Nevertheless, GSDMD, an effector molecule of the NLRP3 inflammasome and key mediator of pyroptosis, showed upregulation in mature neutrophils. Accordingly, neutrophils have been shown to resist the progression of GSDMD-dependent lysis during pyroptosis, maintaining cell viability[218]. In contrast, IL1β and IL18 were only detected on the mRNA level, possibly indicating secretion or degradation of these cytokines.

Moreover, many ubiquitinated peptides derived from NLRP3 inflammasome components showed increasing levels. However, the biological effects of these ubiquitination events during neutrophil differentiation is yet to be determined. Further investigations are required to better understand NLRP3 inflammasome and effector molecule behaviours within the context of neutrophil development. It is possible that, upon maturation, neutrophils act as a reservoir of NLRP3 inflammasome components, allowing for the rapid production and release of pro-inflammatory cytokines in direct response to pathogenic threats. It would be pertinent to decipher if NLRP3 inflammasome complexes are assembling, if they are a feature or necessity of differentiation, and if signalling pathways associated with inflammasome signals are activated.

5.3.3. Novel identified proteins may be markers of mature neutrophils

CD11b is a cell surface marker of myeloid cells, contributing to functions such as adhesion and migration[219]. CD11b shows increased expression in specialised cells compared to progenitor cells. In contrast, CD34 is a marker of cell stemness, and its expression decreases as cells transition from myeloblasts to mature neutrophils[220]. Expression changes in these surface proteins are characteristic of neutrophils undergoing differentiation.

The FCER1G is involved in neutrophil activation[221,222]. Upregulation in mature HoxB8 neutrophils is consistent with its role in neutrophil effector functions.

RHOA is a small GTPase that participates in cell migration, adhesion, division, and survival[223,224]. RHOA contributes to neutrophil response regulation and effector functions, such as the recruitment to inflamed tissue[225,226]. Changes in protein abundance suggest RHOA, and potentially other small GTPases, may play a role during neutrophil differentiation.

SIRT3 is a mitochondrial protein that participates in cellular energy metabolism and ROS production. SIRT3 KO mice showed increased ROS production in their neutrophils compared to WT mice, with no observed changes in NETosis[227]. Additionally, SIRT3 was demonstrated to control neutrophil mitochondria and autophagy in response to pathogens[228]. SIRT3 levels in day 3 and day 5 HoxB8 neutrophils were possibly too low to detect using MS-based proteomics. Interestingly, SIRT3 expression is higher in progenitor cells. SIRT3 has been shown to be important in myogenesis, with higher expression in myoblast progenitor cells, regulating terminal differentiation[229]. Mitochondrial function has been shown to be an aspect of neutrophil differentiation in the literature[230-233]. How SIRT3 contributes to neutrophil differentiation is yet to be explored.

TGM2 is an enzyme involved in protein cross linking. It has been shown to increase in expression during retinoic acid-induced differentiation of promyelocytic NB4 leukaemia cells into granulocytes. Partial translocation of TGM2 to the nucleus was also observed. TGM2 KO in mice showed altered capabilities of certain neutrophil effector functions[234], such as increased cytokine release[235]. The interesting behaviours of TGM2 warrants further research into the potential mechanistic role TGM2 plays during neutrophil differentiation.

Chapter 6: Exploring DUBs in neutrophil differentiation using an HL60 CRISPR phenotypic screen

6.1. Introduction

Inspired by the bacterial defence system against viral attack, CRISPR technology is able to target specific DNA sequences within the genome and modify genes[236]. The CRISPR-Cas9 system is typically used to create double stranded DNA breaks at regions of interest, disrupting gene structure by introducing mutations as a result of DNA repair mechanisms. In humans, DNA is typically 'fixed' through non-homologous end joining (NHEJ) repair[237], potentially resulting in frameshift mutations that lead to NMD of resultant gene transcripts and preventing translation of the corresponding protein. Some organisms can perform homology-directed repair (HDR), and this can be exploited to endogenously introduce specific, targeted mutations by providing an appropriate template[238]. CRISPR–Cas9 can be used to generate gene knockouts by introducing double-strand breaks at the target locus. In the absence of a repair template, breaks are primarily repaired by the error-prone NHEJ pathway, which frequently introduces insertions or deletions that disrupt the reading frame and result in the loss of gene function.

Phenotypic screening is a tool that can identify phenotypic changes that occur due to genetic[239] and/or chemical perturbation[240] without the need for mechanistic insight. This top-down approach can accelerate the selection of functionally-relevant genetic targets or chemical agents that can then be investigated in a more targeted fashion to understand the underlying mechanisms. This approach can ultimately lead to discoveries that may have useful applications, such as the development of therapeutic interventions.

DUBs are key regulators of protein ubiquitination status. To understand their role in neutrophil biology, a phenotypic screening approach was established, optimised, and performed. The nuclei of mature neutrophils are distinct from those of progenitor cells. Upon induction of differentiation, the nucleus becomes lobed and segmented, a well-defined morphological feature. Using an HL60 cell model of neutrophil differentiation, genes encoding the various DUBs were knocked out using CRISPR-Cas9 technology in an arrayed format. Then, the ability of these HL60-DUB KO cells to differentiate into NLCs (dHL60 cells) was assessed using microscopy imaging to visualise the nuclear morphology. Statistical analysis of the image data was used to identify candidate DUBs that appeared to affect the HL60 differentiation process.

6.2. Results

An overview of the experimental workflow for the DUB KO HL60 differentiation screen is depicted in Figure 6.1. Several steps required optimisation and will be discussed in subsequent sections within this chapter. In brief, a two-plasmid system was used to generate HL60-DUB KO cells. First, an HL60-Cas9 cell line was generated via lentiviral transduction. Next, sgRNA plasmids were isolated and packaged into lentivirus particles. Then, these sgRNA lentivirus particles were used to transduce HL60-Cas9 cells to generate HL60-Cas9-DUB KO cell lines. Finally, HL60-Cas9-DUB KO cells were differentiated and analysed using high-throughput microscopy.

Regarding the lentiviral transduction, virus titration was not performed to reduce time and cost, as well as to preserve lentivirus quantities. Instead, Western blot analysis was used to detect molecular changes resulting from transduction with lentivirus particles of interest.

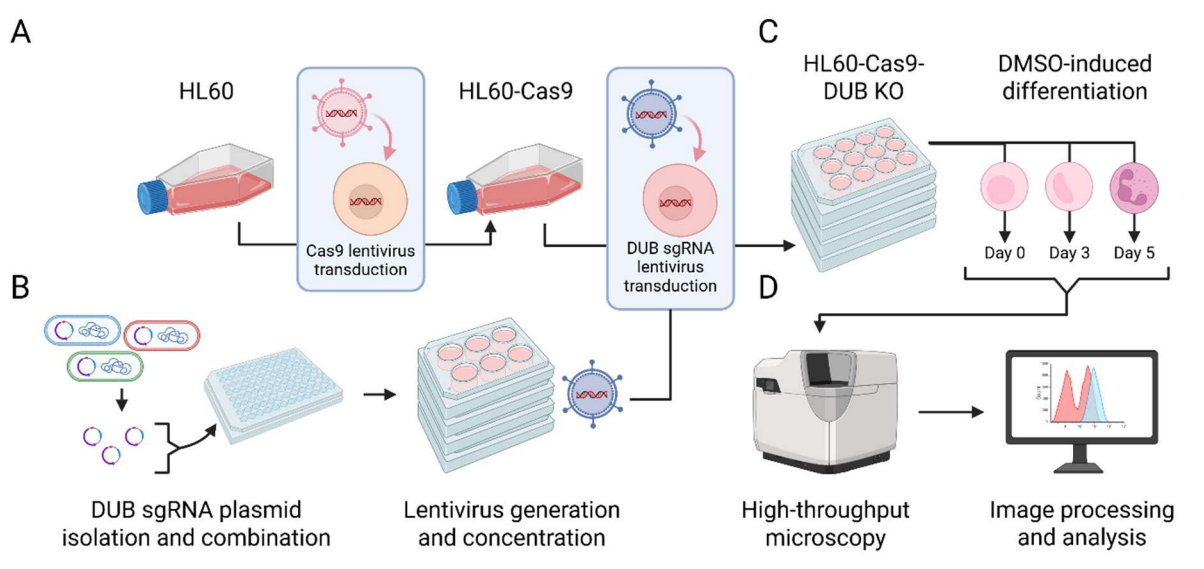


Figure 6.1 ***Experimental workflow for DUB KO HL60 differentiation screen.*** (**A**) HL60-Cas9 cell lines were generated via lentiviral transduction. (**B**) Plasmids encoding individual DUB sgRNAs were isolated, quantified, and combined (minimum 3 sgRNAs for each DUB) for lentivirus particle generation. Lentivirus particles were concentrated and used to transduce HL60-Cas9 cells, generating HL60-Cas9-DUB KO cells in arrays format. (**C**) DUB KO cells were then differentiated using DMSO, samples were collected on day 0, 3, and 5, post-induction, and fixed using paraformaldehyde. (**D**) Nuclei and cell membranes were stained, and cells were imaged on a Celldiscoverer7. Images were processed and analysed using computational approaches. This figure was created using BioRender.

6.2.1. Validation of HL60 differentiation

HL60 cells demonstrate promyelocytic morphology. Various protocols have been established to differentiate HL60 cells into neutrophil-like cells (NLCs). These differentiated HL60 (dHL60) cells exhibit morphological and functional similarity to that of physiological neutrophils. HL60 cells were differentiated using DMSO, and samples were collected at daily intervals for 5 or 7 days. The samples were either fixed using PFA, for microscopy and flow cytometry, or pelleted and lysed for Western blot analysis.

An increase in expression of CD11b, a key myeloid differentiation marker, was observed following DMSO-induced differentiation of HL60 cells (Figure 6.2A). CD11b expression appeared to plateau at day 5 post-induction. Microscopy analysis demonstrated that a certain population of HL60 cells was undergoing differentiation, as indicated by the differences in nuclear morphology; dHL60 cells consisted of segmented, multi-lobed nuclei (Figure 6.2B). Flow cytometry analysis confirmed this, with approximately 30% of dHL60 cells showing increased expression of CD11b compared to HL60 cells and isotype control (Figure 6.3A-C). Median fluorescence intensity (MFI) increased with CD11b labelling between day 0 and 5, indicating increased expression of CD11b on dHL60 CSMs (Figure 6.3C). There was also increased labelling of CXCR4 and decreased labelling of CXCR2, in contrast to primary neutrophils, which typically downregulate CXCR4 and upregulate CXCR2 expression upon maturation[241]. Nevertheless, the data showed HL60 cells were capable of being differentiated using DMSO.

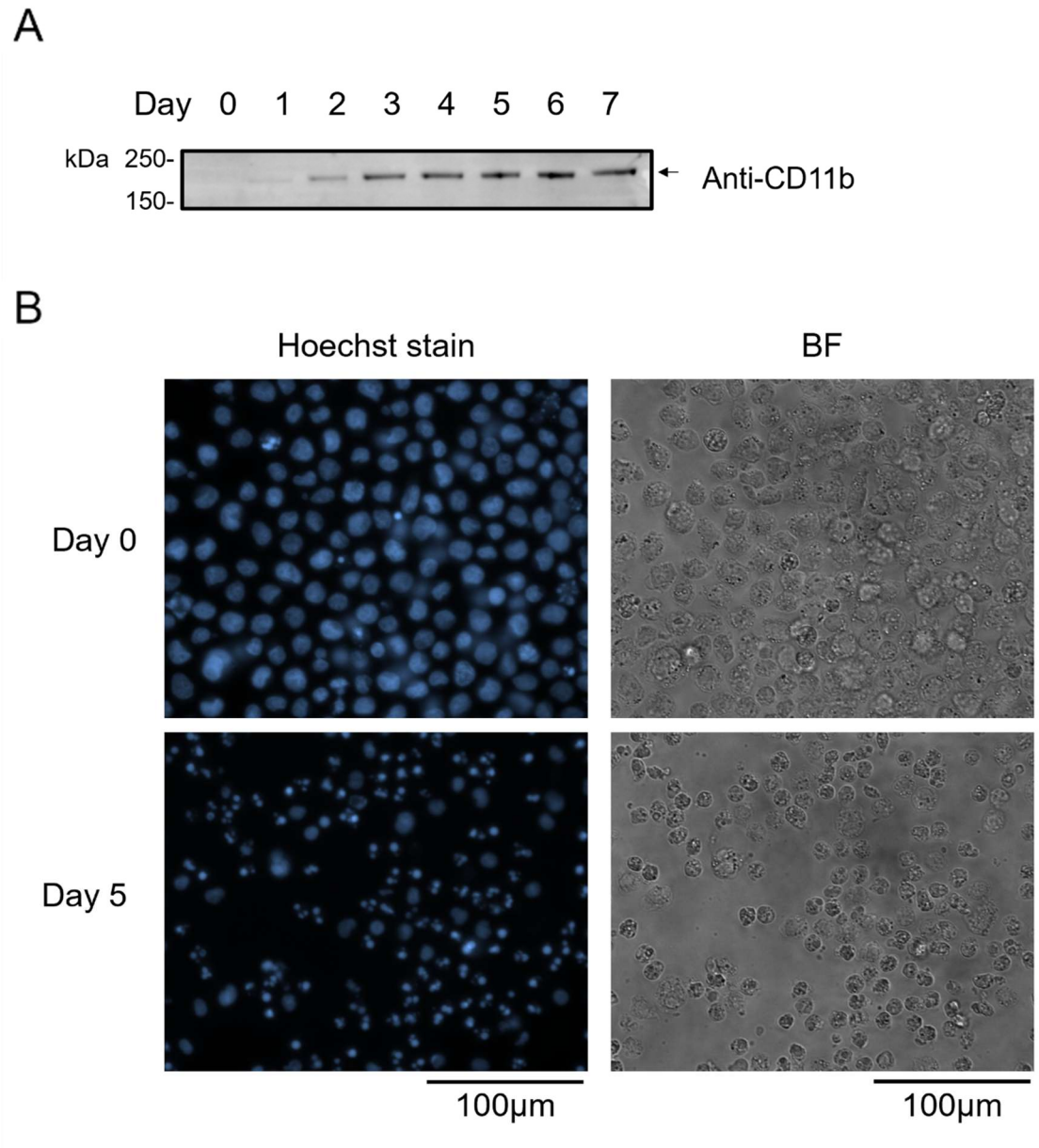


Figure 6.2 ***Characterisation of DMSO-induced HL60 differentiation.*** (**A**) Whole cell lysates from HL60 cells were collected at daily intervals for 7 days and analysed using Western blot. (**B**) Fixed cells from day 0 and da5 post-induction were collected and fixed (n = 1). Fixed cells were stained with Hoescht 33342 (nuclear stain). BF = brightfield.

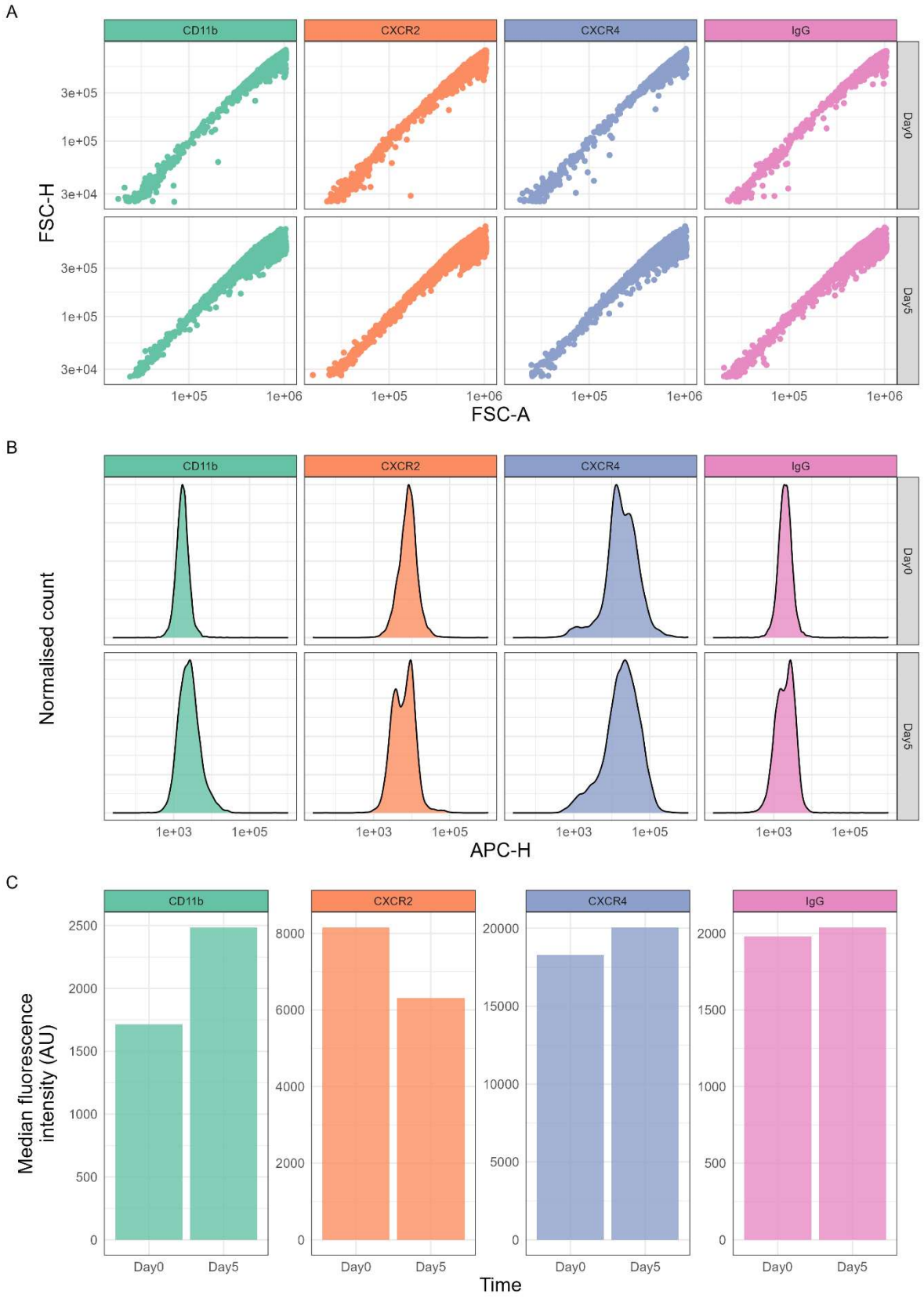


Figure 6.3 ***Flow cytometry analysis of DMSO-induce HL60 differentiation.*** HL60 cells were collected and fixed on day 0 and day 5 post-induction of differentiation (n = 1). Fixed cells were labelled with anti-CD11b, -CXCR2, -CXCR4, and -IgG (isotype control) antibodies conjugated to allophycocyanin (APC). (**A**) Dot plot of FSC-H and FSC-A to identify single cells. Upper panels are day 0, and lower panels are day 5. (**B**) Density plots of labelling, showing normalised counts of APC-H intensities. Upper panels are day 0, and lower panels are day 5. (**C**) Median fluorescence intensities of each antibody labelling and time point.

6.2.2. Generation of the HL60-Cas9 cell line

In order to perform CRISPR-Cas9-based gene KO of the DUBs, a two-plasmid system was employed. Cas9 is first introduced on one plasmid, followed by lentiviral transduction of the sgRNA-expressing plasmid. This contrasts with a one-plasmid system, where both components are delivered to the cell together.

An HL60-Cas9 overexpression cell line was first generated by introducing a Cas9-overexpression plasmid via lentiviral transduction. Two different quantities of lentivirus particles were used to transduce Cas9-encoding plasmids into HL60 cells. 'Low' and 'high' virus titre corresponded to smaller (10 μL) and larger (30 μL) volumes of concentrated lentivirus particles, respectively. Expression of Cas9 protein was confirmed using Western blot analysis. The ability to generate a KO cell line was also tested using USP15 sgRNAs (Figure 6.4). Additionally, the capacity to differentiate the HL60-Cas9 cell line was confirmed (Figure 6.5). Successful USP15 KO and the ability to upregulate CD11b upon DMSO-induced differentiation in the HL60-Cas9 cell lines were observed with both low and high lentivirus titres.

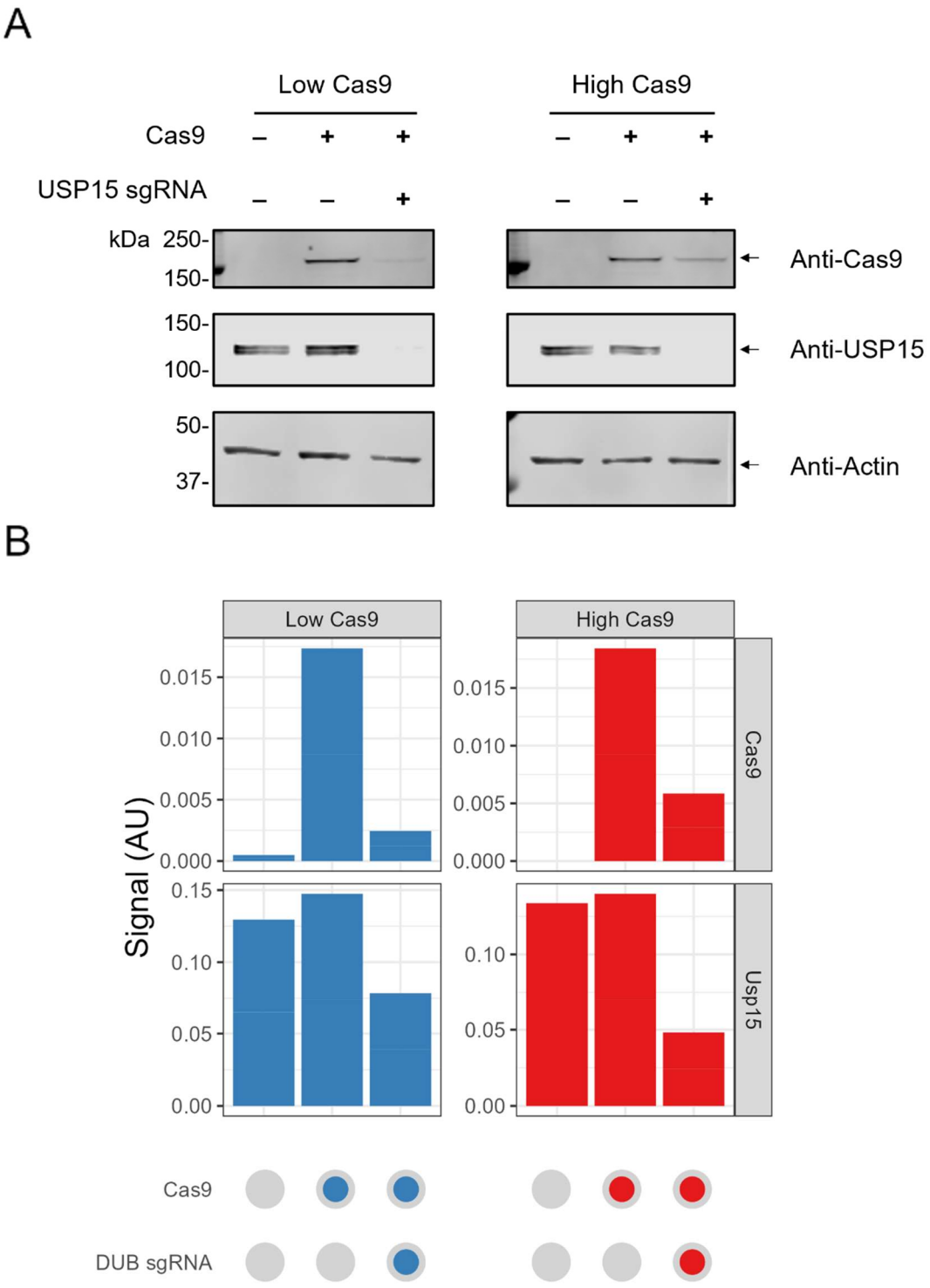


Figure 6.4 ***Generation of HL60-Cas9 cell line.*** (**A**) Western blot validation of Cas9 expression and ability to KO USP15. Actin loading control was included. (**B**) Bar charts representing Western blot signals normalised against actin loading control signals (n = 1).

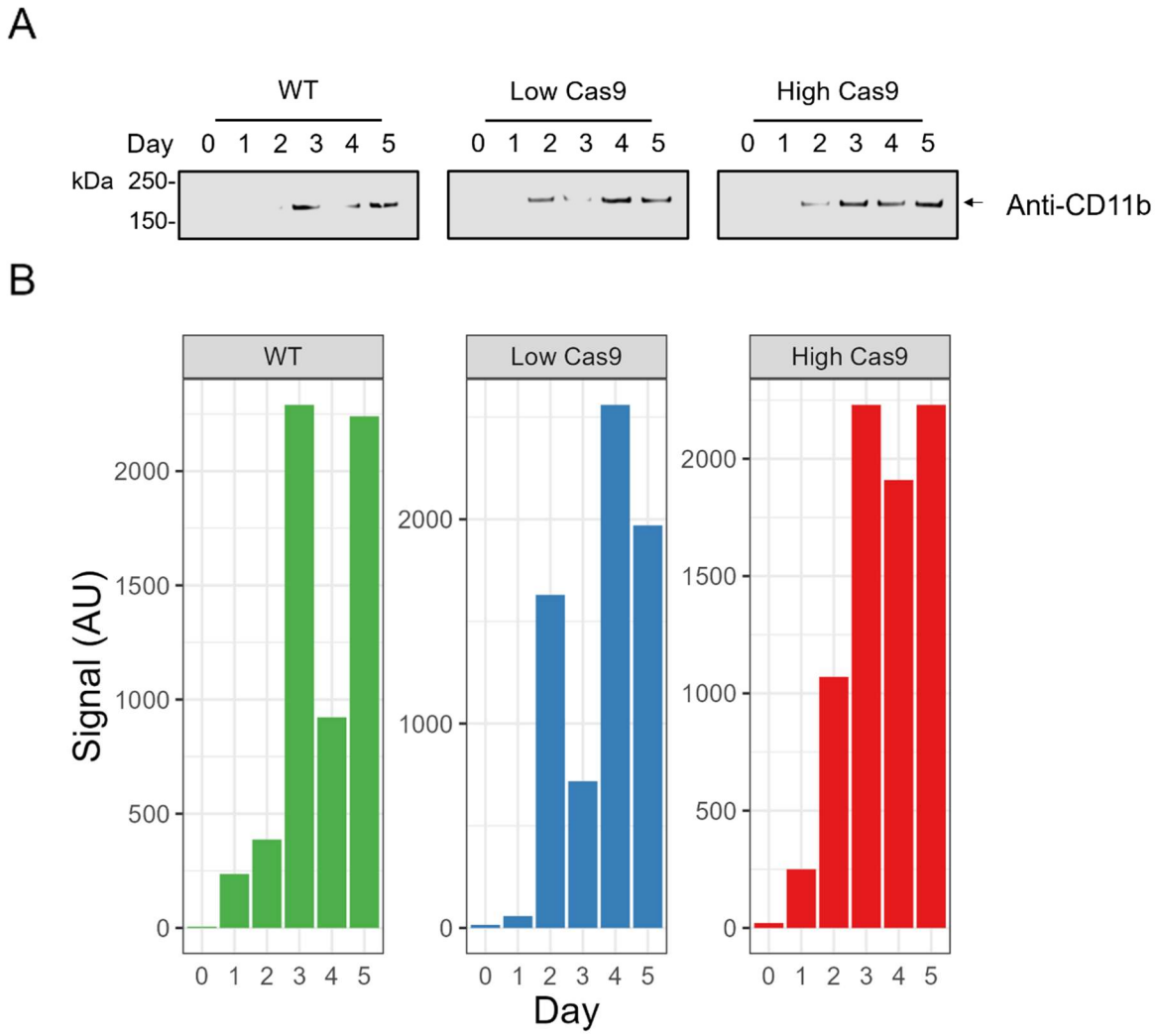


Figure 6.5 ***DMSO-induced differentiation of HL60 WT and HL60-Cas9 cell lines.*** (**A**) Western blot analysis, using anti-CD11b antibody as a differentiation marker. Loading amounts were normalised using protein quantification assay. (**B**) Western blot signals of anti-CD11b (n = 1).

6.2.3. Generating sgRNA lentivirus particles and HL60-Cas9-DUB KO cell lines

The lentivirus generation protocol was optimised for high throughput applications to generate HL60-Cas9 cell lines with individual DUB KOs on the population level in an arrayed format. Plasmids encoding individual sgRNAs targeting 94 different DUBs were kindly provided by Dr. Vincenzo D'Angiolella (Institute of Genetics and Cancer, University of Edinburgh).

Plasmids encoding sgRNAs were isolated, quantified, and combined with other sgRNAs that corresponded to the same gene. Combined sgRNA plasmids were then packaged into lentivirus particles and used to transduced HL60-Cas9 cells in an arrayed format. Transduced cells were cultured in 'recovery medium' to help reduce oxidative stress. Selection of cells involved treatment with blasticidin (Cas9 plasmid) and puromycin (sgRNA plasmid). USP7-, USP10-, and USP15-KO cell lines were verified using Western blot analysis (Figure 6.6). Each DUB exhibited varying levels of protein,

corresponding to the efficiency of gene KO. Nevertheless, the transduction of sgRNAs was successful, and KO cell lines were ready for the differentiation screen.

6.2.4. HL60-Cas9-KO differentiation CRISPR screen

HL60-Cas9-DUB KO cell lines were counted using an NxT Attune flow cytometer, and an equal number of cells were manually dispensed into a 96-well plate. Culture media was supplemented with DMSO to induce differentiation. Cells were collected and fixed using PFA on day 0, 3, and 5 post-induction and stored at 4°C in PBS. Fixed cells were stained using WGA-AF647 (cell membrane stain) and Hoescht 33342 (DNA stain). Stained cells were imaged using a ZEISS Celldiscoverer 7 in wide-field mode, capturing 31 Z-stacks at least six times (i.e. six positions, or scenes) for each well.

Z-stacks for each position were deconvoluted and max intensity projected to achieve a single image per scene, per channel. Image analysis was performed on all scenes using CellProfiler, where the WGA-AF647 channel (cell membrane) was used to identify objects. These objects were then filtered based on roundness and separation from other objects to segment and identify round, distinct cells. The Hoescht 33342 channel (DNA) was masked using the filtered objects (round, distinct cells), and objects were identified that corresponded to the nuclei within each cell. Due to the lobed nature of dHL60 cells, multiple nuclear objects were identified per cell. Therefore, these objects were merged to determine the overall shape of the nucleus. Several features of the nuclei were measured, but analysis primarily focused on the ‘form factor’ measure, a metric that incorporates area and perimeter of an object to calculate ‘roundness’. Form factor measurements for each nucleus from all scenes corresponding to a specific well, i.e., a DUB KO cell line, were combined to allow comparative analysis. See Figure 6.7 for an overview of the image analysis workflow.

Analysis of form factor measurements for WT dHL60 cells revealed a bimodal distribution of values, in contrast to their undifferentiated (HL60) cell counterparts. This suggests that there is an expected proportion of the cells that have differentiated and thus achieved a multi-lobed, segmented nucleus. As such, a form factor threshold was determined to distinguish HL60 and dHL60 populations. An expectation-maximisation algorithm for mixed normal distributions was used to estimate the parameters of the two underlying normal distributions (Figure 6.8A), corresponding to dHL60 (component A) and HL60 cells (component B).

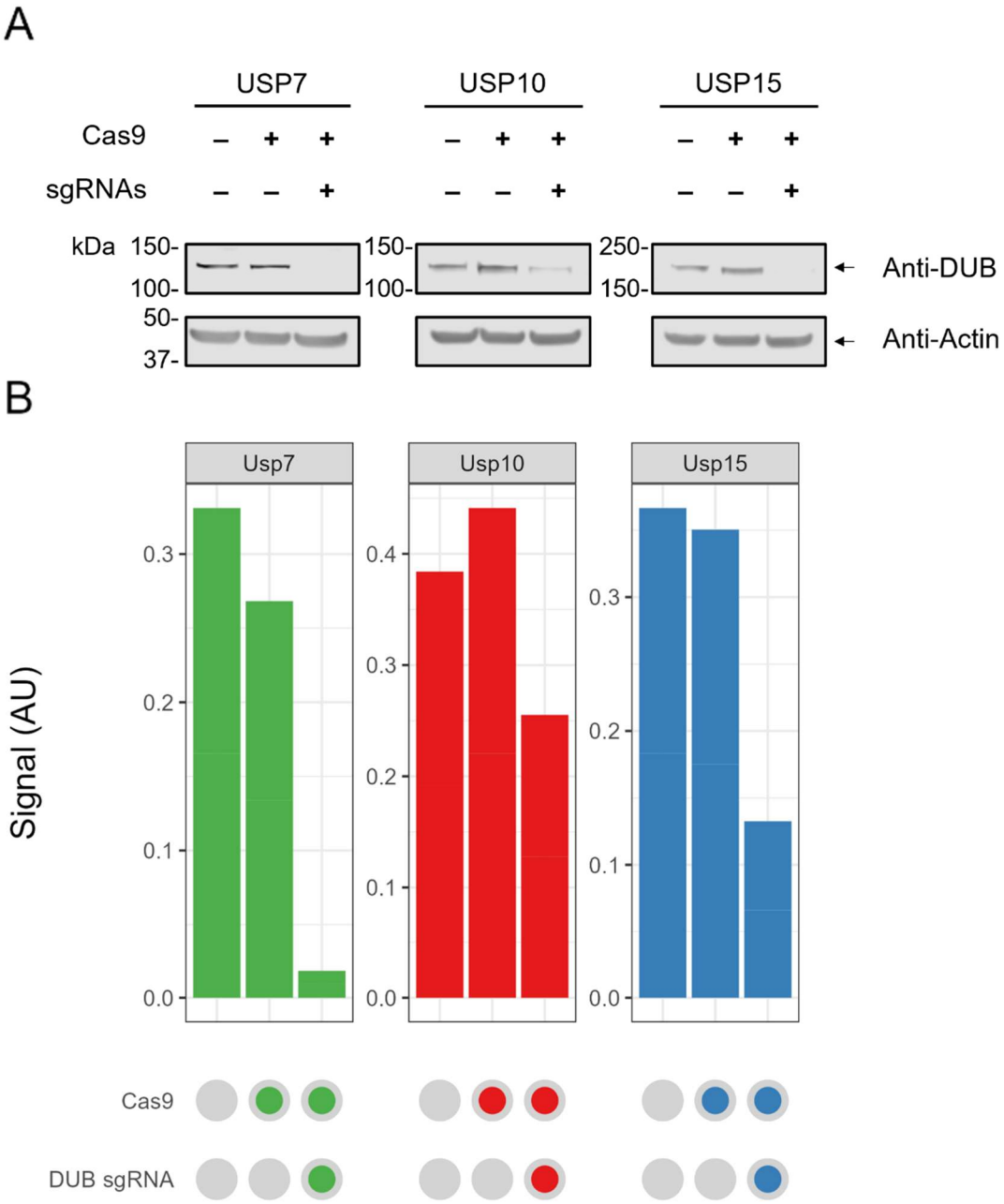


Figure 6.6 ***Verification of DUB KO in select samples from HL60-Cas9 cells transduced with sgRNA lentivirus particles.*** (**A**) Western blot analysis of USP7-, USP-10-, and USP15-KO cell lines. (**B**) Quantitation of Western blot signals normalised against actin loading control signal (n = 1).

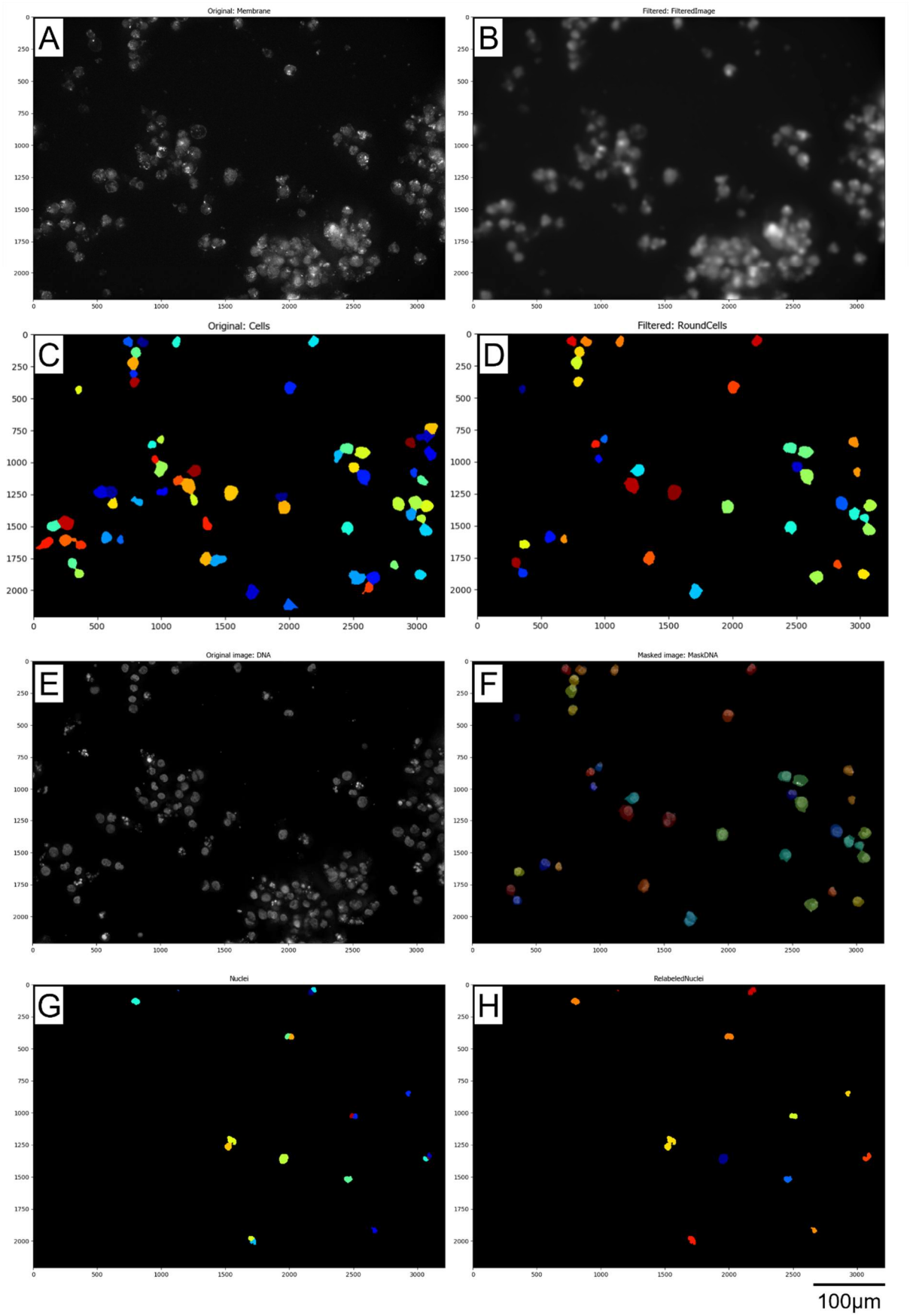


Figure 6.7 ***Image analysis workflow for high-throughput microscopy data from HL60 CRISPR-Cas9 DUB KO screen.*** An example workflow is shown. (**A**) Original image of deconvoluted, max intensity-projected membrane channel (WGA-AF647 stain). (**B**) Gaussian blur of membrane channel. (**C**) Object identificaiton (thresholding and segmentation included) (**D**) Object filtering, retaining circular and 'declumped' objects, corresponding to round, distinct cells. (**E**) Original image of deconvoluted, max intensity-projected DNA (nucleus) channel (Hoechst 33342 stain). (**F**) Filtered membrane objects (cells) were used to mask DNA channel. (**G**) Nuclear objects were identified from masked DNA image. (**H**) Neighbouring nuclear objects were merged to create signle object per cell. Measurements of these objects were then caclulated and exported for analysis.

The probability density functions of the two estimated normal distributions were then calculated and used to determine the probability that a data point belonged to the dHL60 population. By setting this probability to 0.8, a form factor threshold of ~0.55 was assumed to separate the dHL60 and HL60 cell populations (Figure 6.8B). This threshold was applied to the form factor distributions of each DUB KO to calculate the proportion of dHL60 cells at day 0 (Figure 6.9), day 3 (Figure 6.10), and day 5 (Figure 6.11) post-induction of differentiation. Form factor distributions were also compared using the Kolmogorov–Smirnov test, using the Benjamini-Hochberg (BH) procedure for multiple testing FDR control (Figure 6.12A, B).

Several DUB KOs in HL60 resulted in significantly distinct form factor distributions compared to WT HL60 cells on days 3 and 5 post-induction of differentiation (FDR < 0.05). In particular, YOD1, USP7, BRCA1 associated protein BAP1, COP9 signalosome (CSN) complex subunit 6 (COPS6), Valosin-containing protein interacting protein 1 (VCPIP1), and USP47 deletion showed significantly different distributions compared to WT and lower proportions of dHL60 in day 5 (Figure 6.13). In contrast, Ubiquitin thioesterase OTUB2 and USP44 deletion showed higher proportions of dHL60 and significantly distinct distributions of nuclei form factor readings, suggesting enhanced HL60 differentiation.

A DUB may be essential for cell viability, so homozygous KO of such genes would result in cell death. Heterozygous cells may be viable if a gene is haplosufficient. However, any phenotypic consequence of gene KO would require further validation. The Genome Aggregation Database (gnomAD) contains information on genetic variation and, notably, information on the probability of a gene being intolerant to loss-of-function (LoF) variants (pLI)[128]. A pLI greater than 0.9 indicates extreme intolerance of LoF variants, suggesting these genes are likely essential. Moreover, the DepMap project identified genes that, when perturbed, affected the cell viability of various cancer cell lines[129]. Data on these 'Achilles gene effects' relevant to HL60 cells were incorporated into this analysis, with more negative gene effects indicating a more detrimental consequence of gene interference. Finally, mRNA expression levels in HL60 cells for each DUB were obtained from the Human Protein Atlas[130,131], providing evidence for DUB protein expression.

Of the DUB candidates identified, YOD1, VCPIP1, OTUB2, and USP44 showed low pLI (**Error! Reference source not found.**C) and YOD1, USP7, BAP1, USP47, OTUB2, and USP44 demonstrated minimal/no Achilles gene effects (Figure 6.12D). These DUBs are likely to be affecting the HL60 differentiation process rather than general cellular homeostatic conditions. The mRNA expression levels of YOD1, VCPIP1, USP47, OTUB2, and USP44 were relatively low, while USP7, BAP1, and COPS6 showed higher mRNA expression levels (Figure 6.12E).

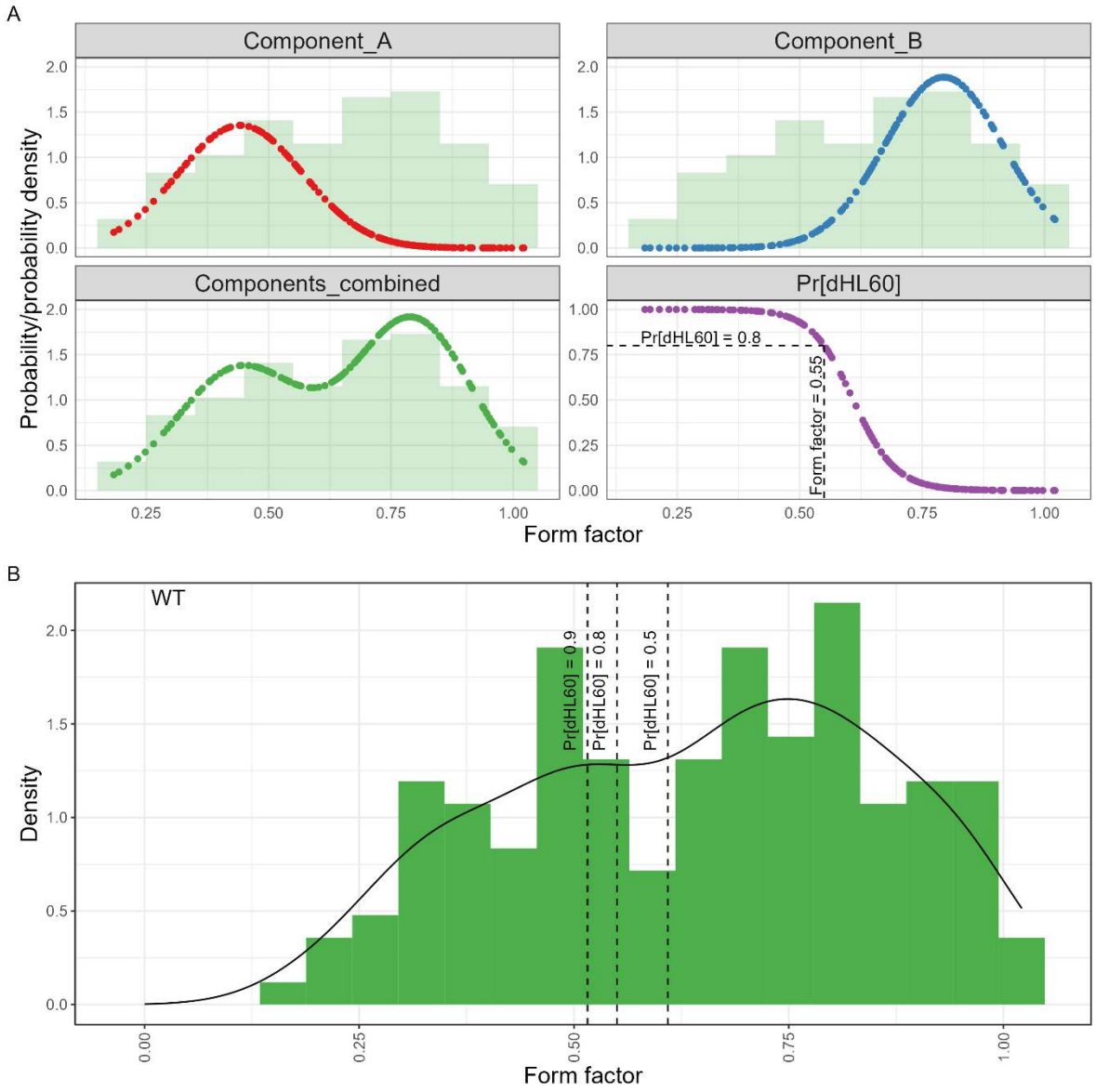


Figure 6.8 ***Determining HL60 and dHL60 populations based on distribution of form factor values.*** (**A**) Expectation-maximisation algorithm identified underlying normal distributions corresponding to dHL60 (component A) and HL60 (component B) populations. Ratio of probability density functions (PDF) of component A to combined components represents probability of value belonging to the dHL60 population (Pr[dHL60]). (**B**) Form factor cutoff values to separate HL60 and dHL60 populations based on probability thresholds calculated from PDFs.

Due to the limited number of cells identified during day 0 in HL60 WT and HL60-Cas9 samples, the proportions of dHL60 cells were assumed to be comparable to the lowest decile of all samples. Log_2FC of the proportion of dHL60 was calculated for days 3 and 5, each compared to day 0 (Figure 6.14). Several DUB KOs exhibited significantly different distributions of form factor values on day 3 of differentiation, compared to WT. However, many of these DUBs did not show significant differences during day 5. Only VCPIP1, OTUB2, and USP7 demonstrated statistical significance during days 3 and 5. Due to the fact that day 3 HL60 cells represent a distinct stage of HL60 differentiation, it may be necessary to consider different mathematical assumptions to capture biological and nuclear morphological information.

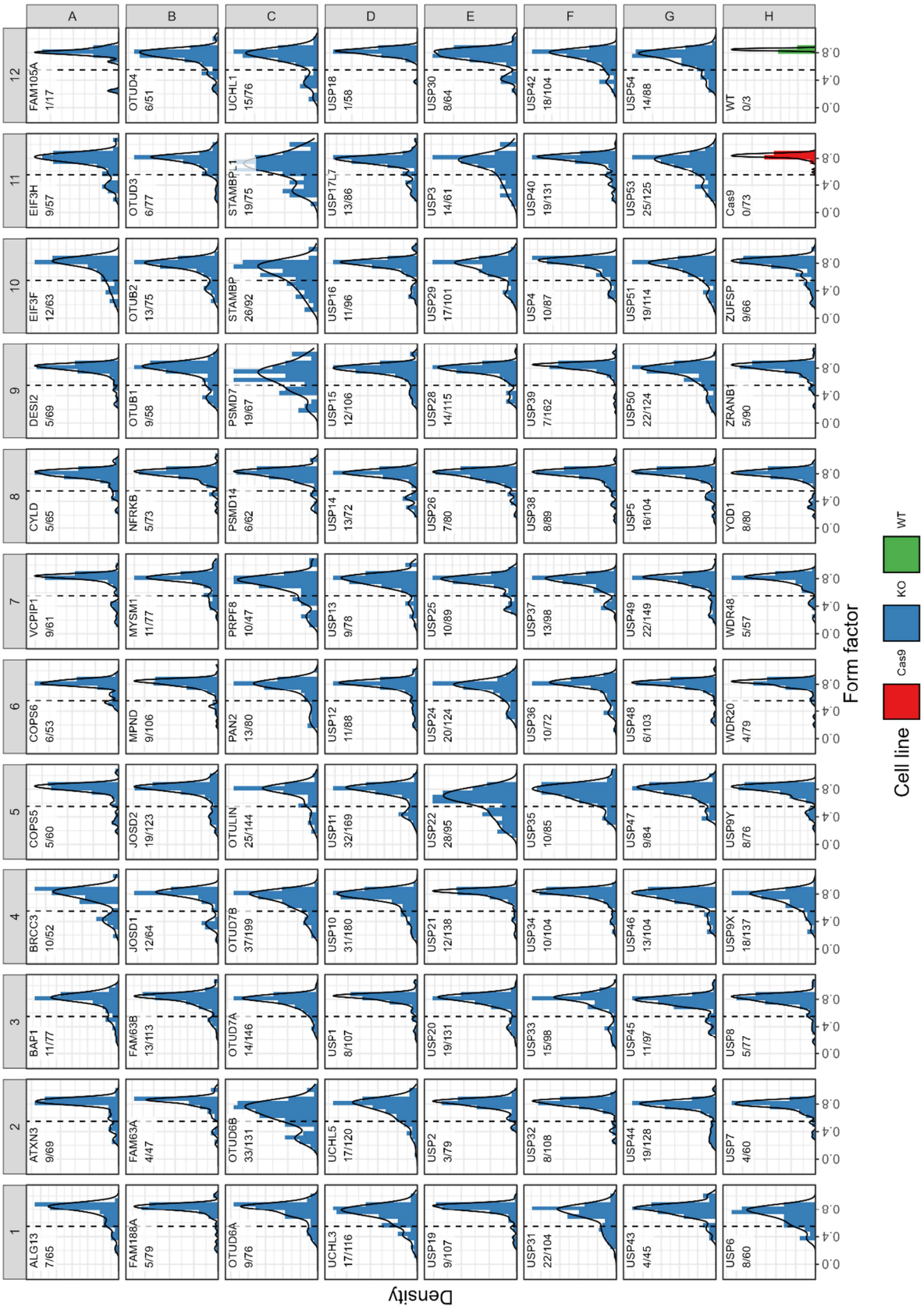


Figure 6.9 ***Form factor distributions of HL60 WT, HL60-Cas9, and HL60-Cas9-DUB KO cell lines on day 0 of DMSO-induced differentiation.*** Dotted line represents separation of HL60 and dHL60 populations, as determined in *Figure 6.8*.

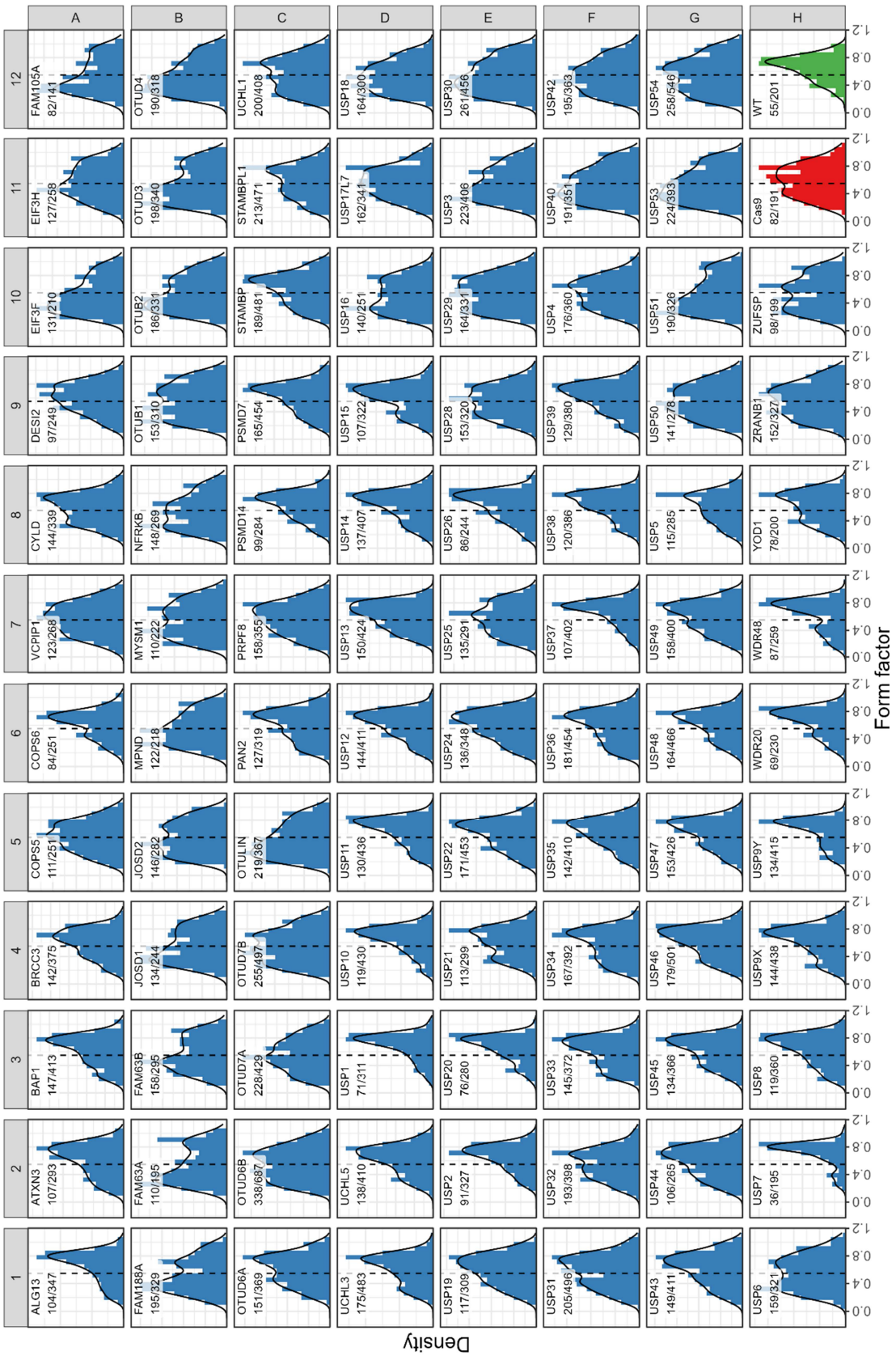


Figure 6.10 ***Form factor distributions of HL60 WT, HL60-Cas9, and HL60-Cas9-DUB KO cell lines on day 3 of DMSO-induced differentiation.*** Dotted line represents separation of HL60 and dHL60 populations, as determined in *Figure 6.8*.

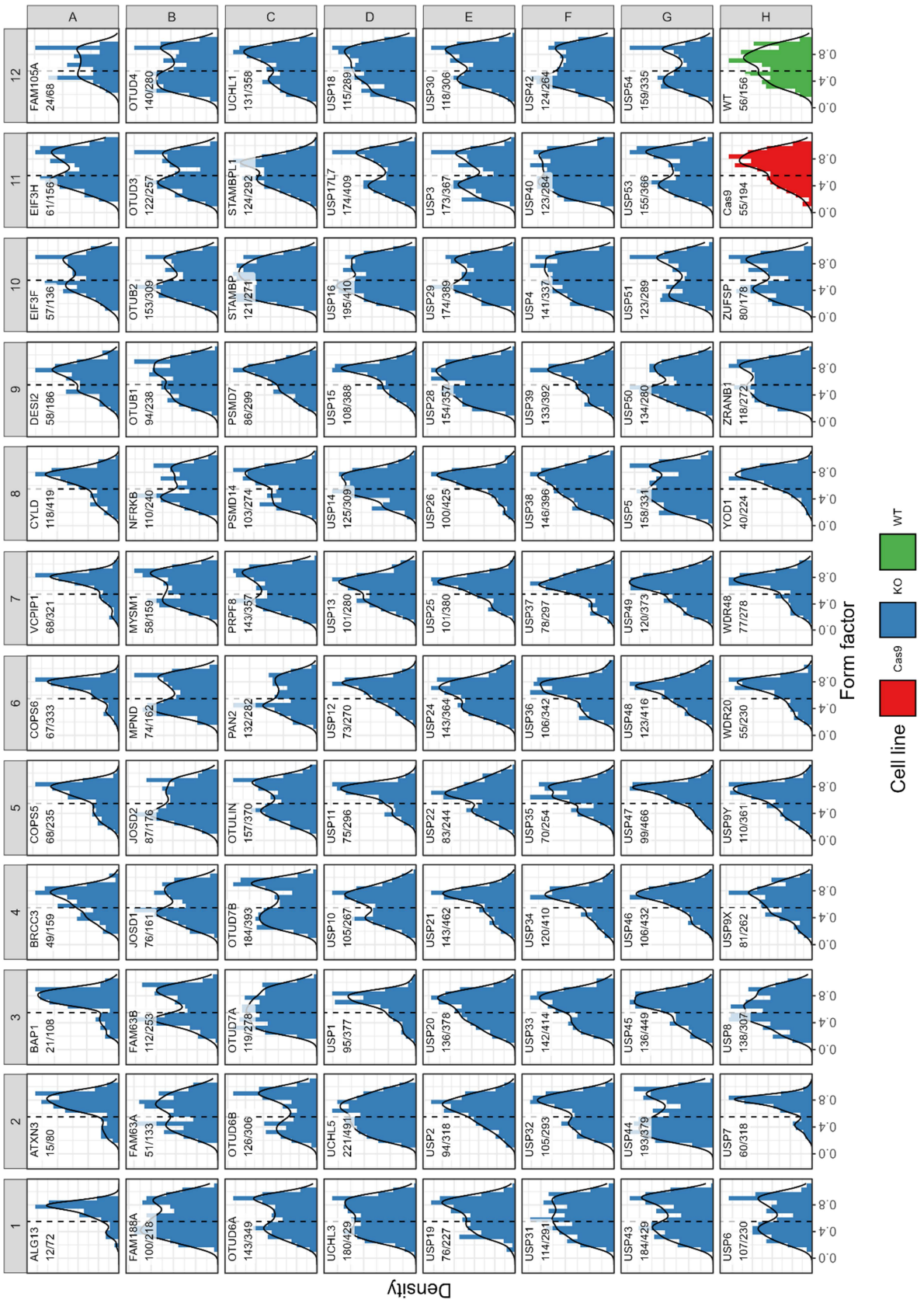


Figure 6.11 ***Form factor distributions of HL60 WT, HL60-Cas9, and HL60-Cas9-DUB KO cell lines on day 5 of DMSO-induced differentiation.*** Dotted line represents separation of HL60 and dHL60 populations, as determined in *Figure 6.8*.

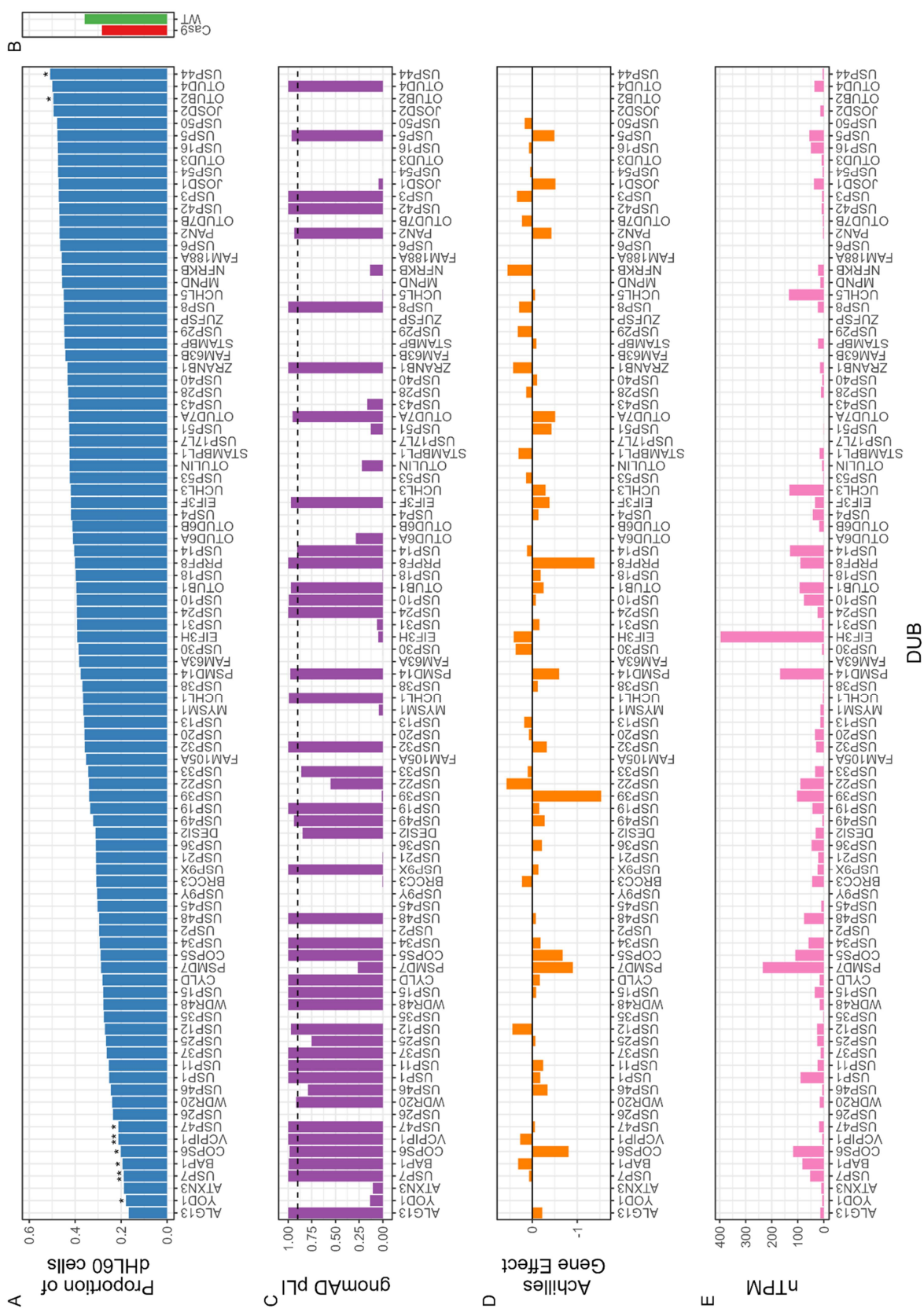


Figure 6.12 ***Proportion of dHL60 cells in HL60-Cas9-DUB KO samples on day 5 of DMSO-induced differentiation, with DUB essentiality and mRNA expression data.*** Proportions of dHL60 for each DUB KO (**A**) and reference HL60 WT and HL60-Cas9 proportions (**B**). FDR thresholds (n = 1) are indicated above each bar (*FDR < 0.05, **FDR < 0.01). (**C**) gnomAD probabilities of loss-of-function intolerance (pLI) for each DUB. pLI > 0.9 indicates a DUB is intolerant of loss-of-function variants. (**D**) Achilles gene effect of RNAi targeting each DUB in HL60 cells. More negative values indicate a DUB is more essential for HL60 viability.(**E**) mRNA levels of each DUB in HL60 cells. Data is from the Human Protein Atlas. nTPM = normalized transcripts per million.

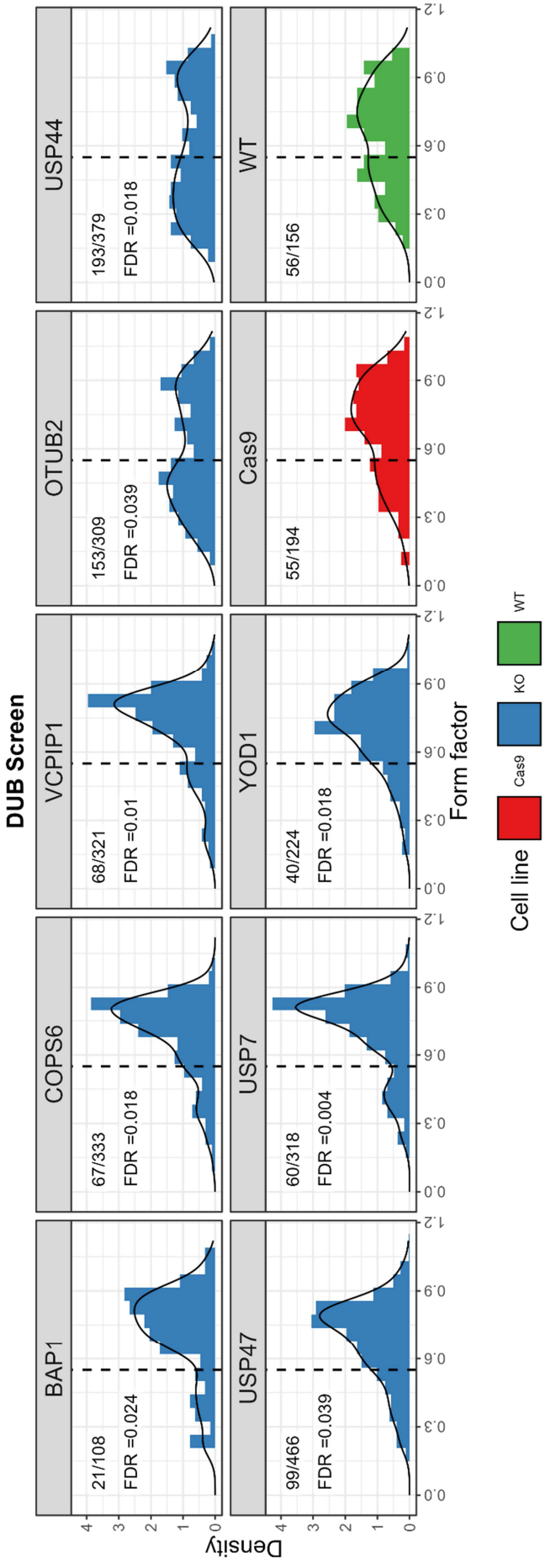


Figure 6.13 ***Significant hits from HL60 differentiation DUB KO screen.*** Distribution of form factor values of DUBs with significant values for Kolmogorov–Smirnov test, with FDR control (n = 1).

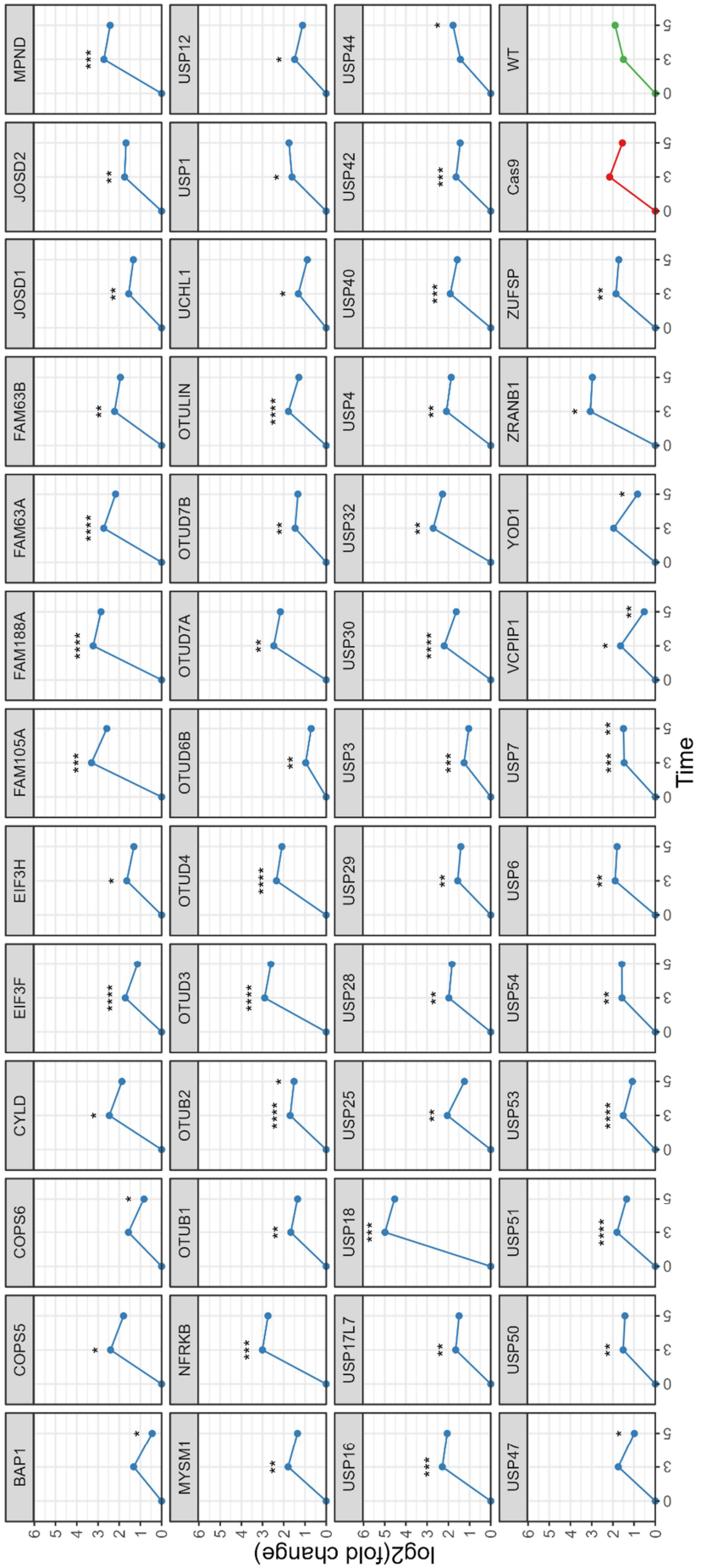


Figure 6.14 ***Fold changes in dHL60 proportions during DMSO-induce HL60 differentiation.*** Proportions of dHL60, based on form factor cut-off, were compared to the proportion of day 0 (HL60) for each cell line. DUB KO samples that showed statistically significant different distributions of form factor values compared to WT are shown above the corresponding time point (*FDR < 0.05, **FDR < 0.01, ***FDR < 0.001, ****FDR < 0.0001).

Western blot validation was performed on HL60 WT, HL60-Cas9, and HL60-COPS6, -USP26, -USP47, and -YOD1 KO cell lines (Figure 6.15). HL60-DUB KO cell lines all showed decreased CD11b signal compared to HL60 WT and HL60-Cas9 cell lines in day-5 dHL60 cells. Day-0 cells appeared to show similar levels of CD11b, suggesting DUB KO does not affect CD11b expression in promyelocytes. From the initial screen data, the USP26 KO cell form factor distribution comparison did not reach statistical significance when compared to WT cells but still showed reduced proportion of dHL60 cells. Nevertheless, Western blot analysis reveals there is indeed a reduction in CD11b expression in day 5 dHL60-USP26 KO cells.

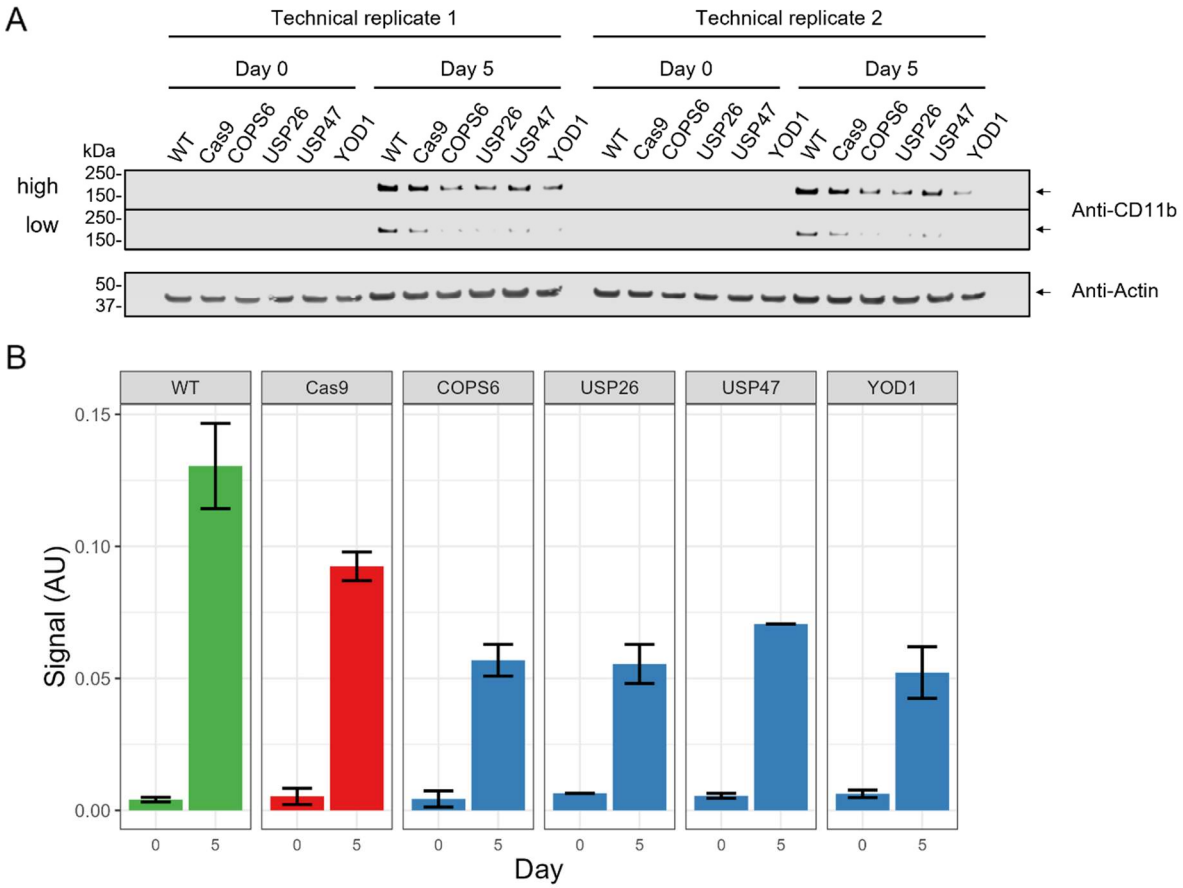


Figure 6.15 ***Western blot validation of CRISPR-Cas9 DUB KO HL60 differentiation screen.*** (**A**) Western blot analysis was performed twice on the same samples (n = 2 technical replicates). Cell lines are indicated above the blot, with DUB names corresponding to their KO cell lines. WT = wildtype. (**B**) Relative signals of CD11b. Anti-CD11b signals were normalised against corresponding anti-Actin signal. Mean signals are shown, and error bars represent the standard deviation.

6.3. Discussion

CRISPR-Cas9 phenotypic screening identified several DUBs that appear to either interfere with or promote HL60 differentiation. This suggests potential roles for these DUBs in regulating neutrophil differentiation, which is relevant to both physiological and pathological contexts involving neutrophil development. At this stage, it remains unclear whether these effects reflect a direct role of the identified DUBs in the differentiation process or whether they arise from indirect consequences such as secondary effects or general cellular fitness defects.

YOD1 was identified in Chapter 4 as a potential regulator of HoxB8 neutrophil differentiation, as determined by the identification of ubiquitin-modified peptides derived from known substrates of YOD1 showing decreasing abundance profiles during HoxB8 differentiation. These substrates included NEDD4, VCP, and ITCH, which are proteins involved in cellular ubiquitination events and protein quality control. Here, YOD1 KO in HL60 cells appeared to affect the differentiation process, resulting in fewer cells with segmented nuclei and decreased expression of CD11b in day-5 dHL60 cells. This provides further evidence that YOD1 may indeed contribute to neutrophil differentiation.

Other DUBs identified present new opportunities to understand regulatory aspects of neutrophil development and are discussed below.

BAP1 is known to be involved in tumour microenvironment modulation, affecting inflammation and the recruitment of various immune cells through signalling pathways such as NF-κB[242,243]. BAP1 expression has been identified in the nuclei of neutrophils[244]. BAP1 roles in DNA repair, cell cycle regulation, and chromatin modification suggest an overarching function of BAP1 protein in transcriptional reprogramming that may contribute to neutrophil differentiation. However, the exact functional roles that BAP1 plays in neutrophil biology remains to be determined.

COPS6 is a member of the CSN, a protein complex involved in DNA repair, cell cycle control, signal transduction, and protein degradation. The CSN controls neddylation (a ubiquitin-like PTM) of notably the cullin-RING-ubiquitin ligases (CRLs), a family of E3 ligases that control a significant proportion of cellular ubiquitination events, essentially affecting protein degradation in the majority of cellular processes[245]. Again, CSN has been described as a regulator of tumour development and immune microenvironment, rather than playing a functional role within immune cells[246-248]. Validation of successful KO of COPS6 and verification of the effect on neutrophil differentiation may reveal new insights into CSN function as well as its role in immune cell differentiation. Interestingly, COPS5 KO did not have a similar effect compared to COPS6 KO, potentially indicating differential roles of these subunits in regulating protein stability and consequent biological pathways.

VCPIP1 has been shown to affect toll-like receptor 4 (TLR4) signalling in macrophages, interacting with interleukin-1 receptor-associated kinases (IRAKs) and stabilising them through prevention of ubiquitination rather than direct deubiquitination[249]. VCPIP1 has also been shown to regulate the NLRP3 inflammasome, and myeloid-specific elimination of VCPIP1 resulted in increased activation of the NLRP3 inflammasome and increased neutrophil tissue invasion in a model of septic acute respiratory distress syndrome[250].

USP7 is involved in regulating diverse biological processes and many signalling pathways, from MAPK to Wnt[251,252]. It has been demonstrated to play a role in many cancers, promoting the proliferation of tumour cells[253]. For example, USP7 deubiquitinates KRAS, removing K48-linked polyubiquitin chains, and stabilising KRAS protein, thus promoting tumorigenesis[254]. Despite high LoF intolerance, USP7 appears to be dispensable in HL60 cells yet interferes with differentiation. The molecular basis underlying this effect is not known and warrants further investigations.

USP47 plays a regulatory role in the inflammatory response by affecting various cellular processes that control cell proliferation and survival. USP47 and USP7 perform similar roles in biological processes, such as NLRP3 inflammasome regulation and DNA repair[255,256]. Since USP47 and USP7 share structural similarity and molecular functions, it may be unsurprising that KO of both DUBs affected HL60 differentiation in similar ways.

OTUB2 plays a role in immune responses and various diseases, such as cancer, through regulating signalling pathways such as Hippo, mTOR, and NF-κB[257,258]. The role of OTUB2 in neutrophil biology remains elusive, yet these data indicate a potential link to neutrophil differentiation.

USP44 regulates cell cycle progression and chromosome segregation[259]. Its role in cell cycle control may contribute to neutrophil differentiation and potentially the unique morphology of the neutrophil nucleus. USP44 KO appeared to increase the proportion of dHL60 cells, therefore removal of this checkpoint may facilitate the establishment of differentiation programs. USP7 and USP44 have been shown to be involved in the maintenance of stem cell pluripotency[260]. Accordingly, their selective KO may impact cell differentiation in specific ways dependant on the underlying mechanisms and consequences of their enzymatic activity.

Taken together, VCPIP1, USP7, and USP47 have previously been demonstrated to regulate the NLRP3 inflammasome. In Chapter 5, several components of the NLRP3 inflammasome were shown to be upregulated on both the mRNA and protein levels, as demonstrated using the HoxB8 neutrophil differentiation model. NLRP3 inflammasome proteins may be vital for neutrophil maturation and stabilised by these DUBs during neutrophil differentiation. This necessitates further investigation to

understand how VCPIP1, USP7, and USP47 affect neutrophil development and whether this has a direct impact on the abundance of NLRP3 inflammasome components or affects other proteins with unforeseen roles in neutrophil biology. Furthermore, OTUB2 and BAP1 are known to influence the NF-κB signalling pathway, a key mediator of NLRP3 inflammasome priming[97]. The evidence suggests that there may be a link between DUB activity, NF-κB signalling, and inflammasome priming within the context of neutrophil biology and development.

Using microscopy as a phenotypic readout for high-throughput screening requires careful consideration of the experimental setup, sample preparation, image acquisition, and data analysis to extract biologically meaningful and relevant features that address the research question at hand. Appropriate controls and normalisation methods allow for the identification of signals. In this work, for example, due to the lack of cells identified in HL60 WT and HL60-Cas9 samples in day 0, an appropriate base line was missing from the data. Certain DUB KOs may affect the progenitor state of cells. Therefore, baseline information would be helpful in identifying such effects.

Using statistical approaches allows for the prioritisation of candidate targets based on evidence acquired during biological experimentation. Looking at the data from the DUB KO screen, KS testing did not result in significant FDR-adjusted p-values for USP26 KO cells, yet visual inspection of the form factor distribution highlighted a noticeable decrease in dHL60 cells (i.e. there were fewer cells with low form factor values). Western blot validation confirmed a decrease in CD11b expression, a key marker of neutrophil maturity, in day 5 dHL60 cells. Solely relying on significance thresholds to inform future experimentation may miss potentially relevant and novel biological phenomena.

Since HL60 cells exhibit promyelocytic leukemia cell-like properties, there are inherent limitations in using dHL60 or NLCs as a proxy for physiologically relevant neutrophils. Altered signalling pathways and metabolism in cancerous progenitor cells are likely causing aberrant cell surface marker expression and affecting the efficiency of DMSO-induced differentiation (approximately 30% of dHL60 cells showed neutrophil nuclear morphology). In addition, flow cytometry analysis revealed increased CD11b expression on the surface of cells, whereas CXCR4 and CXCR2 expression showed opposite changes compared to primary neutrophils. This may be due to technical reasons, such as PFA fixation of cells affecting antibody binding, or inability to recognise the antibody epitopes. Therefore, positive controls would prove useful. Alternatively, the expression of these markers may be indicative of altered promyelocyte biology and resultant phenotypes of dHL60 cells. These limitations need to be considered when analysing and interpreting the results from the arrayed CRISPR-Cas9 DUB KO screen.

Cas9 overexpression and DUB KO were performed on the population level (i.e. single clones were not selected). As a result, a mixed population of HL60 cells will contain different integration sites of Cas9

overexpression and sgRNA plasmids, potentially affecting nearby genomic regions, corresponding pathways of associated genes, and the degrees of zygosity within the population. These consequences may affect the differentiation process of HL60 cells into NLCs, thereby necessitating validation of any hits by clonal selection and DNA sequencing to verify the nature of the gene KO. Validation experiments would include Western blot analysis to confirm DUB KO and the potential effect on differentiation markers associated with HL60 differentiation. Additionally, flow cytometry analysis could also be optimised to include appropriate controls in order to confirm surface expression. Image analysis would also be performed to recapitulate the results from the DUB KO screen.

Regarding image analysis, although the form factor metric was able to capture certain information about nuclear morphology and segmentation, the degree of segmentation (e.g. hyper segmentation) or lack thereof would not be considered. Image analysis pipelines could be optimised to identify such features, potentially using 3D image analysis and leveraging Z-stacked data to identify more details and characteristics of HL60 morphology during differentiation.

Various sources contain information on ‘essential genes’ that are necessary for the viability of cells, and, more broadly, healthy function in organisms. Incorporating such data into the analysis of high-throughput genetic screens can better inform the selection of potential hits for further validation. Essential genes may be causing a phenotypic effect due to altered general cellular processes and homeostasis, rather than contributing to a specific underlying mechanism that participates in neutrophil differentiation. Additionally, the viability of the HL60 cells during DMSO-induced differentiation was not measured; counting cells in each microscopy image would not be indicative of cell viability due to potential bias introduced when selecting positions in wells. This presents several challenges in delineating the functions of DUBs in molecular processes that drive or maintain neutrophil differentiation programmes since a specific DUB’s action may be direct or indirect. The selection of candidate DUBs to validate would reduce the scale of experiments, allowing for discernment of gene essentiality using lower-throughput, more targeted characterisations and approaches, such as clonal selection, DNA sequencing, and cell viability assays.

In addition to gene essentiality, genetic variants related to candidate DUBs could be investigated to see if there is an association with neutrophil or immunological diseases and disorders. This would provide more evidence to support the mechanistic characterisation of a candidate DUB’s role in neutrophil differentiation, highlighting the clinical relevance.

Chapter 7: General discussion and future perspectives

7.1. The role of PTMs in neutrophil differentiation appears dynamic and complex

Overall, this work presents new insights into the dynamics of the protein and PTM landscape during neutrophil differentiation. Specifically, protein ubiquitination and phosphorylation, two common PTMs that are critical in regulating protein fate and function, were investigated. Numerous proteins demonstrate altered abundance and/or PTM levels across different stages of HoxB8 neutrophil differentiation, indicating highly complex and coordinated changes that may be fundamental to neutrophil lineage specification.

Key proteins have already been discussed in the results and discussion sections of chapters corresponding to the HoxB8 neutrophil model (Chapters 3-5). These data provide a comprehensive proteomic and PTMomic dataset, generating many hypotheses and expanding the understanding of neutrophil biology and potentially cell differentiation in general. In this thesis, select proteins and markers were discussed. For example, the role of inflammasome complexes during neutrophil differentiation is yet to be explored (see Section 1.4). Generally, other researchers may be able to use these rich datasets to complement or validate their own research, ultimately furthering scientific understanding of neutrophil biology.

7.2. PTM changes during neutrophil differentiation present novel hypotheses

Ubiquitination is a complex PTM language that requires targeted approaches to understand better the roles that Ub plays in regulating specific protein behaviours and functions, such as stability and localisation. There may be a link between the ubiquitination changes that occur during neutrophil differentiation and the migratory capacity in mature neutrophils. Upon maturation, neutrophils consist of a segmented, multi-lobed nucleus. This distinct nuclear morphology has been shown to facilitate neutrophil migration[261], highlighting the importance of appropriate neutrophil differentiation in enabling effective functions[262]. Concurrently, there is a body of literature that implicates the ubiquitin system in cell migration pathways, notably identifying DUBs, such as USP8[263], USP9X[264], and USP27X[265], and E3 ligases[266], such as HERC1[267], which regulate cell adhesion and migration. Additionally, the DUB USP17 plays essential roles in regulating GTPase subcellular localisation and cell motility[81]. USP17 was shown to be upregulated in activated neutrophils. These findings suggest associations between neutrophil differentiation, cell migration, and the ubiquitin system.

Due to the diverse nature of ubiquitination, which encompasses various chain architectures, linkage types, and chain constituents, a substrate-centric approach may be more feasible. To that end, identifying DUBs that modulate protein ubiquitination within the context of neutrophil differentiation would open avenues towards revealing key protein substrates that are instrumental in neutrophil lineage determination. Additionally, experimentation from the perspective of DUBs provides a more manageable set of genes to investigate, probing roughly 100 genes compared to the approximately 600 genes that encode the E3 ligases.

Affinity purification (AP)-MS analysis can be used to identify proteins that directly or indirectly interact with a DUB of interest[268]. Since PTM enzymes may also form transient interactions with substrates, a proximity labelling approach could be employed to reveal proteins that are within the vicinity of a candidate DUB, inferring potential associations between a DUB and substrates[269]. Previous work has been performed on proximity labelling approaches paired with Ub-modified peptide enrichment, resulting in a more streamlined identification of novel substrates of the DUB USP30[270].

Complementary to these approaches, identified DUBs can also be genetically modified to characterise the substrates upon which they exert their DUB activity. For example, MS-based proteomics can identify which substrates are changing in abundance as a result of gene KO of a DUB. Additionally, a catalytically inactive variant of a DUB[271] can be introduced into cells, and changes in protein abundance or Ub-modified peptide levels can be compared to those of WT DUBs to determine which substrates depend on DUB activity. The use of both WT and catalytically inactive add-back constructs provides a critical control to determine whether the observed effects are attributable to the enzymatic activity of the DUB, rather than to potential scaffolding or non-catalytic functions. For example, USP18 has been shown to perform scaffolding functions in interferon signalling pathways in addition to its deISGylating activity[272]. This approach is essential for establishing confidence that the differentiation defect arises specifically from loss of DUB activity.

Investigation of the exact roles and consequences of ubiquitination and Ub chain architecture can involve several approaches[273]. For example, different Ub chain architectures can be enriched using binding entities[274], and enriched proteins can be analysed via Western blot (if substrates are known) or MS (for unbiased characterisation). Alternatively, specific proteases could be used to remove (specific) Ub chains from substrates for characterisation[275]. Key residues within identified substrates corresponding to the sites of ubiquitination can be mutated, and mutant versions of proteins can be examined to determine consequent effects on a biological process, such as cellular differentiation.

Phosphorylation consists of a covalently linked phosphate group to an amino acid side chain. Despite this simpler covalent structure compared to polymerised Ub chains, the complexity arises from

multiple, potentially ambiguous phosphorylation sites, heterogenous protein populations, high lability of phosphate groups, variable stoichiometries, and context-specific temporal dynamics that modulate biological processes and pathways. Validation of protein phosphorylation can include antibody-based methods, such as Western blot and enzyme-linked immunosorbent assay (ELISA), or broader approaches to detect general protein phosphorylation, such as pro-Q Diamond dye staining and other MS-based approaches[276]. Accordingly, identification of specific proteins, potentially those that are instrumental in signalling pathways of interest, such as the NF-κB and small GTPase signalling pathways, would facilitate the validation of phosphorylation events and dynamics during neutrophil differentiation.

7.3. PTM enzymes associated with the ubiquitin system may play critical roles in regulating neutrophil differentiation

Importantly, PTMs are regulated by different families of enzymes. Ubiquitination status of a protein relies on the interplay between E3 ubiquitin-protein ligases and DUBs, while phosphorylation depends on kinases and phosphatases. Leveraging information about known enzyme-substrate interactions, it is possible to speculate on the potential roles a PTM enzyme may play in a biological process or pathway. However, this approach relies on well-defined, experimentally validated enzyme-substrate interactions that may not be applicable to certain biological contexts or cell types. Alternatively, enzyme-substrate interactions can be predicted based on, for example, the inherent enzyme preference for certain amino acid motifs, leading to substrate recognition. Despite expanding the potential substrate space, these are ultimately predictions that would require further validation. Databases that incorporate more, better-quality, and validated enzyme-substrate interactions would improve predictive power and ability to identify more relevant PTM machinery. Nevertheless, certain PTM enzymes were identified through the analysis of the HoxB8 datasets (Chapter 4) and the CRISPR-Cas9 KO screen targeting the DUBs (Chapter 6).

YOD1 appears to be a promising DUB to follow up in validation experiments. YOD1 KO in HL60 cells demonstrated reduced ability for cells to differentiate into NLCs. Additionally, HoxB8 proteomic and ubiquitomic profiling demonstrated that specific YOD1 substrates showed decreasing ubiquitination levels during differentiation. YOD1 deubiquitinates the E3 ligase NEDD4, supressing NEDD4-induced cell proliferation via the Hippo signalling pathway[186]; stabilises ITCH, also modulating Hippo signalling[187]; and regulates VCP, a protein involved in lysosomal homeostasis and protein quality control[189,190]. YOD1 and USP21 were also demonstrated to have a synergistic effect on the Hippo signalling pathway, with YOD1 affecting USP21 stability, affecting cell proliferation and survival[188]. Interactions between YOD1 and VCP play essential roles in the ERAD pathway[191], which may be a

feature of neutrophil differentiation, as indicated in Chapters 3 and 4. The role of YOD1 in various signalling pathways and biological processes has highlighted YOD1 as a potential therapeutic target in several disease contexts[194]. In the literature, there are some indications of YOD1 participating in neutrophil production, as seen in Yod1^{tm1bNarl}/ Yod1^{tm1bNarl} mice that demonstrated an increase in neutrophil cell numbers[193]. Additionally, YOD1 has been implicated in cell migration within the context of tumour progression[192,277] and keratinocyte biology[278], once again highlighting potential links between neutrophil differentiation and cell migration. Known substrates of YOD1 and indications of YOD1 functions in neutrophils, cell survival, and cell migration lead to speculations of YOD1 being a master regulator of different ubiquitin-associated pathways that may be critical in giving rise to necessary molecular components to enable granulopoiesis.

There are over 600 E3 ligases[279], over 500 kinases[280], and approximately 200 phosphatases[281] in the human genome. The workflow presented for the CRISPR-Cas9 DUB KO screen would need to be optimised for an increased number of targets. This could involve reducing the number of PTM enzyme targets to those expressed in cells or modifying the phenotypic readout to allow for a pooled screen (instead of the arrayed format demonstrated in this work).

The HL60 model provided a means to investigate the roles of DUBs in neutrophil differentiation, allowing for sufficient biological material and genetic manipulation. Several hypotheses have been generated, warranting further investigation to elucidate the underlying molecular mechanisms. Proteomic workflows have already been applied to primary progenitor cells and neutrophils derived from human and mouse tissues. Performing a similar approach to the HoxB8 model, as demonstrated in Chapters 3 and 4, on HL60 cells undergoing differentiation into NLCs would provide valuable data to determine the robustness of the HL60 model in recapitulating the proteomic changes that occur at different developmental stages of the neutrophil lineage. This would provide valuable data that would enhance biological research involving HL60 cells.

7.4. The NLRP3 inflammasome may participate in neutrophil differentiation

Proteins of the NLR signalling pathway were shown to be upregulated as HoxB8 neutrophils matured (Chapter 3). Integrative -omics revealed components of the NLRP3 inflammasome, in particular NLRP3, ASC or PYCARD, and CASP1 increased in mRNA and protein levels during HoxB8 neutrophil differentiation (Chapter 5). Rising levels of GSDMD were detected in the transcriptome and proteome datasets, whereas IL1β and IL18 proteins were absent. These findings suggest that NLRP3 inflammasome upregulation may be a feature or prerequisite for neutrophil maturation. This contrasts with the regulatory mechanism of NLRP3 inflammasome activation in monocytes and macrophages[213],

where a two-stage priming and activation are required to give rise to appropriate gene expression, protein synthesis, and cytokine release[217]. The data suggests that neutrophils upregulate key components of the NLRP3 inflammasome in the absence of typical priming signals. This may be representative of the biological differences between neutrophils and monocytes/macrophages. Neutrophils are typically short-lived leukocytes that primarily circulate in the blood and become recruited to sites of inflammation, providing a rapid and succinct response to pathogenic threats. In contrast, monocytes can differentiate into macrophages that are longer-lived and tissue resident. Differences in NLRP3 inflammasome regulation may be indicative of the differing lifespans and physiological roles of these immune cells.

Further investigations are required to elucidate the underlying mechanisms that lead to this observed increase in gene expression and protein abundance of NLRP3 inflammasome components during neutrophil differentiation. This would involve analysing inflammasome components in progenitor cells, immature neutrophils, and mature neutrophils using approaches such as real-time PCR to determine mRNA levels. Additionally, delineation of the signalling pathways that bring about these changes can provide insights into the regulatory aspects that govern these molecular changes. For example, is the NF-κB pathway or an alternative pathway involved? Are signals that cause gene expression endogenous or exogenous? Do NLRP3 inflammasome components assemble? Are IL1β and IL18 secreted from cells or degraded? What role does GSDMD play in neutrophil viability and pyroptosis, within the context of differentiation? These research questions constitute interesting hypotheses that can be further investigated. For example, secreted cytokines in cell culture supernatants can be detected using ELISA, and ASC/NLRP3 oligomerisation assays can indicate whether NLRP3 inflammasome components are assembling[282].

7.5. Limitations

The work presented in this thesis employed *ex vivo* and *in vitro* cellular models to recapitulate the neutrophil differentiation trajectory for experimental characterisation. HoxB8 limitations have been discussed in the Discussion section of Chapter 3, and those of HL60 cells were discussed in the Discussion section of Chapter 6. Cellular models prove to be an indispensable tool to study biological processes in a controlled manner, providing reproducible and scalable alternatives to working with organisms and whole tissues. Additionally, cell models provides a means to perform genetic perturbations, characterise changes in molecular pathways and phenotypic features, and gain mechanistic insights into bioilogical processes. However, cell models cannot completely recapitulate the exact physiological niche that would occur naturally within an organism. In the case of neutrophils, cell culture and manipulation can potentially lead to aberrant behaviours that would otherwise not be

observed *in vivo* (in the bone marrow, in circulation, and in tissue). Accordingly, future work should strive to involve validation of findings and investigations into human/mouse tissue, *in vivo* models, and primary neutrophil cells, all of which necessitate additional resources and investment into opstimisation. Nevertheless, extraction and genetic manipulation of primary neutrophil cells is difficult since neutrophils are short-lived cells with high turnover. Genetic manipulation requires a dedicated time investment to observe the molecular effects, and during this time the cells may develop unintended adaptations that no longer recapitulate aspects of neutrophil biology. Using the HoxB8 and HL60 cell models mitigated these issues since cells could be cultured, expanded, and differentiated in controlled experiments.

Reproducibility of the data is key to identify true signals. Several experiments, such as flow cytometry, microscopy, and Western blot validations, consisted of a single replicate (n = 1). However, with regards to the CRISPR KO screen, at least 3 distinct sgRNAs were used to generate a mixed population of DUB KO cells for each DUB gene. By using approaches such as flow cytometry or microscopy, consistent changes in a phenotype for each cell provides more confidence that the observed changes are due to DUB KO, irrespective of sgRNA. Nevertheless, technical variation in cell culture and assay plates may introduce biological noise, leading to erroneous conclusions such as false postiives. Repeating such analyses, potentially employing concurrent orthogonal approaches, with independent biological replicates and validated gentic manipulation would provide more confidence in the findings. For example, optimised flow cytometry workflows could be used to validate CSM expression of proteins that were analysed using Western blot for hits identified via microscopy approaches.

7.6. Conclusion and Perspectives

In summary, this thesis provides a comprehensive and unbiased characterisation of global proteomic, ubiquitomic, and phosphoproteomic dynamics during neutrophil differentiation, using the HoxB8 mouse model. These datasets reveal stage-specific patterns of PTM regulation, implicating a coordinated remodelling of signalling networks and effector functions during differentiation. Integrative bioinformatic analyses identified potential candidate regulators, including previously unrecognised E3 ligases, DUBs, kinases, and phosphatases, some of whose activity or modification state strongly correlates with distinct differentiation phases. The targeted CRISPR-Cas9 phenotypic screen identified candidate DUBs that may participate in the regulation of neutrophil differentiation, using an HL60 human model.

Several of the identified PTM enzymes and novel protein markers warrant further experimental validation to assess their mechanistic role in neutrophil lineage specification and function. Notably, investigating the role of DUBs such as YOD1 as well as the NLRP3 inflammasome in neutrophil

differentiation would be worthy endeavours. Logical next steps to investigate the function of YOD1 in neutrophils would involve genetic modification of YOD1 to characterise its contributions to neutrophil differentiation and identification of YOD1 protein substrates. To delineate the role of the NLRP3 inflammasome, transcriptional upregulation should be first verified to then interrogate the inflammasome priming signalling pathways that give rise to transcriptional upregulation of NLRP3 inflammasome components. Moreover, the production of inflammasome effector molecules such as IL1β and IL18 can be investigated to determine whether they are secreted or degraded during neutrophil differentiation.

Integration and visualisation of multi-omics datasets provide indications of connected biological phenomena. Development of analytical bioinformatic tools in this work has applications in other systems biology contexts involving different modalities and -omics approaches. New insights into neutrophil differentiation presented in this work provide further areas of research with potential therapeutic relevance in innate immune response pathways and inflammatory and infectious diseases.

References

1. Khoyratty TE, Ai Z, Ballesteros I, et al. Distinct transcription factor networks control neutrophil-driven inflammation. *Nat Immunol*. 2021;22(9):1093–1106.
2. Servant G, Weiner OD, Neptune ER, Sedat JW, Bourne HR. Dynamics of a chemoattractant receptor in living neutrophils during chemotaxis. *Mol Biol Cell*. 1999;10(4):1163–1178.
3. Kolaczkowska E, Kubes P. Neutrophil recruitment and function in health and inflammation. *Nature reviews immunology*. 2013;13(3):159–175.
4. Mócsai A, Walzog B, Lowell CA. Intracellular signalling during neutrophil recruitment. *Cardiovasc Res*. 2015;107(3):373–385.
5. Bonecchi R, Mantovani A, Jaillon S. Chemokines as regulators of neutrophils: Focus on tumors, therapeutic targeting, and immunotherapy. *Cancers*. 2022;14(3):680.
6. Cheng Y, Ma X, Wei Y, Wei X. Potential roles and targeted therapy of the CXCLs/CXCR2 axis in cancer and inflammatory diseases. *Biochimica et Biophysica Acta (BBA)-Reviews on Cancer*. 2019;1871(2):289–312.
7. Giridharan S, Srinivasan M. Mechanisms of NF-κB p65 and strategies for therapeutic manipulation. *Journal of inflammation research*. 2018:407–419.
8. Lokuta MA, Nuzzi PA, Huttenlocher A. Analysis of neutrophil polarization and chemotaxis. In: *Neutrophil methods and protocols.* Springer; 2007:211–229.
9. Servant G, Weiner OD, Herzmark P, Balla T, Sedat JW, Bourne HR. Polarization of chemoattractant receptor signaling during neutrophil chemotaxis. *Science*. 2000;287(5455):1037–1040.
10. He Y, Li D, Cook SL, et al. Mammalian target of rapamycin and rictor control neutrophil chemotaxis by regulating rac/Cdc42 activity and the actin cytoskeleton. *Mol Biol Cell*. 2013;24(21):3369–3380.
11. Shi Y, Zhang J, Mullin M, Dong B, Alberts AS, Siminovitch KA. The mDial formin is required for neutrophil polarization, migration, and activation of the LARG/RhoA/ROCK signaling axis during chemotaxis. *The Journal of Immunology*. 2009;182(6):3837–3845.
12. Naish E, Wood AJ, Stewart AP, et al. The formation and function of the neutrophil phagosome. *Immunol Rev*. 2023;314(1):158–180.
13. Roberts RE, Hallett MB. Neutrophil cell shape change: Mechanism and signalling during cell spreading and phagocytosis. *International journal of molecular sciences*. 2019;20(6):1383.
14. Winterbourn CC, Kettle AJ, Hampton MB. Reactive oxygen species and neutrophil function. *Annu Rev Biochem*. 2016;85(1):765–792.
15. Biswas S, Sarojini S, Jayaram S, et al. Understanding the role of antimicrobial peptides in neutrophil extracellular traps promoting autoimmune disorders. *Life*. 2023;13(6):1307.
16. Teng T, Ji A, Ji X, Li Y. Neutrophils and immunity: From bactericidal action to being conquered. *Journal of immunology research*. 2017;2017(1):9671604.
17. Naegelen I, Beaume N, Plançon S, Schenten V, Tschirhart EJ, Bréchard S. Regulation of neutrophil degranulation and cytokine secretion: A novel model approach based on linear fitting. *Journal of immunology research*. 2015;2015(1):817038.
18. Faurschou M, Borregaard N. Neutrophil granules and secretory vesicles in inflammation. *Microb Infect*. 2003;5(14):1317–1327.
19. Vorobjeva NV, Chernyak BV. NETosis: Molecular mechanisms, role in physiology and pathology. *Biochemistry (Moscow)*. 2020;85(10):1178–1190.
20. Schreiber A, Rousselle A, Becker JU, Von Mässenhausen A, Linkermann A, Kettritz R. Necroptosis controls NET generation and mediates complement activation, endothelial damage, and autoimmune vasculitis. *Proceedings of the National Academy of Sciences*. 2017;114(45):E9618–E9625.
21. D'Cruz AA, Speir M, Bliss-Moreau M, et al. The pseudokinase MLKL activates PAD4-dependent NET formation in necroptotic neutrophils. *Science signaling*. 2018;11(546):eaao1716.
22. Yipp BG, Petri B, Salina D, et al. Infection-induced NETosis is a dynamic process involving neutrophil multitasking in vivo. *Nat Med*. 2012;18(9):1386–1393.
23. Zhang F, Xia Y, Su J, et al. Neutrophil diversity and function in health and disease. *Signal transduction and targeted therapy*. 2024;9(1):343.
24. Fiedler K, Brunner C. The role of transcription factors in the guidance of granulopoiesis. *American journal of blood research*. 2012;2(1):57.
25. Skokowa J, Cario G, Uenalan M, et al. LEF-1 is crucial for neutrophil granulocytopoiesis and its expression is severely reduced in congenital neutropenia. *Nat Med*. 2006;12(10):1191–1197.
26. de la Luz Sierra M, Sakakibara S, Gasperini P, et al. The transcription factor Gfi1 regulates G-CSF signaling and neutrophil development through the ras activator RasGRP1. *Blood, The Journal of the American Society of Hematology*. 2010;115(19):3970–3979.
27. Gery S, Gombart AF, Fung YK, Koeffler HP. C/EBPϵ interacts with retinoblastoma and E2F1 during granulopoiesis. *Blood*. 2004;103(3):828–835.
28. Hirai H, Zhang P, Dayaram T, et al. C/EBPβ is required for'emergency'granulopoiesis. *Nat Immunol*. 2006;7(7):732–739.
29. Yáñez A, Goodridge HS. Interferon regulatory factor 8 and the regulation of neutrophil, monocyte, and dendritic cell production. *Curr Opin Hematol*. 2016;23(1):11–17.

30. Huang Z, Chen K, Chi Y, et al. Runx1 regulates zebrafish neutrophil maturation via synergistic interaction with c-myb. *J Biol Chem*. 2021;296.
31. Blanter M, Gouwy M, Struyf S. Studying neutrophil function in vitro: Cell models and environmental factors. *Journal of inflammation research*. 2021:141–162.
32. Wang GG, Calvo KR, Pasillas MP, Sykes DB, Häcker H, Kamps MP. Quantitative production of macrophages or neutrophils ex vivo using conditional Hoxb8. *Nature methods*. 2006;3(4):287–293.
33. Salafranca J, Ai Z, Wang L, Udalova IA, van Grinsven E. Analysis of neutrophil morphology and function under genetic perturbation of transcription factors in vitro. In: *Transcription factor regulatory networks.* Springer; 2022:69–86.
34. Shannon JG, Hinnebusch BJ. Characterization and CRISPR/Cas9-mediated genetic manipulation of neutrophils derived from Hoxb8-ER-immortalized myeloid progenitors. *J Leukoc Biol*. 2023;114(1):42–52.
35. Chu JY, McCormick B, Mazelyte G, Michael M, Vermeren S. HoxB8 neutrophils replicate fcγ receptor and integrin-induced neutrophil signaling and functions. *J Leukoc Biol*. 2019;105(1):93–100.
36. Collins SJ. The HL-60 promyelocytic leukemia cell line: Proliferation, differentiation, and cellular oncogene expression. 1987.
37. Millius A, Weiner OD. Manipulation of neutrophil-like HL-60 cells for the study of directed cell migration. In: *Live cell imaging: Methods and protocols.* Springer; 2009:147–158.
38. Bhakta SB, Lundgren SM, Sesti BN, et al. Neutrophil-like cells derived from the HL-60 cell-line as a genetically-tractable model for neutrophil degranulation. *PLoS One*. 2024;19(2):e0297758.
39. Guo Y, Gao F, Wang Q, et al. Differentiation of HL-60 cells in serum-free hematopoietic cell media enhances the production of neutrophil extracellular traps. *Experimental and Therapeutic Medicine*. 2021;21(4):353.
40. Kim K, Seoh JY, Cho SJ. Phenotypic and functional analysis of HL-60 cells used in opsonophagocytic-killing assay for streptococcus pneumoniae. *J Korean Med Sci*. 2015;30(2):145–150.
41. Narimani M, Sharifi M, Jalili A. Knockout of BIRC5 gene by CRISPR/Cas9 induces apoptosis and inhibits cell proliferation in leukemic cell lines, HL60 and KG1. *Blood and lymphatic cancer: targets and therapy*. 2019:53–61.
42. MacDonald RJ, Shrimp JH, Jiang H, Zhang L, Lin H, Yen A. Probing the requirement for CD38 in retinoic acid-induced HL-60 cell differentiation with a small molecule dimerizer and genetic knockout. *Scientific reports*. 2017;7(1):17406.
43. Bao S, He D, Zeng J, Zhang Y, Chen S. Effect of miRNA-19a antisense oligonucleotide and ara-C on the proliferation and apoptosis of HL60 cells. *Annals of Translational Medicine*. 2019;7(12):260.
44. Fu S, Fu Y, Chen F, Hu Y, Quan B, Zhang J. Overexpression of MYCT1 inhibits proliferation and induces apoptosis in human acute myeloid leukemia HL-60 and KG-1a cells in vitro and in vivo. *Frontiers in Pharmacology*. 2018;9:1045.
45. Yi T, Wenli L, Jianfeng Z, Huizhen X, Wu L. Transfection efficiency comparison of oligonucleotide and plasmid to the HL-60 cell line with liposomes. *Current Medical Science*. 2005;25(1):24–25.
46. Al-Otaibi NA, Cassoli JS, Slater NK, Rahmoune H. Molecular characterization of human leukemia 60 (HL-60) cells as a model of acute myelogenous leukemia post cryopreservation. In: *Pre-clinical models: Techniques and protocols.* Springer; 2018:239–247.
47. Liu K, Yeh I, Hsu Y, Yen M. Transcriptomic analysis of lipoteichoic acid-treated undifferentiated and neutrophil-like differentiated HL-60 cells. *Experimental and Therapeutic Medicine*. 2024;27(4):158.
48. Rincón E, Rocha-Gregg BL, Collins SR. A map of gene expression in neutrophil-like cell lines. *BMC Genomics*. 2018;19(1):573.
49. Scieszka D, Lin Y, Li W, Choudhury S, Yu Y, Freire M. NETome: A model to decode the human genome and proteome of neutrophil extracellular traps. *Scientific Data*. 2022;9(1):702.
50. Thimmappa PY, Nair AS, D'silva S, et al. Neutrophils display distinct post-translational modifications in response to varied pathological stimuli. *Int Immunopharmacol*. 2024;132:111950.
51. Chen Y, Ren W, Wang Q, He Y, Ma D, Cai Z. The regulation of necroptosis by ubiquitylation. *Apoptosis*. 2022;27(9):668–684.
52. Swatek KN, Komander D. Ubiquitin modifications. *Cell Res*. 2016;26(4):399–422.
53. Chen ZJ, Sun LJ. Nonproteolytic functions of ubiquitin in cell signaling. *Mol Cell*. 2009;33(3):275–286.
54. Kelsall IR. Non-lysine ubiquitylation: Doing things differently. *Frontiers in Molecular Biosciences*. 2022;9:1008175.
55. Clague MJ, Urbé S. Ubiquitin: Same molecule, different degradation pathways. *Cell*. 2010;143(5):682–685.
56. Zhao L, Zhao J, Zhong K, Tong A, Jia D. Targeted protein degradation: Mechanisms, strategies and application. *Signal transduction and targeted therapy*. 2022;7(1):113.
57. Chauhan D, Catley L, Li G, et al. A novel orally active proteasome inhibitor induces apoptosis in multiple myeloma cells with mechanisms distinct from bortezomib. *Cancer cell*. 2005;8(5):407–419.
58. Clague MJ, Urbé S, Komander D. Breaking the chains: Deubiquitylating enzyme specificity begets function. *Nature reviews Molecular cell biology*. 2019;20(6):338–352.
59. Sahtoe DD, Sixma TK. Layers of DUB regulation. *Trends Biochem Sci*. 2015;40(8):456–467.
60. Leznicki P, Kulathu Y. Mechanisms of regulation and diversification of deubiquitylating enzyme function. *J Cell Sci*. 2017;130(12):1997–2006.
61. Mevissen TE, Komander D. Mechanisms of deubiquitinase specificity and regulation. *Annu Rev Biochem*. 2017;86(1):159–192.
62. Grou CP, Pinto MP, Mendes AV, Domingues P, Azevedo JE. The de novo synthesis of ubiquitin: Identification of deubiquitinases acting on ubiquitin precursors. *Scientific reports*. 2015;5(1):1–16.

63. Snyder NA, Silva GM. Deubiquitinating enzymes (DUBs): Regulation, homeostasis, and oxidative stress response. *J Biol Chem*. 2021;297(3).
64. Farshi P, Deshmukh RR, Nwankwo JO, et al. Deubiquitinases (DUBs) and DUB inhibitors: A patent review. *Expert opinion on therapeutic patents*. 2015;25(10):1191–1208.
65. Liu J, Leung CT, Liang L, et al. Deubiquitinases in cancers: Aspects of proliferation, metastasis, and apoptosis. *Cancers*. 2022;14(14):3547.
66. Ronau JA, Beckmann JF, Hochstrasser M. Substrate specificity of the ubiquitin and ubl proteases. *Cell Res*. 2016;26(4):441–456.
67. Zheng L, Wang L, Pang Y, Sun L, Shi L. Recent advances in the development of deubiquitinases inhibitors as antitumor agents. *Eur J Med Chem*. 2024;266:116161.
68. Clague MJ, Barsukov I, Coulson JM, Liu H, Rigden DJ, Urbé S. Deubiquitylases from genes to organism. *Physiol Rev*. 2013;93(3):1289–1315.
69. Henning NJ, Boike L, Spradlin JN, et al. Deubiquitinase-targeting chimeras for targeted protein stabilization. *Nature chemical biology*. 2022;18(4):412–421.
70. Chen Y, Xue H, Jin J. Applications of protein ubiquitylation and deubiquitylation in drug discovery. *J Biol Chem*. 2024;300(5):107264.
71. Nishi H, Hashimoto K, Panchenko AR. Phosphorylation in protein-protein binding: Effect on stability and function. *Structure*. 2011;19(12):1807–1815.
72. Ardito F, Giuliani M, Perrone D, Troiano G, Lo Muzio L. The crucial role of protein phosphorylation in cell signaling and its use as targeted therapy. *Int J Mol Med*. 2017;40(2):271–280.
73. Wu C, Ba Q, Lu D, et al. Global and site-specific effect of phosphorylation on protein turnover. *Developmental cell*. 2021;56(1):111–124. e6.
74. Cheng H, Qi RZ, Paudel H, Zhu H. Regulation and function of protein kinases and phosphatases. *Enzyme research*. 2011;2011:794089.
75. Pellarin I, Dall'Acqua A, Favero A, et al. Cyclin-dependent protein kinases and cell cycle regulation in biology and disease. *Signal Transduction and Targeted Therapy*. 2025;10(1):11.
76. Zhu Y, Huang Y, Ji Q, et al. Interplay between extracellular matrix and neutrophils in diseases. *Journal of immunology research*. 2021;2021(1):8243378.
77. Gambardella L, Vermeren S. Molecular players in neutrophil chemotaxis—focus on PI3K and small GTPases. *J Leukoc Biol*. 2013;94(4):603–612.
78. Yin G, Huang J, Petela J, et al. Targeting small GTPases: Emerging grasps on previously untamable targets, pioneered by KRAS. *Signal transduction and targeted therapy*. 2023;8(1):212.
79. Wang H, Zhang Y, Ozdamar B, et al. Regulation of cell polarity and protrusion formation by targeting RhoA for degradation. *Science*. 2003;302(5651):1775–1779.
80. Oberoi TK, Dogan T, Hocking JC, et al. IAPs regulate the plasticity of cell migration by directly targeting Rac1 for degradation. *EMBO J*. 2012;31(1):14–28.
81. de La Vega M, Kelvin AA, Dunican DJ, et al. The deubiquitinating enzyme USP17 is essential for GTPase subcellular localization and cell motility. *Nature communications*. 2011;2(1):259.
82. Song J, Zhang Y, Bai Y, et al. The deubiquitinase OTUD1 suppresses secretory neutrophil polarization and ameliorates immunopathology of periodontitis. *Advanced Science*. 2023;10(30):2303207.
83. Ou Q, Fang J, Zhang Z, et al. TcpC inhibits neutrophil extracellular trap formation by enhancing ubiquitination mediated degradation of peptidylarginine deiminase 4. *Nature Communications*. 2021;12(1):3481.
84. Liu H, Jing G, Wu S, et al. Acod1 promotes PAD4 ubiquitination via UBR5 alkylation to modulate NETosis and exert protective effects in sepsis. *Advanced Science*. 2025:e11652.
85. Arneth B. Defects in ubiquitination and NETosis and their associations with human diseases. *Pathology*. 2021;53(4):439–445.
86. Barrera-Vargas A, Gómez-Martín D, Carmona-Rivera C, et al. Differential ubiquitination in NETs regulates macrophage responses in systemic lupus erythematosus. *Ann Rheum Dis*. 2018;77(6):944–950.
87. László L, Doherty FJ, Watson A, et al. Immunogold localisation of ubiquitin-protein conjugates in primary (azurophilic) granules of polymorphonuclear neutrophils. *FEBS Lett*. 1991;279(2):175–178.
88. Suresh B, Lee J, Kim K, Ramakrishna S. The importance of ubiquitination and deubiquitination in cellular reprogramming. *Stem cells international*. 2016;2016(1):6705927.
89. Olguín HC. The gentle side of the UPS: Ubiquitin-proteasome system and the regulation of the myogenic program. *Frontiers in Cell and Developmental Biology*. 2022;9:821839.
90. Pan Y, Tang Y, Gu H, Ge W. Ubiquitin modification in osteogenic differentiation and bone formation: From mechanisms to clinical significance. *Frontiers in Cell and Developmental Biology*. 2022;10:1033223.
91. Xu P, Scott DC, Xu B, et al. FBXO11-mediated proteolysis of BAHD1 relieves PRC2-dependent transcriptional repression in erythropoiesis. *Blood, The Journal of the American Society of Hematology*. 2021;137(2):155–167.
92. Federzoni EA, Gloor S, Jin J, et al. Linking the SUMO protease SENP5 to neutrophil differentiation of AML cells. *Leukemia research reports*. 2015;4(1):32–35.
93. Forrer P, Palianina D, Stühler C, et al. Unveiling signaling pathways inducing MHC class II expression in neutrophils. *Frontiers in immunology*. 2024;15:1444558.

94. Ma Y, Chen X, Wang M, et al. Ruxolitinib targets JAK-STAT signaling to modulate neutrophil activation in refractory macrophage activation syndrome. *Blood Journal*. 2025:blood. 2024024362.
95. Panwar V, Singh A, Bhatt M, et al. Multifaceted role of mTOR (mammalian target of rapamycin) signaling pathway in human health and disease. *Signal transduction and targeted therapy*. 2023;8(1):375.
96. Itakura A, McCarty OJ. Pivotal role for the mTOR pathway in the formation of neutrophil extracellular traps via regulation of autophagy. *American Journal of Physiology-Cell Physiology*. 2013;305(3):C348–C354.
97. Liu T, Zhang L, Joo D, Sun S. NF-κB signaling in inflammation. *Signal transduction and targeted therapy*. 2017;2(1):1–9.
98. McDonald PP, Bald A, Cassatella MA. Activation of the NF-κB pathway by inflammatory stimuli in human neutrophils. *Blood, The Journal of the American Society of Hematology*. 1997;89(9):3421–3433.
99. Nakajima T, Kinoshita S, Sasagawa T, et al. Phosphorylation at threonine-235 by a ras-dependent mitogen-activated protein kinase cascade is essential for transcription factor NF-IL6. *Proceedings of the National Academy of Sciences*. 1993;90(6):2207–2211.
100. Pilipuk GP, Galigniana MD, Schwartz J. Subnuclear localization of C/EBPβ is regulated by growth hormone and dependent on MAPK. *J Biol Chem*. 2003;278(37):35668–35677.
101. Piwien-Pilipuk G, Van Mater D, Ross SE, MacDougald OA, Schwartz J. Growth hormone regulates phosphorylation and function of CCAAT/enhancer-binding protein β by modulating akt and glycogen synthase kinase-3. *J Biol Chem*. 2001;276(22):19664–19671.
102. Ramji DP, Foka P. CCAAT/enhancer-binding proteins: Structure, function and regulation. *Biochem J*. 2002;365(3):561–575.
103. Ai Z, Udalova IA. Transcriptional regulation of neutrophil differentiation and function during inflammation. *J Leukoc Biol*. 2020;107(3):419–430.
104. Baillet A, Hograindleur M, El Benna J, et al. Unexpected function of the phagocyte NADPH oxidase in supporting hyperglycolysis in stimulated neutrophils: Key role of 6-phosphofructo-2-kinase. *The FASEB Journal.* 2017;31(2):663–673.
105. Jeon J, Hong C, Kim EY, Lee JM. Current understanding on the metabolism of neutrophils. *Immune Network*. 2020;20(6):e46.
106. Min KI, Byeon S. Diagnosis and management of neutropenia. *Blood Research*. 2025;60(1):30.
107. Wu X, Chen J, Chen Y, et al. Targeting deltex E3 ubiquitin ligase 2 inhibits tumor-associated neutrophils and sensitizes hepatocellular carcinoma cells to immunotherapy. *Advanced Science*. 2025;12(7):2408233.
108. Gong Z, Li Q, Shi J, et al. Immunosuppressive reprogramming of neutrophils by lung mesenchymal cells promotes breast cancer metastasis. *Science immunology*. 2023;8(80):eadd5204.
109. Bissenova S, Ellis D, Mathieu C, Gysemans C. Neutrophils in autoimmunity: When the hero becomes the villain. *Clin Exp Immunol*. 2022;210(2):128–140.
110. Devaprasad A, Ai Z, Mukherjee AK, et al. Zfp263 is a transcriptional checkpoint of neutrophil development. *bioRxiv*. 2024:2024.08. 22.609132.
111. Demichev V, Messner CB, Vernardis SI, Lilley KS, Ralser M. DIA-NN: Neural networks and interference correction enable deep proteome coverage in high throughput. *Nature methods*. 2020;17(1):41–44.
112. Yu F, Teo GC, Kong AT, et al. Analysis of DIA proteomics data using MSFragger-DIA and FragPipe computational platform. *Nature communications*. 2023;14(1):4154.
113. Geiszler DJ, Polasky DA, Yu F, Nesvizhskii AI. Detecting diagnostic features in MS/MS spectra of post-translationally modified peptides. *Nature Communications*. 2023;14(1):4132.
114. Geiszler DJ, Kong AT, Avtonomov DM, Yu F, da Veiga Leprevost F, Nesvizhskii AI. PTM-shepherd: Analysis and summarization of post-translational and chemical modifications from open search results. *Molecular & Cellular Proteomics*. 2021;20.
115. Teo GC, Polasky DA, Yu F, Nesvizhskii AI. Fast deisotoping algorithm and its implementation in the MSFragger search engine. *Journal of proteome research*. 2020;20(1):498–505.
116. Yu F, Haynes SE, Teo GC, Avtonomov DM, Polasky DA, Nesvizhskii AI. Fast quantitative analysis of timsTOF PASEF data with MSFragger and IonQuant. *Molecular & Cellular Proteomics*. 2020;19(9):1575–1585.
117. Yu F, Teo GC, Kong AT, et al. Identification of modified peptides using localization-aware open search. *Nature communications*. 2020;11(1):4065.
118. Kong AT, Leprevost FV, Avtonomov DM, Mellacheruvu D, Nesvizhskii AI. MSFragger: Ultrafast and comprehensive peptide identification in mass spectrometry–based proteomics. *Nature methods*. 2017;14(5):513–520.
119. Li K, Teo GC, Yang KL, Yu F, Nesvizhskii AI. diaTracer enables spectrum-centric analysis of diaPASEF proteomics data. *Nature Communications*. 2025;16(1):95.
120. R Core Team R. R: A language and environment for statistical computing. 2020.
121. Villanueva RAM, Chen ZJ. No title. *ggplot2: elegant graphics for data analysis*. 2019.
122. Langfelder P, Horvath S. WGCNA: An R package for weighted correlation network analysis. *BMC Bioinformatics*. 2008;9(1):559.
123. Wu T, Hu E, Xu S, et al. clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *The innovation*. 2021;2(3).
124. Hornbeck PV, Kornhauser JM, Latham V, et al. 15 years of PhosphoSitePlus®: Integrating post-translationally modified sites, disease variants and isoforms. *Nucleic Acids Res*. 2019;47(D1):D433–D441.
125. Damle NP, Köhn M. The human DEPhOsphorylation database DEPOD: 2019 update. *Database*. 2019;2019:baz133.

126. Wang X, Li Y, He M, et al. UbiBrowser 2.0: A comprehensive resource for proteome-wide known and predicted ubiquitin ligase/deubiquitinase–substrate interactions in eukaryotic species. *Nucleic Acids Res*. 2022;50(D1):D719–D728.
127. Liska O, Bohár B, Hidas A, et al. TFLink: An integrated gateway to access transcription factor–target gene interactions for multiple species. *Database*. 2022;2022:baac083.
128. Chen S, Francioli LC, Goodrich JK, et al. A genomic mutational constraint map using variation in 76,156 human genomes. *Nature*. 2024;625(7993):92–100.
129. Arafeh R, Shibue T, Dempster JM, Hahn WC, Vazquez F. The present and future of the cancer dependency map. *Nature Reviews Cancer*. 2025;25(1):59–73.
130. Uhlén M, Fagerberg L, Hallström BM, et al. Tissue-based map of the human proteome. *Science*. 2015;347(6220):1260419.
131. Jin H, Zhang C, Zwahlen M, et al. Systematic transcriptional analysis of human cell lines for gene expression landscape and tumor representation. *Nature communications*. 2023;14(1):5417.
132. Stirling DR, Swain-Bowden MJ, Lucas AM, Carpenter AE, Cimini BA, Goodman A. CellProfiler 4: Improvements in speed, utility and usability. *BMC Bioinformatics*. 2021;22(1):433.
133. Delobel P, Ginter B, Rubio E, Balabanian K, Lazennec G. CXCR2 intrinsically drives the maturation and function of neutrophils in mice. *Frontiers in Immunology*. 2022;13:1005551.
134. Ellison 3rd RT, Giehl TJ, LaForce FM. Damage of the outer membrane of enteric gram-negative bacteria by lactoferrin and transferrin. *Infect Immun*. 1988;56(11):2774–2781.
135. Schroll A, Eller K, Feistritzer C, et al. Lipocalin-2 ameliorates granulocyte functionality. *Eur J Immunol*. 2012;42(12):3346–3357.
136. Lin M, Jackson P, Tester AM, et al. Matrix metalloproteinase-8 facilitates neutrophil migration through the corneal stromal matrix by collagen degradation and production of the chemotactic peptide pro-gly-pro. *The American journal of pathology*. 2008;173(1):144–153.
137. Quinn MT, Gauss KA. Structure and regulation of the neutrophil respiratory burst oxidase: Comparison with nonphagocyte oxidases. *J Leukoc Biol*. 2004;76(4):760–781.
138. Malech HL, DeLeo FR, Quinn MT. The role of neutrophils in the immune system: An overview. *Neutrophil methods and protocols*. 2014:3–10.
139. Venkatakrishnan V, Dieckmann R, Loke I, et al. Glycan analysis of human neutrophil granules implicates a maturation-dependent glycosylation machinery. *J Biol Chem*. 2020;295(36):12648–12660.
140. Chamovitz DA. Revisiting the COP9 signalosome as a transcriptional regulator. *EMBO Rep*. 2009;10(4):352–358.
141. Sun SC. CYLD: A tumor suppressor deubiquitinase regulating NF-κB activation and diverse biological processes. *Cell Death & Differentiation*. 2010;17(1):25–34.
142. L'Estrange-Stranieri E, Gottschalk TA, Wright MD, Hibbs ML. The dualistic role of lyn tyrosine kinase in immune cell signaling: Implications for systemic lupus erythematosus. *Frontiers in Immunology*. 2024;15:1395427.
143. Zhou Y, Zhang Y, Yu W, et al. Immunomodulatory role of spleen tyrosine kinase in chronic inflammatory and autoimmune diseases. *Immunity, Inflammation and Disease*. 2023;11(7):e934.
144. Leclair HM, Dubois SM, Azzi S, Dwyer J, Bidère N, Gavard J. Control of CXCR2 activity through its ubiquitination on K327 residue. *BMC cell biology*. 2014;15(1):38.
145. Khandagale A, Holmlund T, Entesarian M, et al. Severe congenital neutropenia-associated JAGN1 mutations unleash a calpain-dependent cell death programme in myeloid cells. *Br J Haematol*. 2021;192(1):200–211.
146. Fumagalli L, Zhang H, Baruzzi A, Lowell CA, Berton G. The src family kinases hck and fgr regulate neutrophil responses to N-formyl-methionyl-leucyl-phenylalanine. *The Journal of Immunology*. 2007;178(6):3874–3885.
147. Sheikh E, Tran T, Vranic S, Levy A, Bonfil RD. Role and significance of c-KIT receptor tyrosine kinase in cancer: A review. *Bosnian journal of basic medical sciences*. 2022;22(5):683.
148. Edling CE, Hallberg B. C-kit—a hematopoietic cell essential receptor tyrosine kinase. *Int J Biochem Cell Biol*. 2007;39(11):1995–1998.
149. Miyazawa K, Toyama K, Gotoh A, Hendrie PC, Mantel C, Broxmeyer HE. Ligand-dependent polyubiquitination of c-kit gene product: A possible mechanism of receptor down modulation in M07e cells. 1994.
150. Masson K, Heiss E, Band H, Rönnstrand L. Direct binding of cbl to Tyr568 and Tyr936 of the stem cell factor receptor/c-kit is required for ligand-induced ubiquitination, internalization and degradation. *Biochem J*. 2006;399(1):59–67.
151. Lubbers R, Van Essen MF, Van Kooten C, Trouw LA. Production of complement components by cells of the immune system. *Clinical & Experimental Immunology*. 2017;188(2):183–194.
152. Camous L, Roumenina L, Bigot S, et al. Complement alternative pathway acts as a positive feedback amplification of neutrophil activation. *Blood, The Journal of the American Society of Hematology*. 2011;117(4):1340–1349.
153. Zheng X, Zeng F, Lei Y, et al. YBX1: An RNA/DNA-binding protein that affects disease progression. *Frontiers in Oncology*. 2025;15:1635209.
154. Jayavelu AK, Schnöder TM, Perner F, et al. Splicing factor YBX1 mediates persistence of JAK2-mutated neoplasms. *Nature*. 2020;588(7836):157–163.
155. Shibata T, Watari K, Kawahara A, et al. Targeting phosphorylation of Y-box–binding protein YBX1 by TAS0612 and everolimus in overcoming antiestrogen resistance. *Molecular cancer therapeutics*. 2020;19(3):882–894.
156. Prabhu L, Mundade R, Wang B, et al. Critical role of phosphorylation of serine 165 of YBX1 on the activation of NF-κB in colon cancer. *Oncotarget*. 2015;6(30):29396.

157. Li X, Chen G, Liu B, et al. PLK1 inhibition promotes apoptosis and DNA damage in glioma stem cells by regulating the nuclear translocation of YBX1. *Cell death discovery*. 2023;9(1):68.
158. Mao S, Xie C, Liu Y, et al. Apurinic/apyrimidinic endodeoxyribonuclease 1 (APE1) promotes stress granule formation via YBX1 phosphorylation in ovarian cancer. *Cellular and Molecular Life Sciences*. 2024;81(1):113.
159. Hermert D, Martin IV, Reiss LK, et al. The nucleic acid binding protein YB-1–controlled expression of CXCL-1 modulates kidney damage in liver fibrosis. *Kidney Int*. 2020;97(4):741–752.
160. Alidousty C, Rauen T, Hanssen L, et al. Calcineurin-mediated YB-1 dephosphorylation regulates CCL5 expression during monocyte differentiation. *J Biol Chem*. 2014;289(31):21401–21412.
161. Ohe S, Kubota Y, Yamaguchi K, et al. ERK-mediated NELF-A phosphorylation promotes transcription elongation of immediate-early genes by releasing promoter-proximal pausing of RNA polymerase II. *Nature communications*. 2022;13(1):7476.
162. Sharma A, Takata H, Shibahara K, Bubulya A, Bubulya PA. Son is essential for nuclear speckle organization and cell cycle progression. *Mol Biol Cell*. 2010;21(4):650–663.
163. Wu J, Wu H, Tsai D, et al. Temporal regulation of Lsp1 O-GlcNAcylation and phosphorylation during apoptosis of activated B cells. *Nature Communications*. 2016;7(1):12526.
164. Wu Y, Zhan L, Ai Y, et al. MAPKAPK2-mediated LSP1 phosphorylation and FMLP-induced neutrophil polarization. *Biochem Biophys Res Commun*. 2007;358(1):170–175.
165. Maxeiner S, Shi N, Schalla C, et al. Crucial role for the LSP1–myosin1e bimolecular complex in the regulation of fcγ receptor–driven phagocytosis. *Mol Biol Cell*. 2015;26(9):1652–1664.
166. Sánchez-Barrena MJ, Vallis Y, Clatworthy MR, et al. Bin2 is a membrane sculpting N-BAR protein that influences leucocyte podosomes, motility and phagocytosis. *PloS one*. 2012;7(12):e52401.
167. Li C, Xie S, Zhang S, et al. BIN2 inhibition suppress ovarian cancer progression meanwhile protect ovarian function through downregulating HDAC1 and RPS6 phosphorylation respectively. *Clinical and Translational Medicine*. 2024;14(10):e70051.
168. Schroeder T, Just U. Notch signalling via RBP-J promotes myeloid differentiation. *EMBO J*. 2000.
169. Baumann M, Frye T, Naqvi T, Gomez-Cambronero J. Normal neutrophil maturation is associated with selective loss of MAP kinase activation by G-CSF. *Leuk Res*. 2005;29(1):73–78.
170. Darley RL, Burnett AK. Mutant RAS inhibits neutrophil but not macrophage differentiation and allows continued growth of neutrophil precursors. *Exp Hematol*. 1999;27(11):1599–1608.
171. Fernández IF, Blanco S, Lozano J, Lazo PA. VRK2 inhibits mitogen-activated protein kinase signaling and inversely correlates with ErbB2 in human breast cancer. *Mol Cell Biol*. 2010;30(19):4687–4697.
172. Blanco S, Klimcakova L, Vega FM, Lazo PA. The subcellular localization of vaccinia-related kinase-2 (VRK2) isoforms determines their different effect on p53 stability in tumour cell lines. *The FEBS journal*. 2006;273(11):2487–2504.
173. Xu Q, Xu Y, Pan B, et al. TTK is a favorable prognostic biomarker for triple-negative breast cancer survival. *Oncotarget*. 2016;7(49):81815.
174. Leeksma AC, Derks IA, Garrick B, et al. SMG1, a nonsense-mediated mRNA decay (NMD) regulator, as a candidate therapeutic target in multiple myeloma. *Molecular oncology*. 2023;17(2):284–297.
175. Song J, Lan J, Tang J, Luo N. PTPN2 in the immunity and tumor immunotherapy: A concise review. *International Journal of Molecular Sciences*. 2022;23(17):10025.
176. Spalinger MR, Manzini R, Hering L, et al. PTPN2 regulates inflammasome activation and controls onset of intestinal inflammation and colon cancer. *Cell reports*. 2018;22(7):1835–1848.
177. Masaki T, Shimada M. Decoding the phosphatase code: Regulation of cell proliferation by calcineurin. *International journal of molecular sciences*. 2022;23(3):1122.
178. Potel CM, Lemeer S, Heck AJ. Phosphopeptide fragmentation and site localization by mass spectrometry: An update. *Anal Chem*. 2018;91(1):126–141.
179. Kohler D, Tsai T, Verschueren E, et al. MSstatsPTM: Statistical relative quantification of posttranslational modifications in bottom-up mass spectrometry-based proteomics. *Molecular & cellular proteomics*. 2023;22(1).
180. Lorente JS, Sokolov AV, Ferguson G, Schiöth HB, Hauser AS, Gloriam DE. GPCR drug discovery: New agents, targets and indications. *Nature Reviews Drug Discovery*. 2025:1–22.
181. Dahlgren C, Holdfeldt A, Lind S, et al. Neutrophil signaling that challenges dogmata of G protein-coupled receptor regulated functions. *ACS Pharmacology & Translational Science*. 2020;3(2):203–220.
182. Futosi K, Mócsai A. Tyrosine kinase signaling pathways in neutrophils. *Immunol Rev*. 2016;273(1):121–139.
183. Fiedler K, Sindrilaru A, Terszowski G, et al. Neutrophil development and function critically depend on bruton tyrosine kinase in a mouse model of X-linked agammaglobulinemia. *Blood, The Journal of the American Society of Hematology*. 2011;117(4):1329–1339.
184. Futosi K, Fodor S, Mócsai A. Reprint of neutrophil cell surface receptors and their intracellular signal transduction pathways. *Int Immunopharmacol*. 2013;17(4):1185–1197.
185. Jacobs CF, Eldering E, Kater AP. Kinase inhibitors developed for treatment of hematologic malignancies: Implications for immune modulation in COVID-19. *Blood Advances*. 2021;5(3):913–925.
186. Park J, Kim S, Cho H, Lee S, Baek K. YOD1 deubiquitinates NEDD4 involved in the hippo signaling pathway. *Cellular Physiology & Biochemistry (Cell Physiol Biochem Press GmbH & Co.KG)*. 2020;54(1).
187. Kim Y, Kim W, Song Y, et al. Deubiquitinase YOD1 potentiates YAP/TAZ activities through enhancing ITCH stability. *Proceedings of the National Academy of Sciences*. 2017;114(18):4691–4696.

188. Park S, Baek K. Synergistic effect of YOD1 and USP21 on the hippo signaling pathway. *Cancer Cell International*. 2023;23(1):209.
189. Papadopoulos C, Kirchner P, Bug M, et al. VCP/p97 cooperates with YOD 1, UBXD 1 and PLAA to drive clearance of ruptured lysosomes by autophagy. *EMBO J*. 2017;36(2):135–150.
190. Ernst R, Mueller B, Ploegh HL, Schlieker C. The otubain YOD1 is a deubiquitinating enzyme that associates with p97 to facilitate protein dislocation from the ER. *Mol Cell*. 2009;36(1):28–38.
191. Sasset L, Petris G, Cesaratto F, Burrone OR. The VCP/p97 and YOD1 proteins have different substrate-dependent activities in endoplasmic reticulum-associated degradation (ERAD). *J Biol Chem*. 2015;290(47):28175–28188.
192. Wu Y, Duan Y, Han W, et al. Deubiquitinase YOD1 suppresses tumor progression by stabilizing E3 ligase TRIM33 in head and neck squamous cell carcinoma. *Cell Death & Disease*. 2023;14(8):517.
193. Koscielny G, Yaikhom G, Iyer V, et al. The international mouse phenotyping consortium web portal, a unified point of access for knockout mice and related phenotyping data. *Nucleic Acids Res*. 2014;42(D1):D802–D809.
194. Zhao J, Guo X, Li H, et al. Regulation of disease signaling by YOD1: Potential implications for therapeutic strategies. *Cancer Cell International*. 2025;25(1):232.
195. Chakraborty A, Tweardyabc DJ. Stat3 and G-CSF-induced myeloid differentiation. *Leuk Lymphoma*. 1998;30(5-6):433–442.
196. Zhang M, Meng Y, Ying Y, et al. Selective activation of STAT3 and STAT5 dictates the fate of myeloid progenitor cells. *Cell Death Discovery*. 2023;9(1):274.
197. Quinton LJ, Jones MR, Simms BT, et al. Functions and regulation of NF-κB RelA during pneumococcal pneumonia. *The Journal of Immunology*. 2007;178(3):1896–1903.
198. Wang D, Paz-Priel I, Friedman AD. NF-κB p50 regulates C/EBPα expression and inflammatory cytokine-induced neutrophil production. *The Journal of Immunology*. 2009;182(9):5757–5762.
199. Chen J, Bai Y, Xue K, et al. CREB1-driven CXCR4hi neutrophils promote skin inflammation in mouse models and human patients. *Nature Communications*. 2023;14(1):5894.
200. Yuan X, Hao X, Chan HL, et al. CREB-binding protein/P300 bromodomain inhibition reduces neutrophil accumulation and activates antitumor immunity in triple-negative breast cancer. *JCI insight*. 2024;9(20):e182621.
201. Kim IS, Gao Y, Welte T, et al. Immuno-subtyping of breast cancer reveals distinct myeloid cell profiles and immunotherapy resistance mechanisms. *Nat Cell Biol*. 2019;21(9):1113–1126.
202. Casbon A, Reynaud D, Park C, et al. Invasive breast cancer reprograms early myeloid differentiation in the bone marrow to generate immunosuppressive neutrophils. *Proceedings of the National Academy of Sciences*. 2015;112(6):E566–E575.
203. Yang J, Ge H, Poulton CJ, et al. Histone modification signature at myeloperoxidase and proteinase 3 in patients with anti-neutrophil cytoplasmic autoantibody-associated vasculitis. *Clinical epigenetics*. 2016;8(1):85.
204. Pollin G, Mathison AJ, de Assuncao TM, et al. Ehmt2 inactivation in pancreatic epithelial cells shapes the transcriptional landscape and inflammation response of the whole pancreas. *Frontiers in Genetics*. 2024;15:1412767.
205. Park S, Kim J. A short guide to histone deacetylases including recent progress on class II enzymes. *Exp Mol Med*. 2020;52(2):204–212.
206. Tan J, Yan Y, Wang X, Jiang Y, Xu HE. EZH2: Biology, disease, and structure-based drug discovery. *Acta Pharmacol Sin*. 2014;35(2):161–174.
207. Gunawan M, Venkatesan N, Loh JT, et al. The methyltransferase Ezh2 controls cell adhesion and migration through direct methylation of the extranuclear regulatory protein talin. *Nat Immunol*. 2015;16(5):505–516.
208. Chen Y, Gong W, Dai W, Jiang H, Xu X. E2F1/2/4 mRNA is associated with immune infiltration and are potential biomarkers for the prognosis of human gastric carcinoma. *Translational Cancer Research*. 2021;10(6):2801.
209. Wang H, Guo M, Wei H, Chen Y. Targeting p53 pathways: Mechanisms, structures and advances in therapy. *Signal transduction and targeted therapy*. 2023;8(1):92.
210. O'connor L, Gilmour J, Bonifer C. The role of the ubiquitously expressed transcription factor Sp1 in tissue-specific transcriptional regulation and in disease. *Yale J Biol Med*. 2016;89(4):513.
211. Hoogendijk AJ, Pourfarzad F, Aarts CE, et al. Dynamic transcriptome-proteome correlation networks reveal human myeloid differentiation and neutrophil-specific programming. *Cell reports*. 2019;29(8):2505–2519. e4.
212. Pinu FR, Beale DJ, Paten AM, et al. Systems biology and multi-omics integration: Viewpoints from the metabolomics research community. *Metabolites*. 2019;9(4):76.
213. Abderrazak A, Syrovets T, Couchie D, et al. NLRP3 inflammasome: From a danger signal sensor to a regulatory node of oxidative stress and inflammatory diseases. *Redox biology*. 2015;4:296–307.
214. Van Bruggen S, Jarrot P, Thomas E, et al. NLRP3 is essential for neutrophil polarization and chemotaxis in response to leukotriene B4 gradient. *Proceedings of the National Academy of Sciences*. 2023;120(35):e2303814120.
215. Fukui S, Fukui S, Van Bruggen S, et al. NLRP3 inflammasome activation in neutrophils directs early inflammatory response in murine peritonitis. *Scientific Reports*. 2022;12(1):21313.
216. Münzer P, Negro R, Fukui S, et al. NLRP3 inflammasome assembly in neutrophils is supported by PAD4 and promotes NETosis under sterile conditions. *Frontiers in Immunology*. 2021;12:683803.
217. Kelley N, Jeltema D, Duan Y, He Y. The NLRP3 inflammasome: An overview of mechanisms of activation and regulation. *International journal of molecular sciences*. 2019;20(13):3328.
218. Dubyak GR, Miller BA, Pearlman E. Pyroptosis in neutrophils: Multimodal integration of inflammasome and regulated cell death signaling pathways. *Immunol Rev*. 2023;314(1):229–249.

219. Lundgren-Åkerlund E, Olofsson AM, Berger E, Arfors K. CD1 1b/CD18-Dependent polymorphonuclear leucocyte interaction with matrix proteins in adhesion and migration. *Scand J Immunol*. 1993;37(5):569–574.
220. Ng LG, Liu Z, Kwok I, Ginhoux F. Origin and heterogeneity of tissue myeloid cells: A focus on GMP-derived monocytes and neutrophils. *Annu Rev Immunol*. 2023;41(1):375–404.
221. Yang R, Chen Z, Liang L, et al. Fc fragment of IgE receptor ig (FCER1G) acts as a key gene involved in cancer immune infiltration and tumour microenvironment. *Immunology*. 2023;168(2):302–319.
222. Mócsai A, Abram CL, Jakus Z, Hu Y, Lanier LL, Lowell CA. Integrin signaling in neutrophils and macrophages uses adaptors containing immunoreceptor tyrosine-based activation motifs. *Nat Immunol*. 2006;7(12):1326–1333.
223. Zhou X, Zheng Y. Cell type-specific signaling function of RhoA GTPase: Lessons from mouse gene targeting. *J Biol Chem*. 2013;288(51):36179–36188.
224. Gorla M, Guleria DS. Rho GTPase signaling: A molecular switchboard for regulating the actin cytoskeleton in axon guidance. *J Cell Physiol*. 2025;240(1):e70005.
225. Jennings RT, Strengert M, Hayes P, et al. RhoA determines disease progression by controlling neutrophil motility and restricting hyperresponsiveness. *Blood, The Journal of the American Society of Hematology*. 2014;123(23):3635–3645.
226. Kolaczkowska E. Riding the RhoA rodeo: The role of Myo9b in neutrophils. *Blood*. 2025;146(10):1150–1152.
227. Hayashi H, Cherpokova D, Martinod K, et al. Sirt3 deficiency does not affect venous thrombosis or NETosis despite mild elevation of intracellular ROS in platelets and neutrophils in mice. *PLoS One*. 2017;12(12):e0188341.
228. Kim TS, Jin YB, Kim YS, et al. SIRT3 promotes antimycobacterial defenses by coordinating mitochondrial and autophagic functions. *Autophagy*. 2019;15(8):1356–1375.
229. Abdel Khalek W, Cortade F, Ollendorff V, et al. SIRT3, a mitochondrial NAD -dependent deacetylase, is involved in the regulation of myoblast differentiation. *PLoS One*. 2014;9(12):e114388.
230. Riffelmacher T, Clarke A, Richter FC, et al. Autophagy-dependent generation of free fatty acids is critical for normal neutrophil differentiation. *Immunity*. 2017;47(3):466–480. e5.
231. Tanimura A, Miyoshi K, Horiguchi T, Hagita H, Fujisawa K, Noma T. Mitochondrial activity and unfolded protein response are required for neutrophil differentiation. *Cellular Physiology and Biochemistry*. 2018;47(5):1936–1950.
232. Liu W, Hsu AY, Wang Y, et al. Mitofusin-2 regulates leukocyte adhesion and β2 integrin activation. *J Leukoc Biol*. 2022;111(4):771–791.
233. Cao Z, Zhao M, Sun H, Hu L, Chen Y, Fan Z. Roles of mitochondria in neutrophils. *Frontiers in immunology*. 2022;13:934444.
234. Balajthy Z, Csomós K, Vámosi G, Szántó A, Lanotte M, Fésüs L. Tissue-transglutaminase contributes to neutrophil granulocyte differentiation and functions. *Blood*. 2006;108(6):2045–2054.
235. Bae GH, Kim YS, Park JY, et al. Unique characteristics of lung-resident neutrophils are maintained by PGE2/PKA/Tgm2-mediated signaling. *Blood, The Journal of the American Society of Hematology*. 2022;140(8):889–899.
236. Pacesa M, Pelea O, Jinek M. Past, present, and future of CRISPR genome editing technologies. *Cell*. 2024;187(5):1076–1100.
237. Song B, Yang S, Hwang G, Yu J, Bae S. Analysis of NHEJ-based DNA repair after CRISPR-mediated DNA cleavage. *International journal of molecular sciences*. 2021;22(12):6397.
238. Liao H, Wu J, VanDusen NJ, Li Y, Zheng Y. CRISPR-Cas9-mediated homology-directed repair for precise gene editing. *Molecular Therapy Nucleic Acids*. 2024;35(4).
239. Bock C, Datlinger P, Chardon F, et al. High-content CRISPR screening. *Nature Reviews Methods Primers*. 2022;2(1):8.
240. Brändli AW. Chemical screening and toxicity testing. *Cold Spring Harbor Protocols*. 2023;2023(4):pdb. top098251.
241. Eash KJ, Greenbaum AM, Gopalan PK, Link DC. CXCR2 and CXCR4 antagonistically regulate neutrophil trafficking from murine bone marrow. *J Clin Invest*. 2010;120(7):2423–2431.
242. Zhang C, Wu S. BAP1 mutations inhibit the NF-κB signaling pathway to induce an immunosuppressive microenvironment in uveal melanoma. *Molecular Medicine*. 2023;29(1):126.
243. Figueiredo CR, Kalirai H, Sacco JJ, et al. Loss of BAP1 expression is associated with an immunosuppressive microenvironment in uveal melanoma, with implications for immunotherapy development. *J Pathol*. 2020;250(4):420–439.
244. Cigognetti M, Lonardi S, Fisogni S, et al. BAP1 (BRCA1-associated protein 1) is a highly specific marker for differentiating mesothelioma from reactive mesothelial proliferations. *Modern Pathology*. 2015;28(8):1043–1057.
245. Dubiel W, Chaithongyot S, Dubiel D, Naumann M. The COP9 signalosome: A multi-DUB complex. *Biomolecules*. 2020;10(7):1082.
246. Du W, Zhu Z, Jiang X, Kang M, Pei D. COPS6 promotes tumor progression and reduces CD8 T cell infiltration by repressing IL-6 production to facilitate tumor immune evasion in breast cancer. *Acta Pharmacol Sin*. 2023;44(9):1890–1905.
247. Wu T, Ji M, Luo L. Mechanisms and potential applications of COPS6 in pan-cancer therapy. *World Journal of Clinical Oncology*. 2024;15(3):367.
248. Wang S, Zhuo G, Wang L, et al. Computational exploration of the significance of COPS6 in cancer: Functional and clinical relevance across tumor types. *World Journal of Clinical Oncology*. 2023;14(11):479.
249. Lou L, Shen J, Zhang J, et al. VCPIP1 potentiates innate immune responses by non-catalytically reducing the ubiquitination of IRAK1/2. *Cell Reports*. 2025;44(7).
250. Xu H, Li W, Lin Y, Sheng S, Zhang Z. Deubiquitinase VCPIP1 modulates the activation of NLRP3 inflammasome in septic ARDS. *American Journal of Respiratory and Critical Care Medicine*. 2025;211(Abstracts):A7453.
251. Ying H, Zhang B, Cao G, Wang Y, Zhang X. Role for ubiquitin-specific protease 7 (USP7) in the treatment and the immune response to hepatocellular carcinoma: Potential mechanisms. *Translational Cancer Research*. 2023;12(11):3016.

252. Pozhidaeva A, Bezsonova I. USP7: Structure, substrate specificity, and inhibition. *DNA repair*. 2019;76:30–39.
253. Al-Eidan A, Wang Y, Skipp P, Ewing RM. The USP7 protein interaction network and its roles in tumorigenesis. *Genes & Diseases*. 2022;9(1):41–50.
254. Huang B, Cao D, Yuan X, et al. USP7 deubiquitinates KRAS and promotes non-small cell lung cancer. *Cell reports*. 2024;43(11).
255. Nie L, Wang C, Liu X, et al. USP7 substrates identified by proteomics analysis reveal the specificity of USP7. *Genes Dev*. 2022;36(17-18):1016–1030.
256. Palazón-Riquelme P, Worboys JD, Green J, et al. USP7 and USP47 deubiquitinases regulate NLRP3 inflammasome activation. *EMBO Rep*. 2018;19(10):e44766.
257. Li J, Zhang N, Li M, Hong T, Meng W, Ouyang T. The emerging role of OTUB2 in diseases: From cell signaling pathway to physiological function. *Frontiers in Cell and Developmental Biology*. 2022;10:820781.
258. Ren W, Xu Z, Chang Y, et al. Pharmaceutical targeting of OTUB2 sensitizes tumors to cytotoxic T cells via degradation of PD-L1. *Nature communications*. 2024;15(1):9.
259. Zhang Y, Foreman O, Wigle DA, et al. USP44 regulates centrosome positioning to prevent aneuploidy and suppress tumorigenesis. *J Clin Invest*. 2012;122(12):4362–4374.
260. Suresh B, Lee J, Kim H, Ramakrishna S. Regulation of pluripotency and differentiation by deubiquitinating enzymes. *Cell Death & Differentiation*. 2016;23(8):1257–1264.
261. Shen C, Mulder E, Buitenwerf W, et al. Nuclear segmentation facilitates neutrophil migration. *J Cell Sci*. 2023;136(11):jcs260768.
262. Manley HR, Keightley MC, Lieschke GJ. The neutrophil nucleus: An important influence on neutrophil migration and function. *Frontiers in immunology*. 2018;9:2867.
263. Zhao Y, Peng D, Liu Y, et al. Usp8 promotes tumor cell migration through activating the JNK pathway. *Cell Death & Disease*. 2022;13(3):286.
264. Liu L, Yao D, Zhang P, et al. Deubiquitinase USP9X promotes cell migration, invasion and inhibits apoptosis of human pancreatic cancer. *Oncol Rep*. 2017;38(6):3531–3537.
265. Lambies G, Miceli M, Martínez-Guillamon C, et al. TGFβ-activated USP27X deubiquitinase regulates cell migration and chemoresistance via stabilization of Snail1. *Cancer Res*. 2019;79(1):33–46.
266. Huang C. Roles of E3 ubiquitin ligases in cell adhesion and migration. *Cell adhesion & migration*. 2010;4(1):10–18.
267. Pedrazza L, Schneider T, Bartrons R, Ventura F, Rosa JL. The ubiquitin ligase HERC1 regulates cell migration via RAF-dependent regulation of MKK3/p38 signaling. *Scientific Reports*. 2020;10(1):824.
268. Wu S, Zhang S, Liu C, Fernie AR, Yan S. Recent advances in mass spectrometry-based protein interactome studies. *Molecular & Cellular Proteomics*. 2025;24(1):100887.
269. Pfeiffer CT, Paulo JA, Gygi SP, Rockman HA. Proximity labeling for investigating protein-protein interactions. In: *Methods in cell biology.* Vol 169. Elsevier; 2022:237–266.
270. Damianou A, Jones HB, Grigoriou A, et al. Integrative proximal-ubiquitomics profiling for deubiquitinase substrate discovery applied to USP30. *Cell Chemical Biology*. 2025;32(5):736–751. e8.
271. Keijzer N, Priyanka A, Stijf-Bultsma Y, Fish A, Gersch M, Sixma TK. Variety in the USP deubiquitinase catalytic mechanism. *Life Science Alliance*. 2024;7(4).
272. Jové V, Wheeler H, Lee CW, et al. Type I interferon regulation by USP18 is a key vulnerability in cancer. *Iscience*. 2024;27(4).
273. French ME, Koehler CF, Hunter T. Emerging functions of branched ubiquitin chains. *Cell discovery*. 2021;7(1):6.
274. Kadimisetty K, Sheets KJ, Gross PH, Zerr MJ, Ouazia D. Tandem ubiquitin binding entities (TUBEs) as tools to explore ubiquitin-proteasome system and PROTAC drug discovery. In: *Targeted protein degradation: Methods and protocols.* Springer; 2021:185–202.
275. Swatek KN, Usher JL, Kueck AF, et al. Insights into ubiquitin chain architecture using ub-clipping. *Nature*. 2019;572(7770):533–537.
276. Han Y, Zhong L, Ren F. A simple method for the preparation of positive samples to preliminarily determine the quality of phosphorylation-specific antibody. *Plos one*. 2022;17(7):e0272138.
277. Liu J, Lu Y, Zhu R, et al. The deubiquitinase YOD1 suppresses tumor progression by stabilizing ZNF24 in clear cell renal carcinoma. *Cell Death & Disease*. 2025;16(1):334.
278. Qiang JU, Li MX, Gang C, et al. Overexpression of YOD1 promotes the migration of human oral keratinocytes by enhancing TGF-β3 signaling. *Biomedical and Environmental Sciences*. 2018;31(7):499–506.
279. Liu Y, Yang J, Wang T, et al. Expanding PROTACtable genome universe of E3 ligases. *Nature communications*. 2023;14(1):6509.
280. Duong-Ly KC, Peterson JR. The human kinome and kinase inhibition. *Current protocols in pharmacology*. 2013;60(1):2.9. 1–2.9. 14.
281. Sacco F, Perfetto L, Castagnoli L, Cesareni G. The human phosphatase interactome: An intricate family portrait. *FEBS Lett*. 2012;586(17):2732–2739.
282. Liang Z, Damianou A, Vendrell I, et al. Proximity proteomics reveals UCH-L1 as an essential regulator of NLRP3-mediated IL-1β production in human macrophages and microglia. *Cell Reports*. 2024;43(5).

Appendices

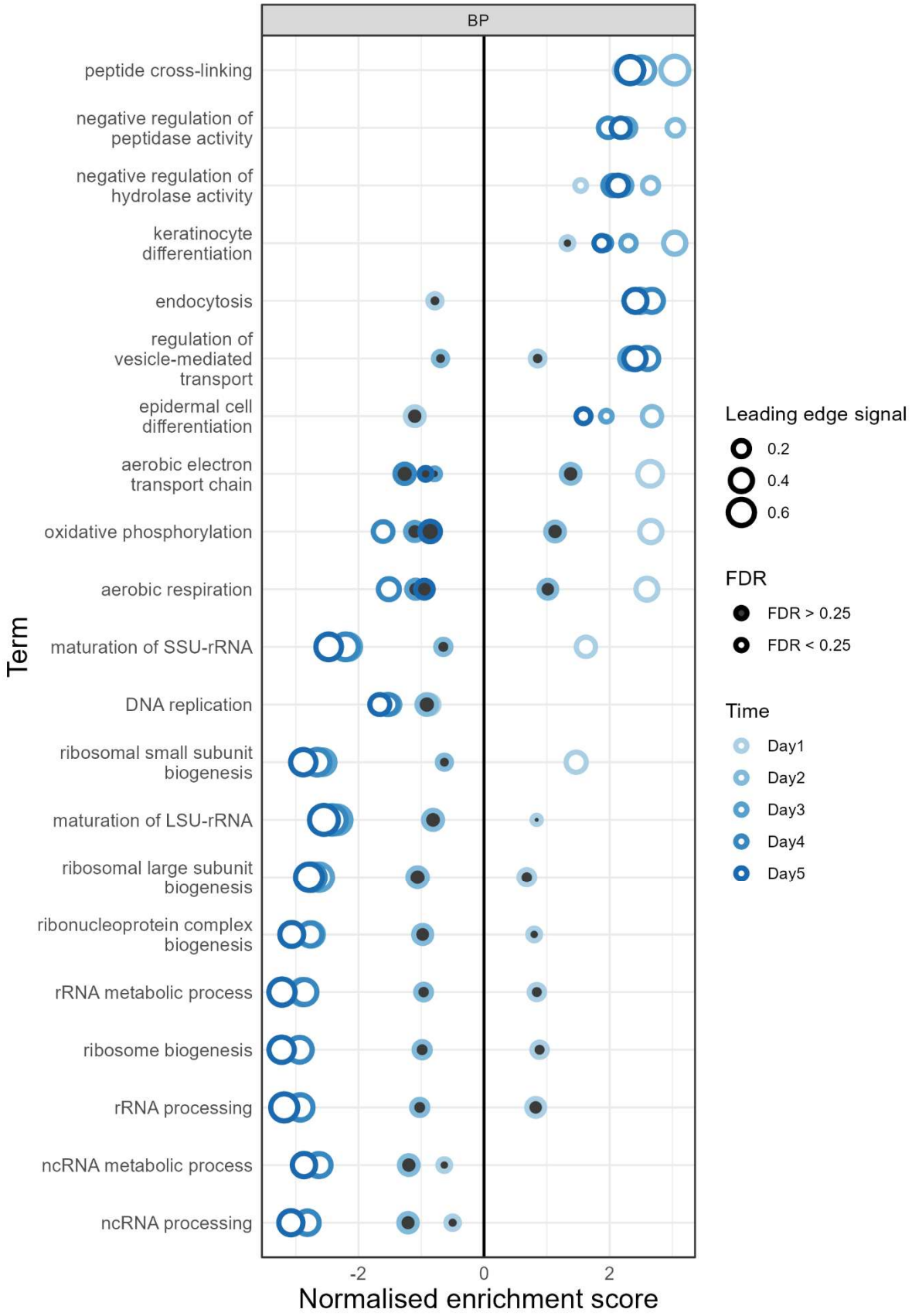
Chapter 2

RStudio session information and R packages used in this thesis are shown below.

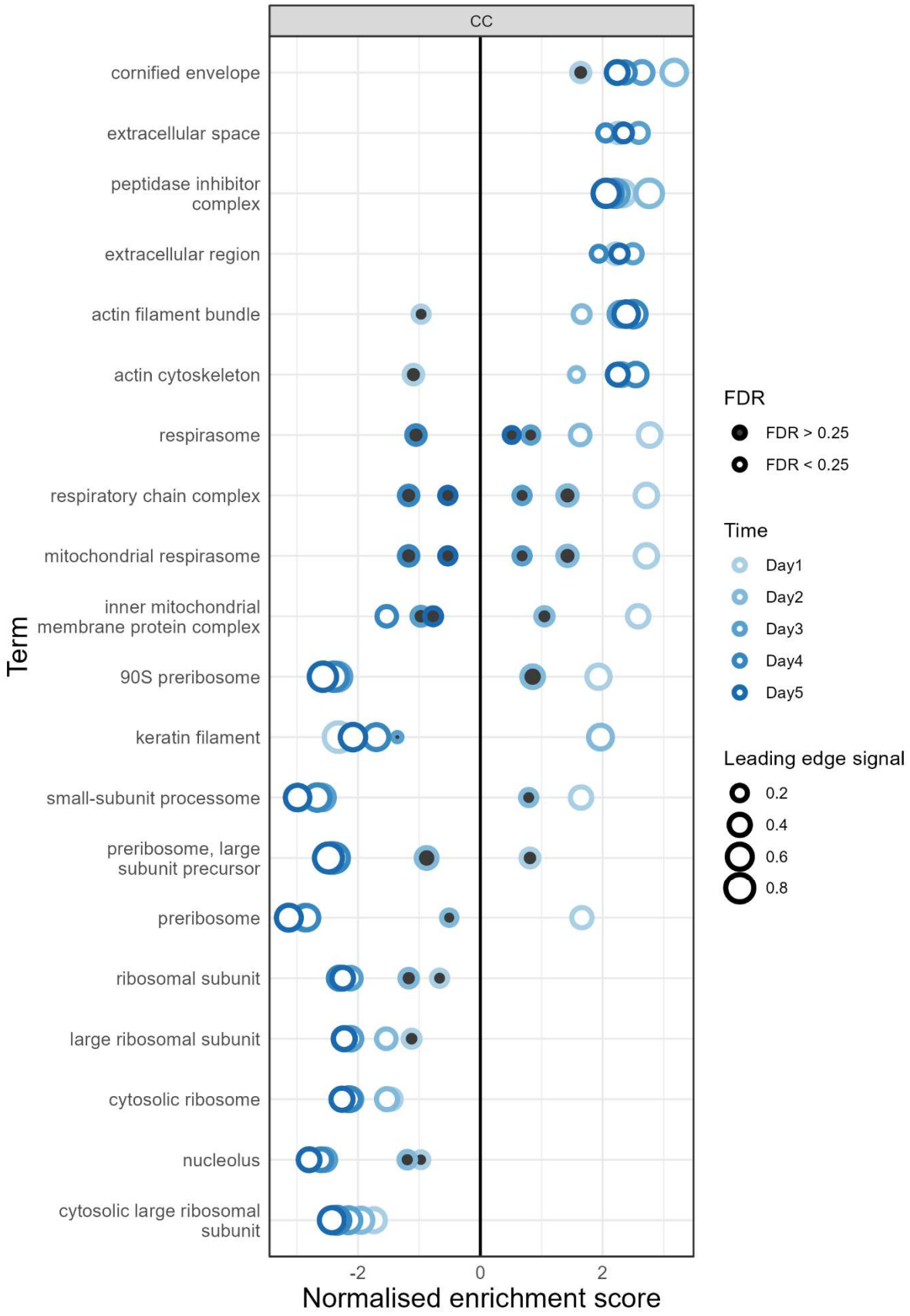
```
R version 4.3.2 (2023-10-31 ucrt)
Platform: x86_64-w64-mingw32/x64 (64-bit)
Running under: Windows 11 x64 (build 26100)

Matrix products: default


locale:
[1] LC_COLLATE=English_United Kingdom.utf8 LC_CTYPE=English_United
Kingdom.utf8 LC_MONETARY=English_United Kingdom.utf8
[4] LC_NUMERIC=C       LC_TIME=English_United Kingdom.utf8

time zone: Europe/London
tzcode source: internal

attached base packages:
[1] stats  graphics grDevices utils  datasets methods base

other attached packages:
 [1] WGCNA_1.73    fastcluster_1.3.0  dynamicTreeCut_1.63-1 readxl_1.4.5
patchwork_1.3.0
 [6] lubridate_1.9.4  tidyverse_2.0.0  purrr_1.0.4   ggsignif_0.6.4
readr_2.1.5
[11] reshape2_1.4.4   ggnewscale_0.5.2  ggprism_1.0.7   DOSE_3.28.2
ggpubr_0.6.0
[16] pcds_0.1.8    MKmisc_1.9    clusterProfiler_4.10.1 limma_3.58.1
rstatix_0.7.2
[21] FactoMineR_2.11  factoextra_1.0.7  impute_1.76.0   tibble_3.2.1
rlang_1.1.3
[26] RColorBrewer_1.1-3  zoo_1.8-14    gtools_3.9.5   data.table_1.17.2
forcats_1.0.0
[31] ggh4x_0.3.0   ggplot2_3.5.2   tidyr_1.3.1   dplyr_1.1.4   plyr_1.8.9
[36] stringr_1.5.1

loaded via a namespace (and not attached):
 [1] splines_4.3.2   bitops_1.0-9   ggplotify_0.1.2   cellranger_1.1.0
polyclip_1.10-7
 [6] preprocessCore_1.64.0 rpart_4.1.21   lifecycle_1.0.4   Rdpack_2.6.4
tcltk_4.3.2
 [11] doParallel_1.0.17  lattice_0.21-9   MASS_7.3-60    flashClust_1.01-2
backports_1.5.0
 [16] magrittr_2.0.3   rmarkdown_2.29   Hmisc_5.2-3    plot3D_1.4.2
cowplot_1.2.0
 [21] DBI_1.2.3    abind_1.4-8    zlibbioc_1.48.2   ggraph_2.2.1
BiocGenerics_0.48.1
 [26] RCurl_1.98-1.17   nnet_7.3-19    yulab.utils_0.2.1  tweenr_2.0.3
rappdirs_0.3.3
```

```
 [31] misc3d_0.9-1   GenomeInfoDbData_1.2.11 IRanges_2.36.0
S4Vectors_0.40.2  enrichplot_1.22.0
 [36] ggrepel_0.9.6   tidytree_0.4.6   codetools_0.2-19  DT_0.33
ggforce_0.4.2
 [41] tidyselect_1.2.1  aplot_0.2.8    farver_2.1.2   viridis_0.6.5
base64enc_0.1-3
 [46] matrixStats_1.5.0  stats4_4.3.2   jsonlite_2.0.0   tidygraph_1.3.1
Formula_1.2-5
 [51] survival_3.5-7   iterators_1.0.14  emmeans_1.11.2
systemfonts_1.2.3  foreach_1.5.2
 [56] tools_4.3.2    treeio_1.26.0   Rcpp_1.0.14    glue_1.7.0
gridExtra_2.3
 [61] xfun_0.53    qvalue_2.34.0   GenomeInfoDb_1.38.8  withr_3.0.2
fastmap_1.2.0
 [66] digest_0.6.37   timechange_0.3.0  R6_2.6.1    gridGraphics_0.5-1
estimability_1.5.1
 [71] colorspace_2.1-1  GO.db_3.18.0   RSQLite_2.3.11   generics_0.1.4
robustbase_0.99-4-1
 [76] graphlayouts_1.2.2  httr_1.4.7    htmlwidgets_1.6.4
scatterpie_0.2.5  scatterplot3d_0.3-44
 [81] pkgconfig_2.0.3   gtable_0.3.6   blob_1.2.4    S7_0.2.0
XVector_0.42.0
 [86] shadowtext_0.1.6  htmltools_0.5.8.1  carData_3.0-5   fgsea_1.28.0
multcompView_0.1-10
 [91] scales_1.4.0   Biobase_2.62.0   leaps_3.2    png_0.1-8
ggfun_0.2.0
 [96] knitr_1.50    rstudioapi_0.17.1  tzdb_0.5.0    uuid_1.2-1
checkmate_2.3.2
[101] coda_0.19-4.1   nlme_3.1-163   cachem_1.1.0   parallel_4.3.2
HDO.db_0.99.1
[106] foreign_0.8-85   arrow_21.0.0.1   AnnotationDbi_1.64.1 pillar_1.11.0
grid_4.3.2
[111] vctrs_0.6.5    car_3.1-3    xtable_1.8-4   cluster_2.1.4
htmlTable_2.4.3
[116] evaluate_1.0.4   mvtnorm_1.3-3   cli_3.6.2    compiler_4.3.2
crayon_1.5.3
[121] interp_1.1-6   fs_1.6.6    ggiraph_0.9.0   stringi_1.8.7
viridisLite_0.4.2
[126] deldir_2.0-4   BiocParallel_1.36.0  assertthat_0.2.1
Biostrings_2.70.3  lazyeval_0.2.2
[131] GOSemSim_2.28.1   Matrix_1.6-1.1   hms_1.1.3    bit64_4.6.0-1
KEGGREST_1.42.0
[136] statmod_1.5.0   rbibutils_2.3   igraph_2.1.4   broom_1.0.9
memoise_2.0.1
[141] ggtree_3.10.1   fastmatch_1.1-6   DEoptimR_1.1-4   bit_4.6.0
ape_5.8-1
[146] gson_0.1.0
```

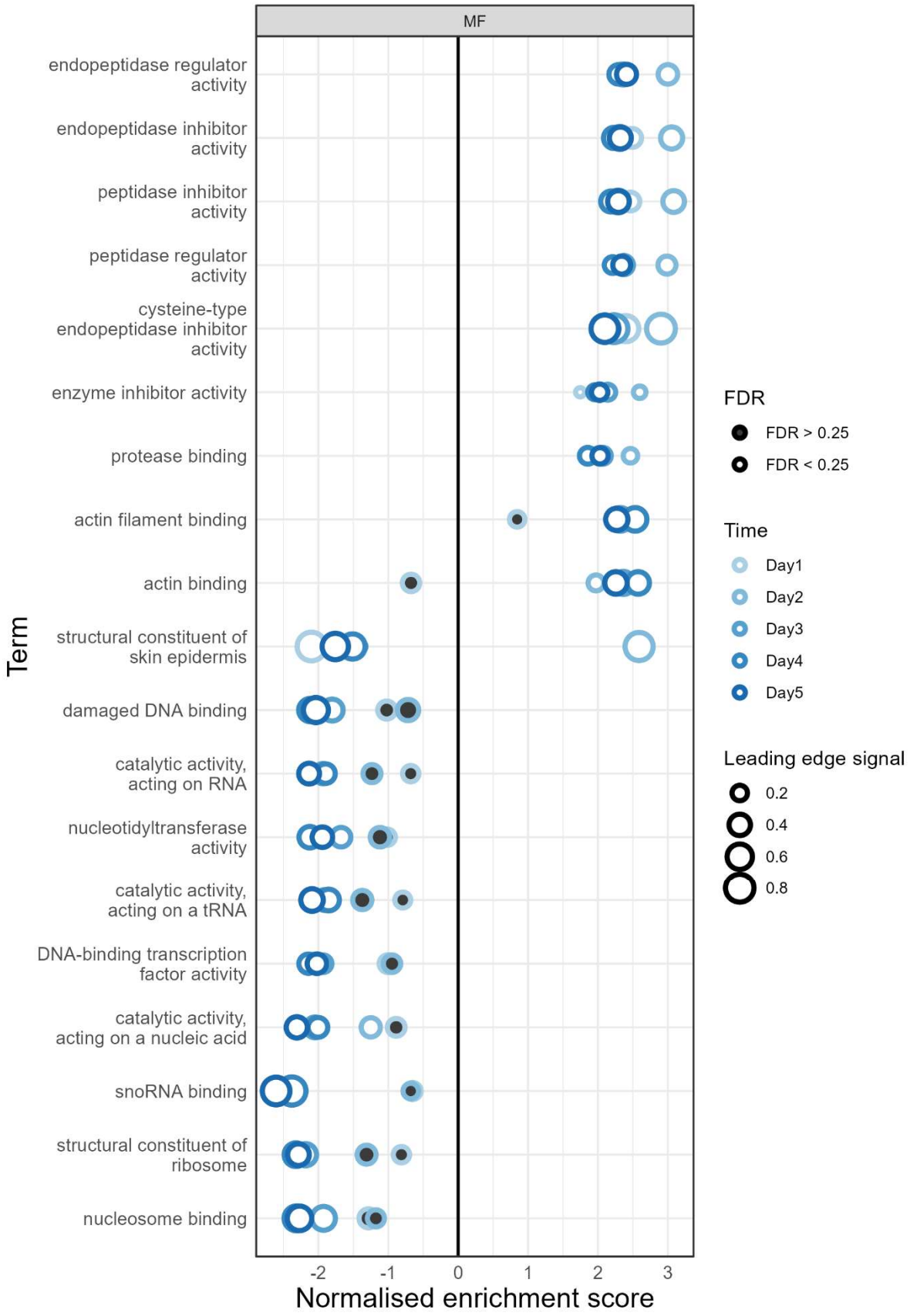
Chapter 3



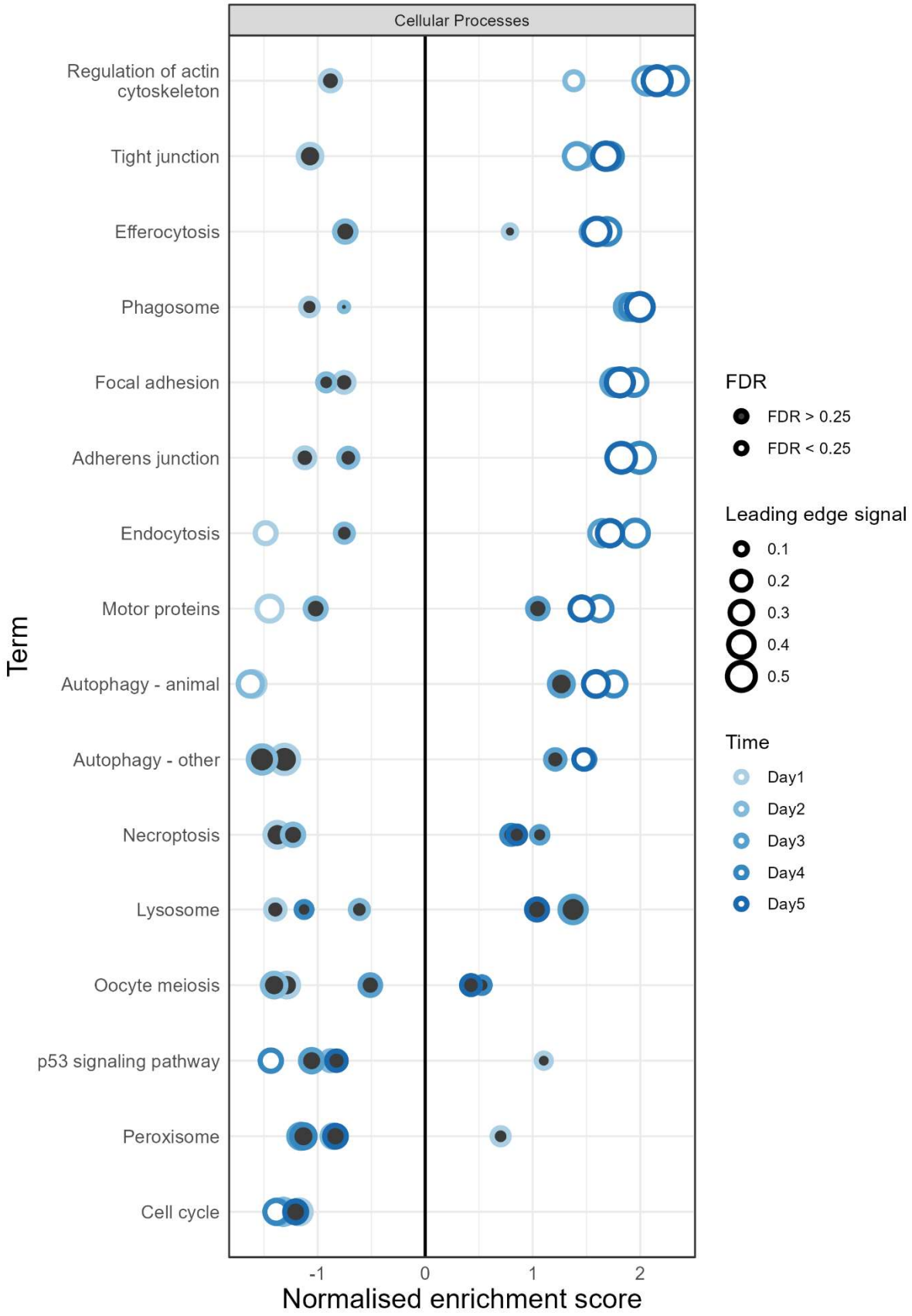
Supplementary Figure 1 ***Gene set enrichment analysis (GSEA) of nuclear fraction data from HoxB8 neutrophil differentiation (GOBP).*** Genes were pre-ranked based on log2(fold change) of protein abundance compared to day 0. Stroke colour represents the time point, size indicates leading edge signal, and permutation-based FDR is indicated by fill colour. Gene ontology (GO) biological process (BP) terms are shown.



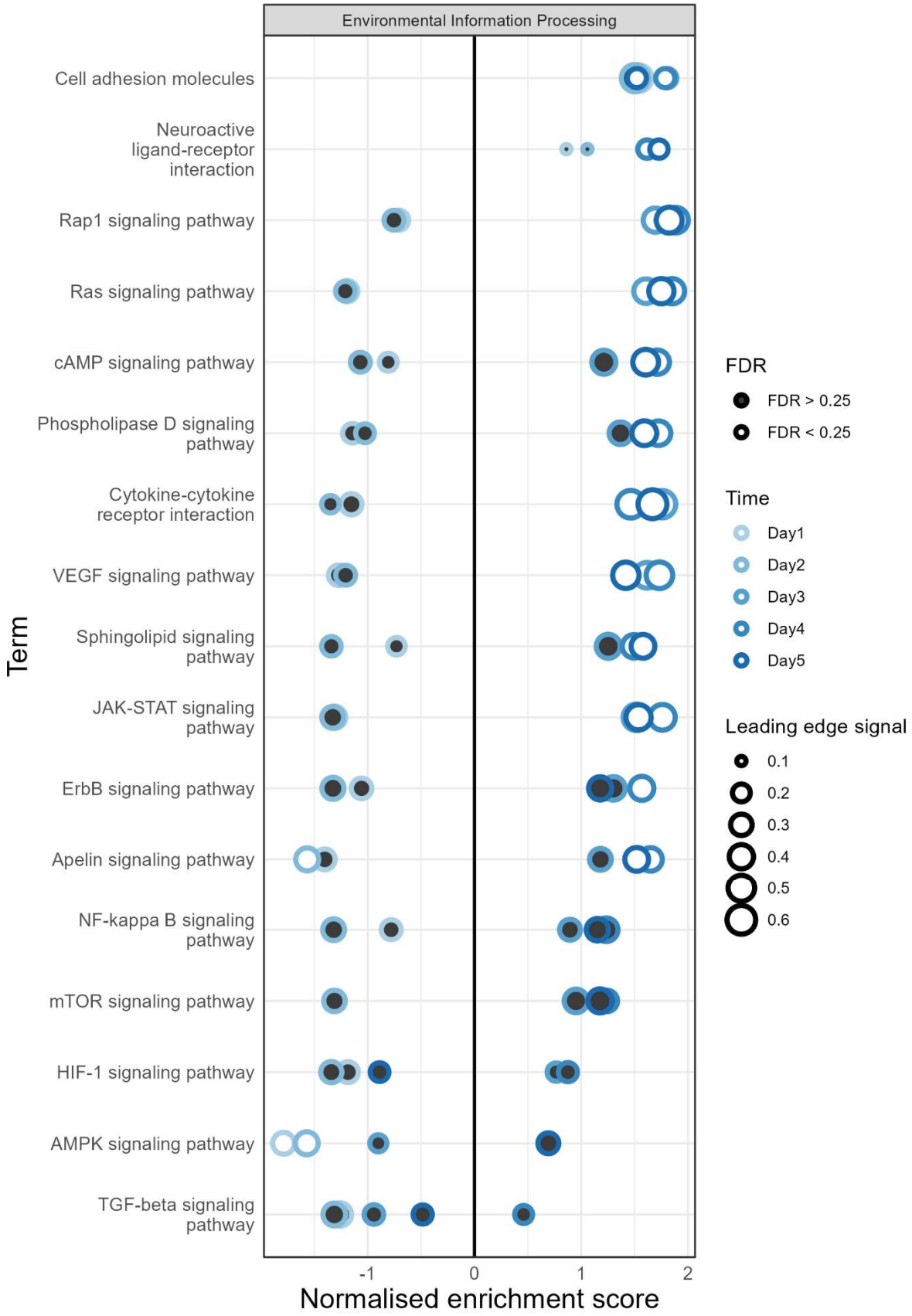
Supplementary Figure 2 ***Gene set enrichment analysis (GSEA) of nuclear fraction data from HoxB8 neutrophil differentiation (GOCC).*** Genes were pre-ranked based on log2(fold change) of protein abundance compared to day 0. Stroke colour represents the time point, size indicates leading edge signal, and permutation-based FDR is indicated by fill colour. Gene ontology (GO) cellular component (CC) terms are shown.



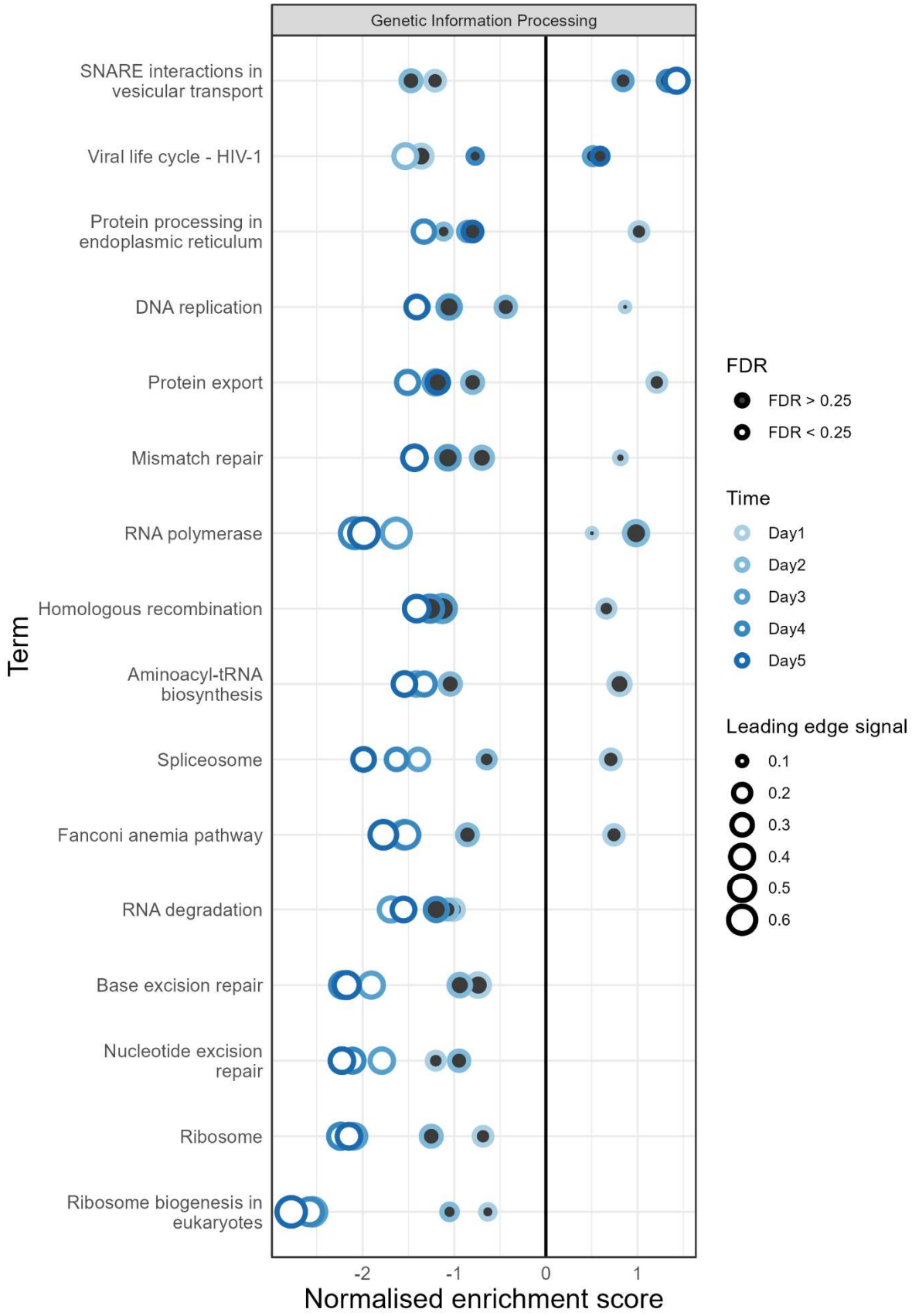
Supplementary Figure 3 ***Gene set enrichment analysis (GSEA) of nuclear fraction data from HoxB8 neutrophil differentiation (GOMF).*** Genes were pre-ranked based on log2(fold change) of protein abundance compared to day 0. Stroke colour represents the time point, size indicates leading edge signal, and permutation-based FDR is indicated by fill colour. Gene ontology (GO) molecular function (MF) terms are shown.



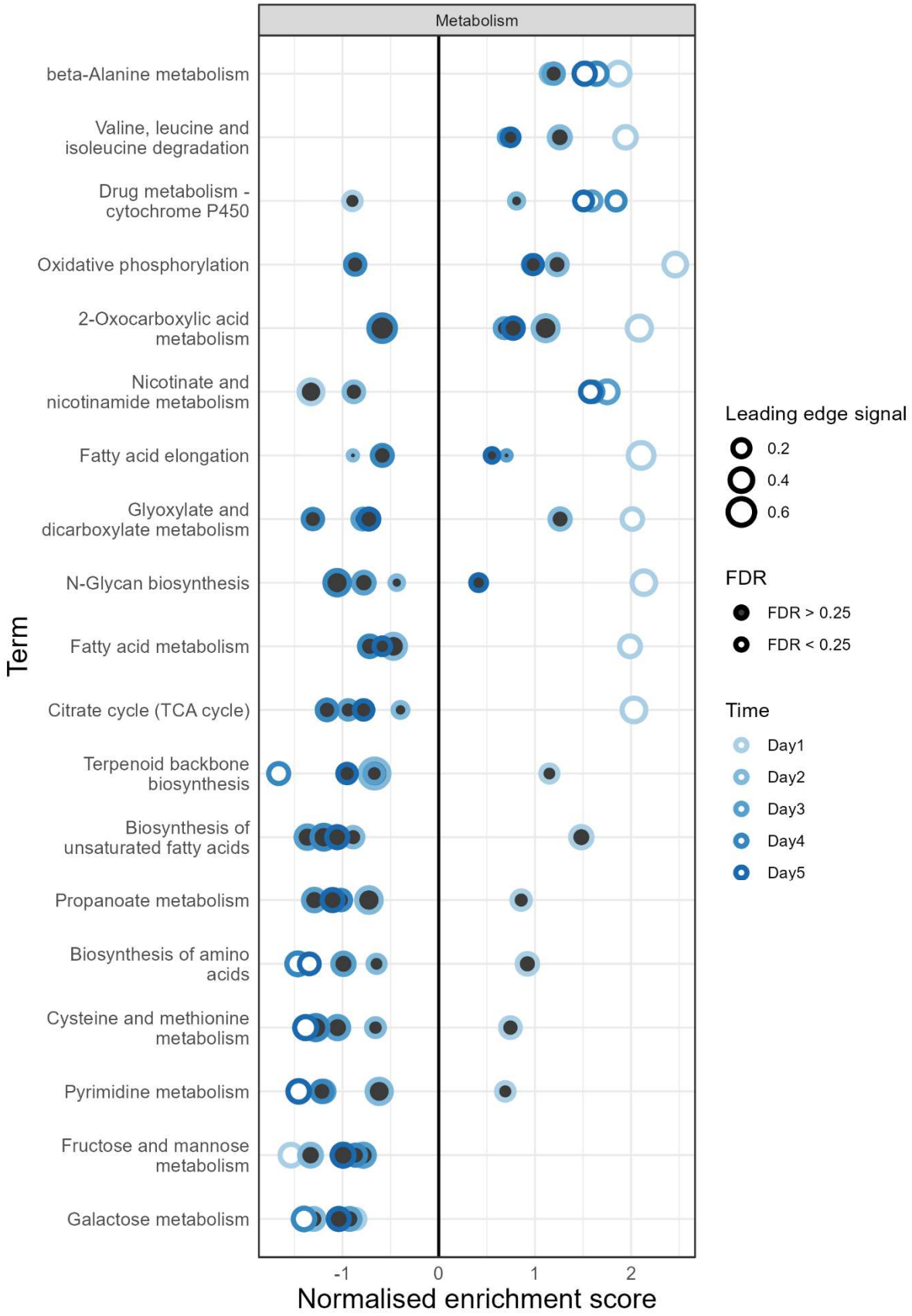
Supplementary Figure 4 ***Gene set enrichment analysis (GSEA) of nuclear fraction data from HoxB8 neutrophil differentiation (KEGG, Cellular Processes).*** Genes were pre-ranked based on log2(fold change) of protein abundance compared to day 0. Stroke colour represents the time point, size indicates leading edge signal, and permutation-based FDR is indicated by fill colour. The KEGG pathway subcategory 'Cellular Processes' terms are shown.



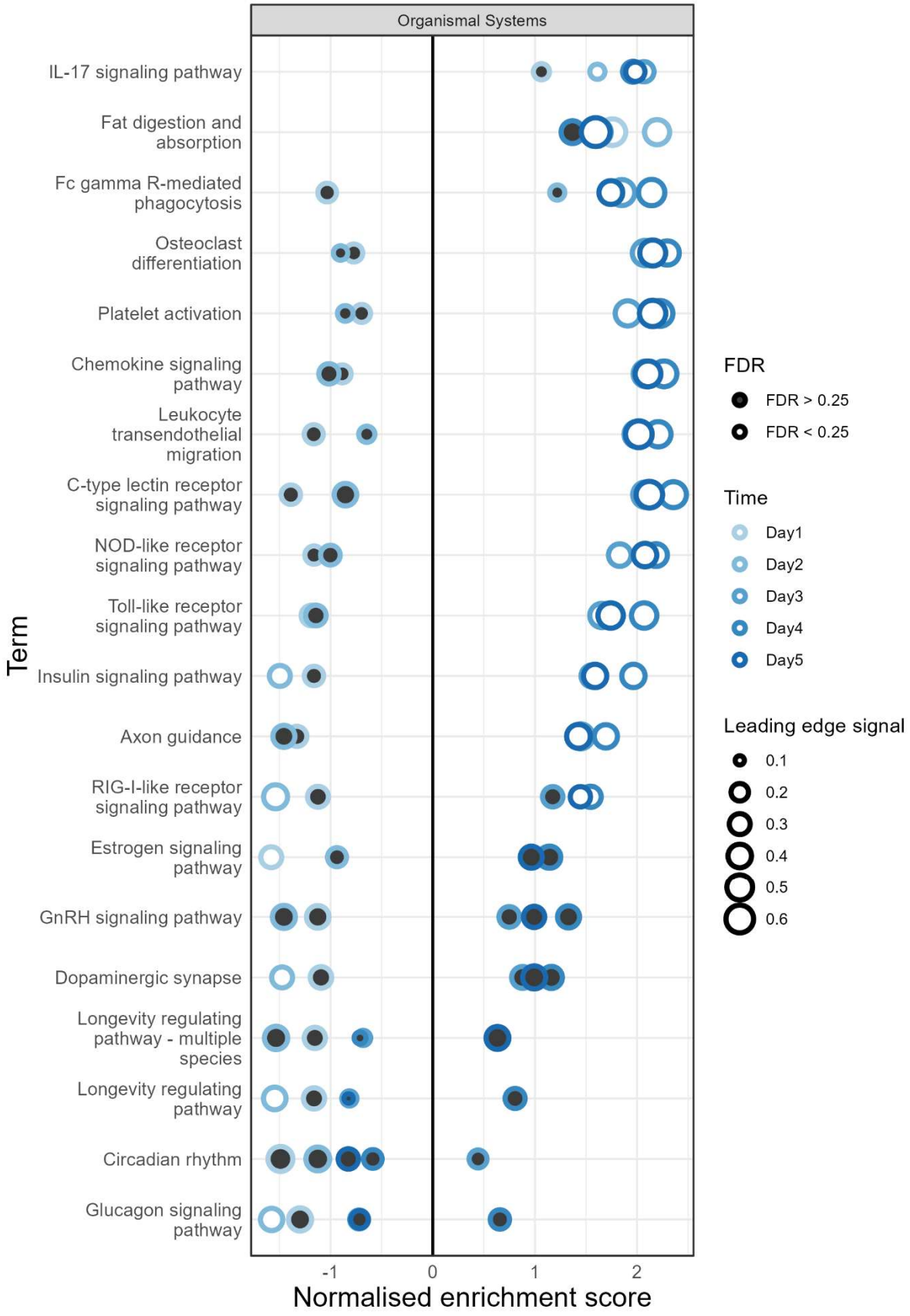
Supplementary Figure 5 ***Gene set enrichment analysis (GSEA) of nuclear fraction data from HoxB8 neutrophil differentiation (KEGG, Environmental Information Processing).*** Genes were pre-ranked based on log2(fold change) of protein abundance compared to day 0. Stroke colour represents the time point, size indicates leading edge signal, and permutation-based FDR is indicated by fill colour. The KEGG pathway subcategory 'Environmental Information Processing' terms are shown.



Supplementary Figure 6 ***Gene set enrichment analysis (GSEA) of nuclear fraction data from HoxB8 neutrophil differentiation (KEGG, Genetic Information Processing).*** Genes were pre-ranked based on log2(fold change) of protein abundance compared to day 0. Stroke colour represents the time point, size indicates leading edge signal, and permutation-based FDR is indicated by fill colour. The KEGG pathway subcategory 'Genetic Information Processing' terms are shown.



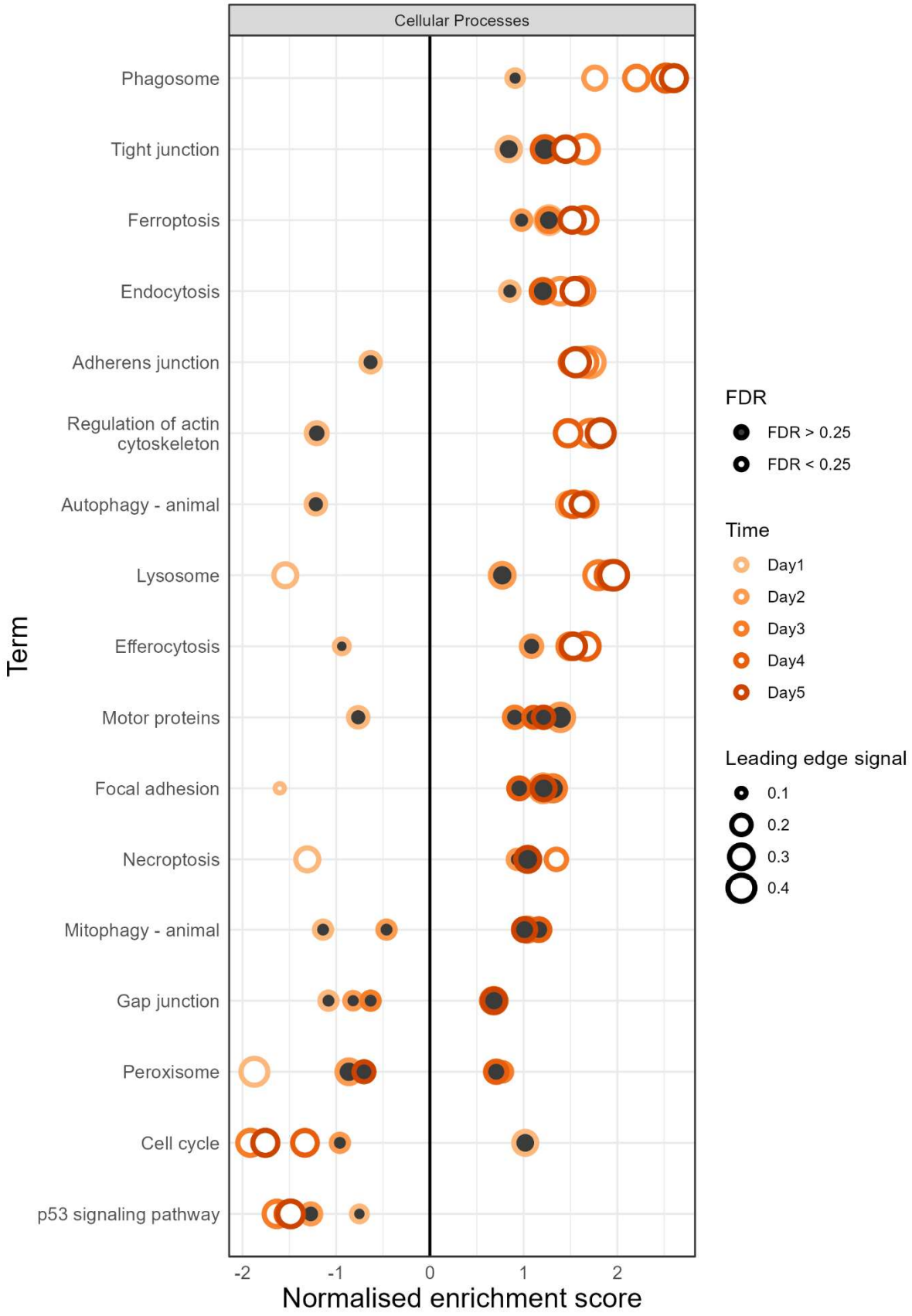
Supplementary Figure 7 ***Gene set enrichment analysis (GSEA) of nuclear fraction data from HoxB8 neutrophil differentiation (KEGG, Metabolism).*** Genes were pre-ranked based on log2(fold change) of protein abundance compared to day 0. Stroke colour represents the time point, size indicates leading edge signal, and permutation-based FDR is indicated by fill colour. The KEGG pathway subcategory 'Metabolism' terms are shown.



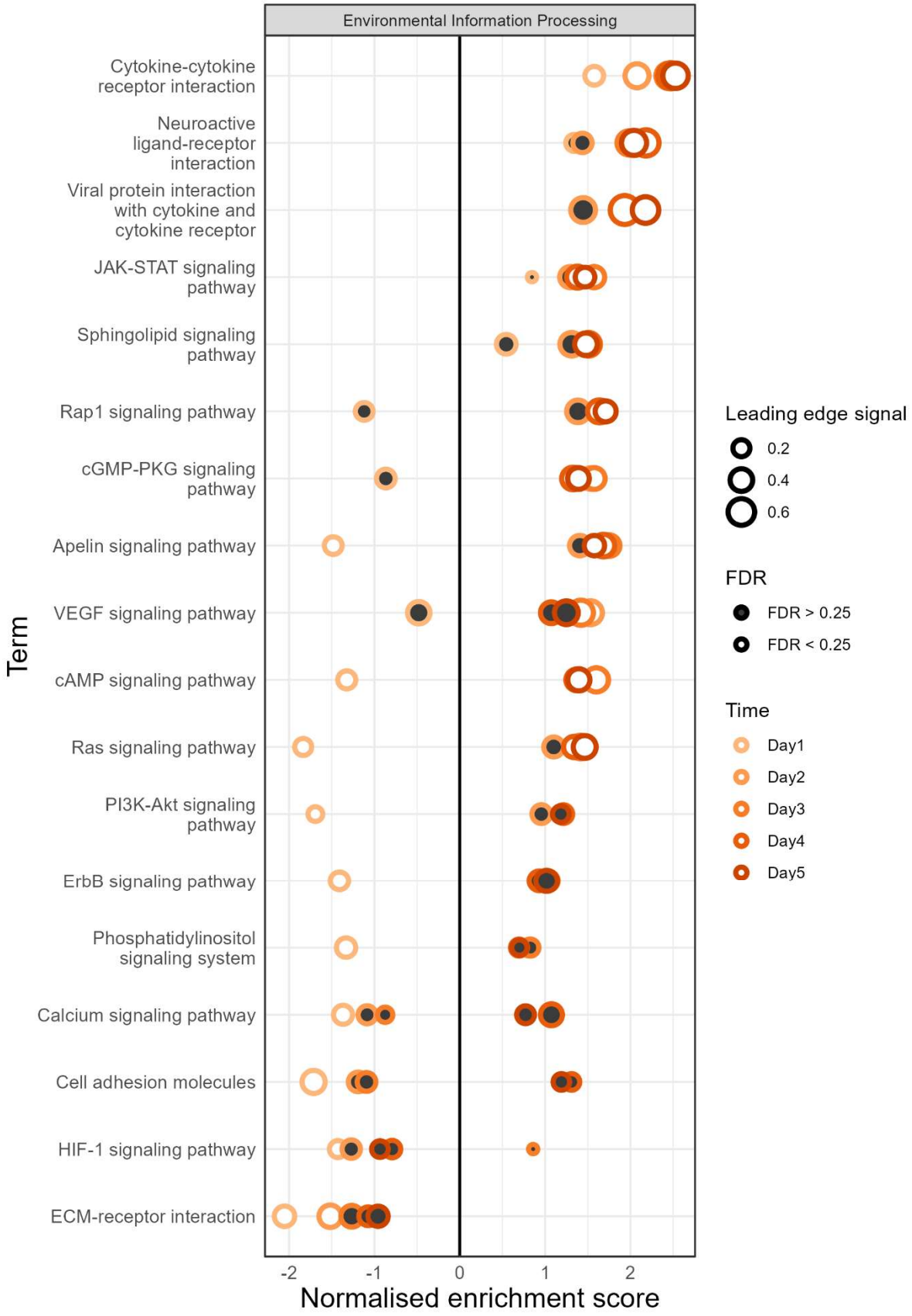
Supplementary Figure 8 ***Gene set enrichment analysis (GSEA) of nuclear fraction data from HoxB8 neutrophil differentiation (KEGG, Organismal Systems).*** Genes were pre-ranked based on log2(fold change) of protein abundance compared to day 0. Stroke colour represents the time point, size indicates leading edge signal, and permutation-based FDR is indicated by fill colour. The KEGG pathway subcategory 'Organismal Systems' terms are shown.

Supplementary Figure 9 ***Gene set enrichment analysis (GSEA) of total proteome data from HoxB8 neutrophil differentiation (GOCC).*** Genes were pre-ranked based on log2(fold change) of protein abundance compared to day 0. Stroke colour represents the time point, size indicates leading edge signal, and permutation-based FDR is indicated by fill colour. Gene ontology (GO) cellular component (CC) terms are shown.

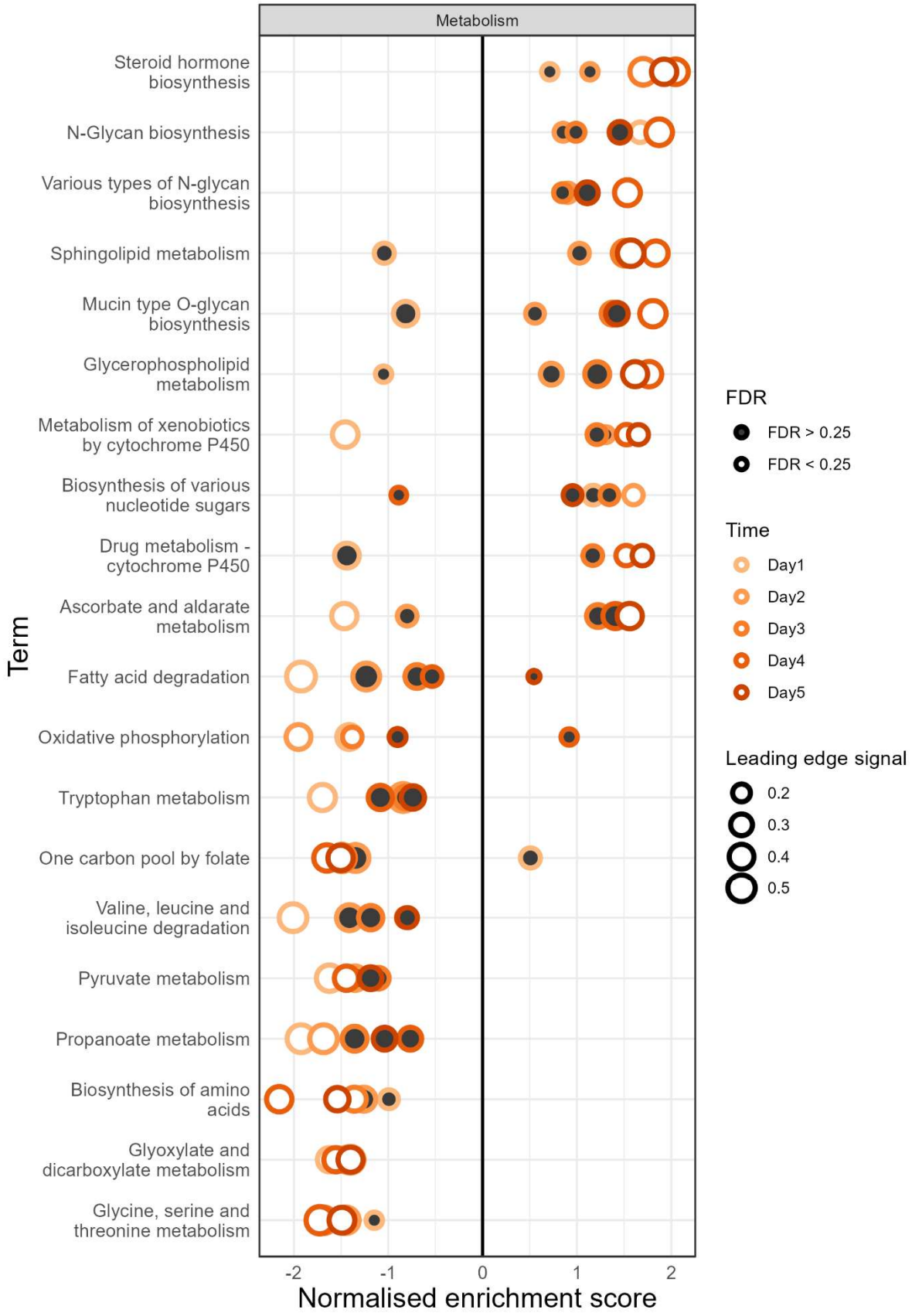
Supplementary Figure 10 ***Gene set enrichment analysis (GSEA) of total proteome data from HoxB8 neutrophil differentiation (GOMF).*** Genes were pre-ranked based on log2(fold change) of protein abundance compared to day 0. Stroke colour represents the time point, size indicates leading edge signal, and permutation-based FDR is indicated by fill colour. Gene ontology (GO) molecular function (MF) terms are shown.



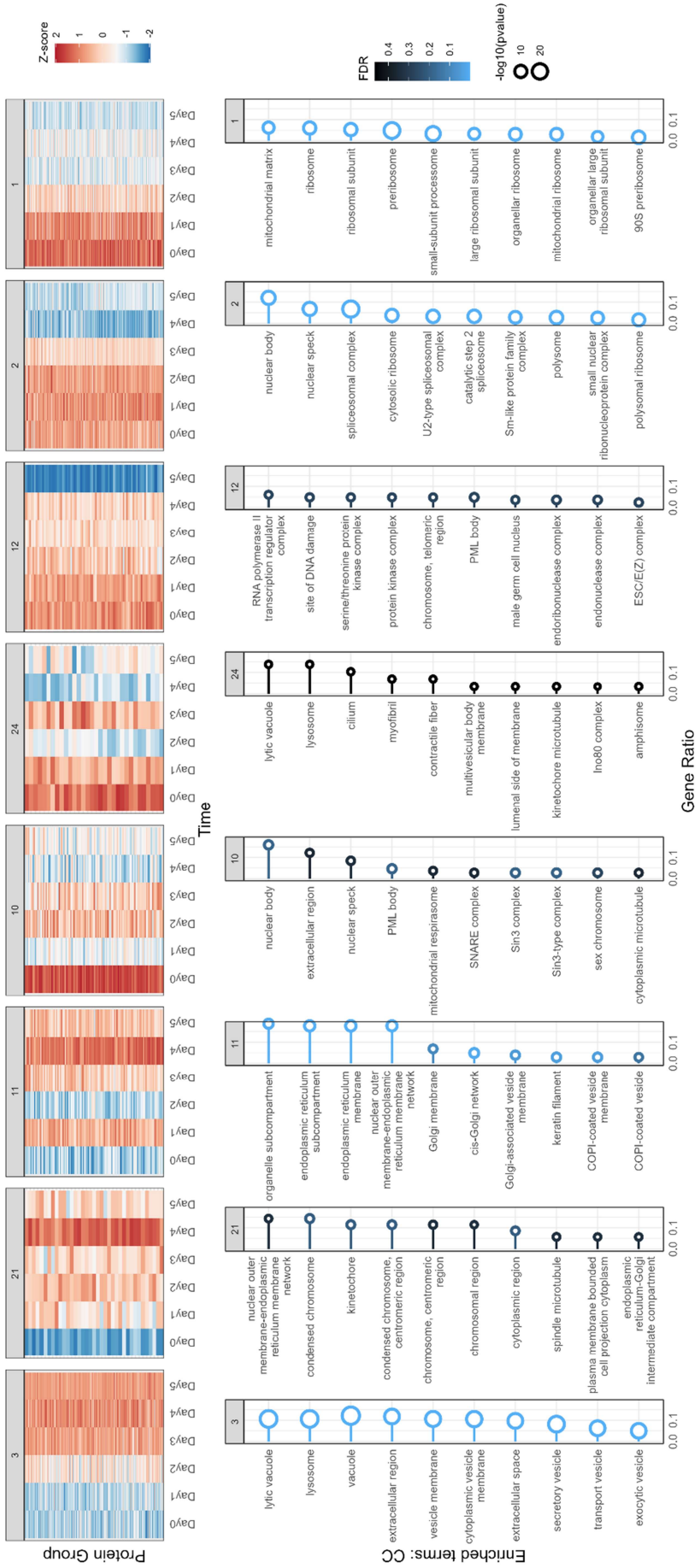
Supplementary Figure 11 ***Gene set enrichment analysis (GSEA) of total proteome data from HoxB8 neutrophil differentiation (KEGG, Cellular Processes).*** Genes were pre-ranked based on log2(fold change) of protein abundance compared to day 0. Stroke colour represents the time point, size indicates leading edge signal, and permutation-based FDR is indicated by fill colour. The KEGG pathway subcategory 'Cellular Processes' terms are shown.



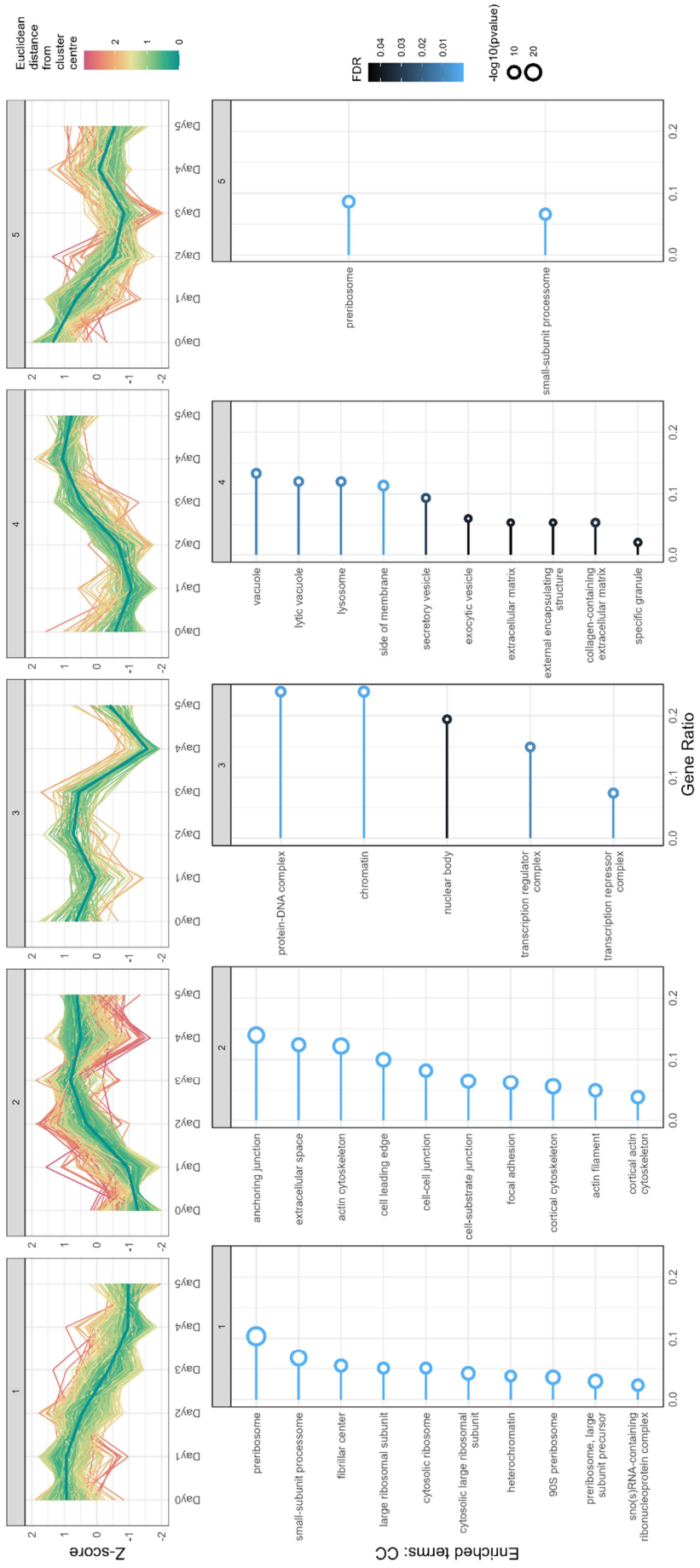
Supplementary Figure 12 ***Gene set enrichment analysis (GSEA) of total proteome data from HoxB8 neutrophil differentiation (KEGG, Environmental Information Processing).*** Genes were pre-ranked based on log2(fold change) of protein abundance compared to day 0. Stroke colour represents the time point, size indicates leading edge signal, and permutation-based FDR is indicated by fill colour.. The KEGG pathway subcategory 'Environmental Information Processing' terms are shown.



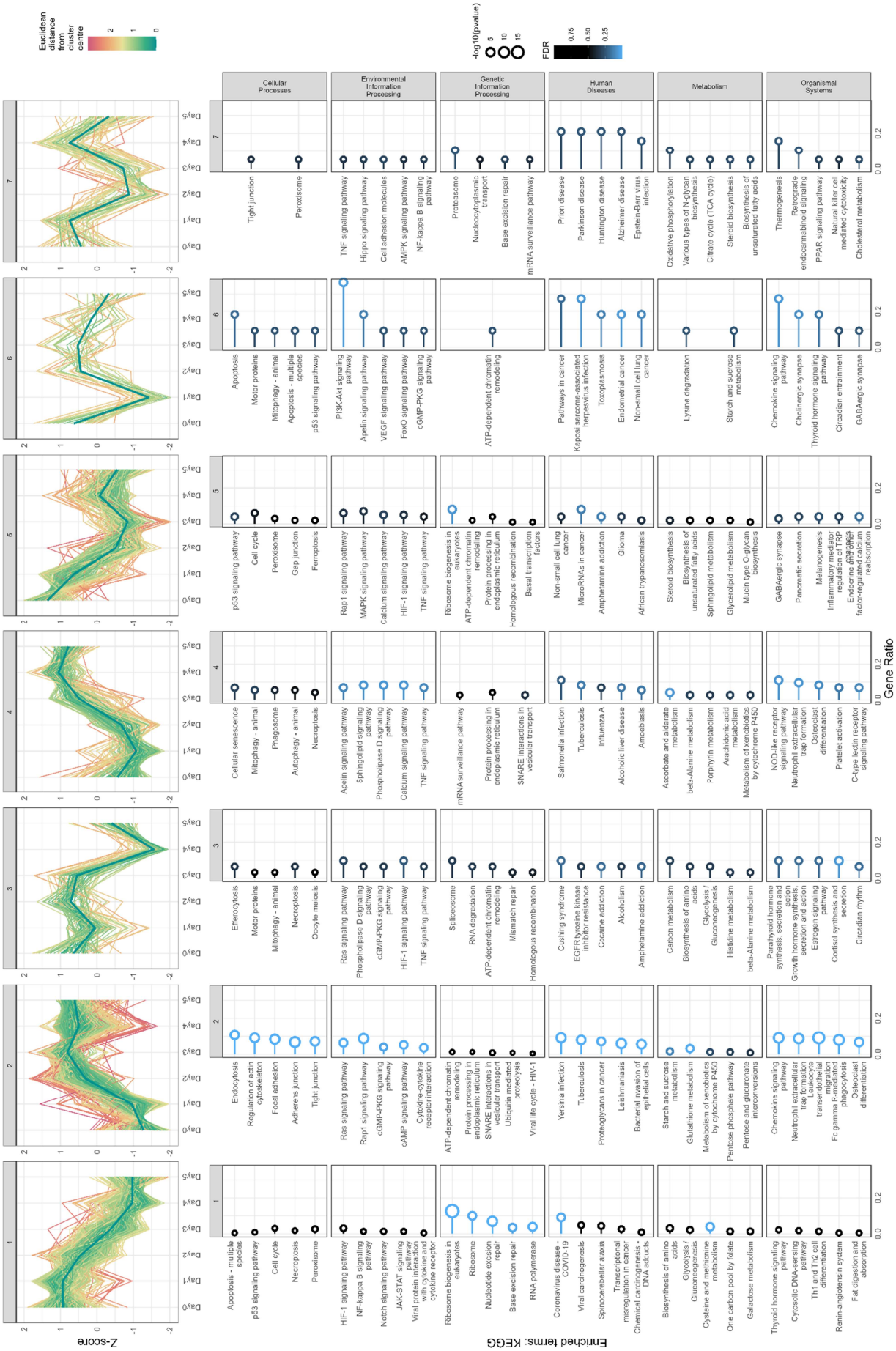
Supplementary Figure 13 ***Gene set enrichment analysis (GSEA) of total proteome data from HoxB8 neutrophil differentiation (KEGG, Metabolism).*** Genes were pre-ranked based on log2(fold change) of protein abundance compared to day 0. Stroke colour represents the time point, size indicates leading edge signal, and permutation-based FDR is indicated by fill colour. The KEGG pathway subcategory 'Metabolism' terms are shown.



Supplementary Figure 14 ***Functional enrichment analysis of select modules identified from WCNA show upregulation and downregulation of neutrophil-associated cellular components.*** Over-representation analysis was performed on each module identified by WCNA, using the total proteome as the background proteins. Point size and colour represent -log10(p-value) and FDR, respectively. Gene ontology (GO) cellular components (CC) terms are shown.

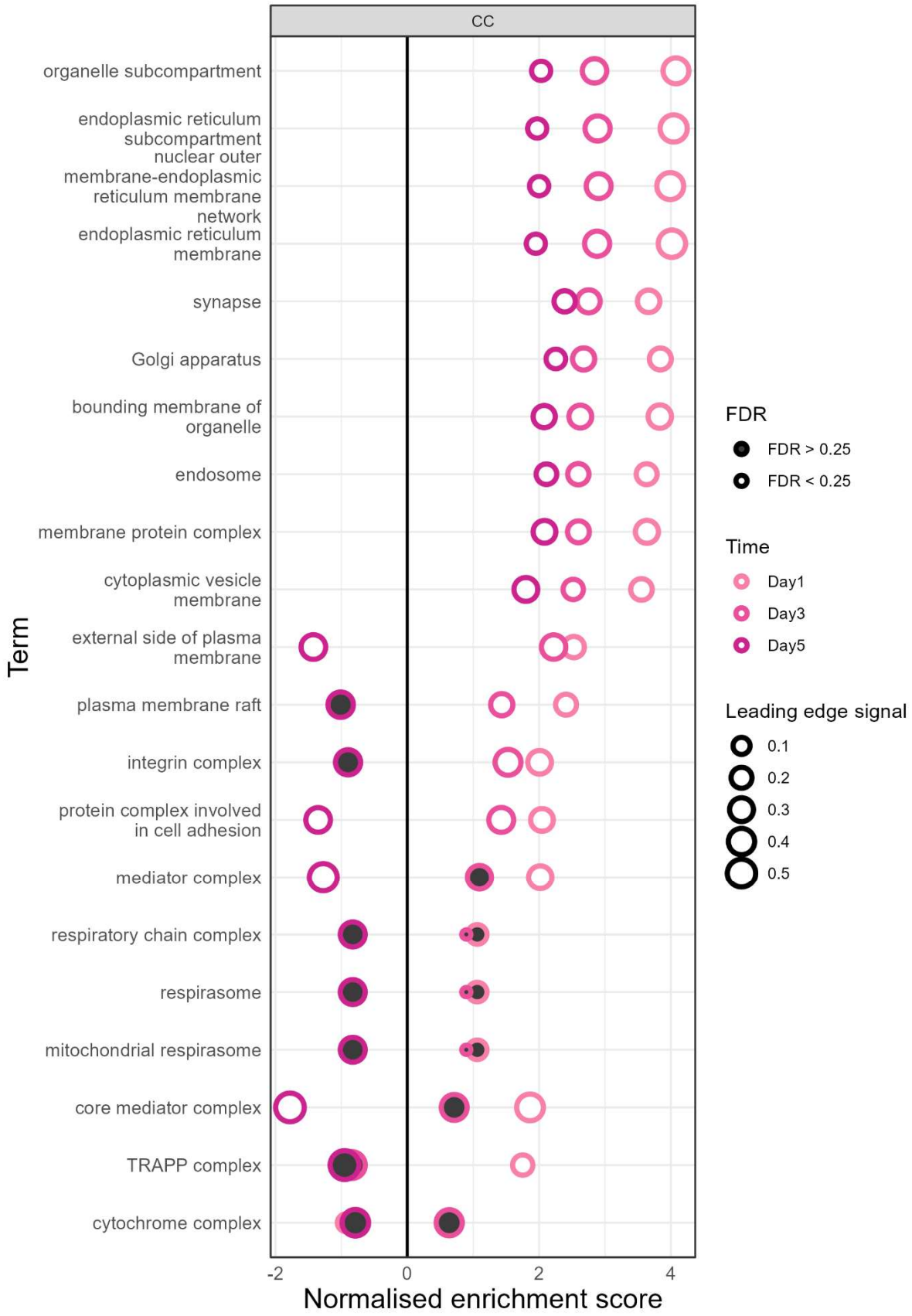


Supplementary Figure 15 ***Functional enrichment analysis of clustered protein profiles of DAPs (GOCC).*** Over-representation analysis was performed on each cluster, using the total proteome as the background proteins. Point size and colour represent -log10(p-value) and FDR, respectively. Gene ontology (GO) cellular component (CC) terms are shown.



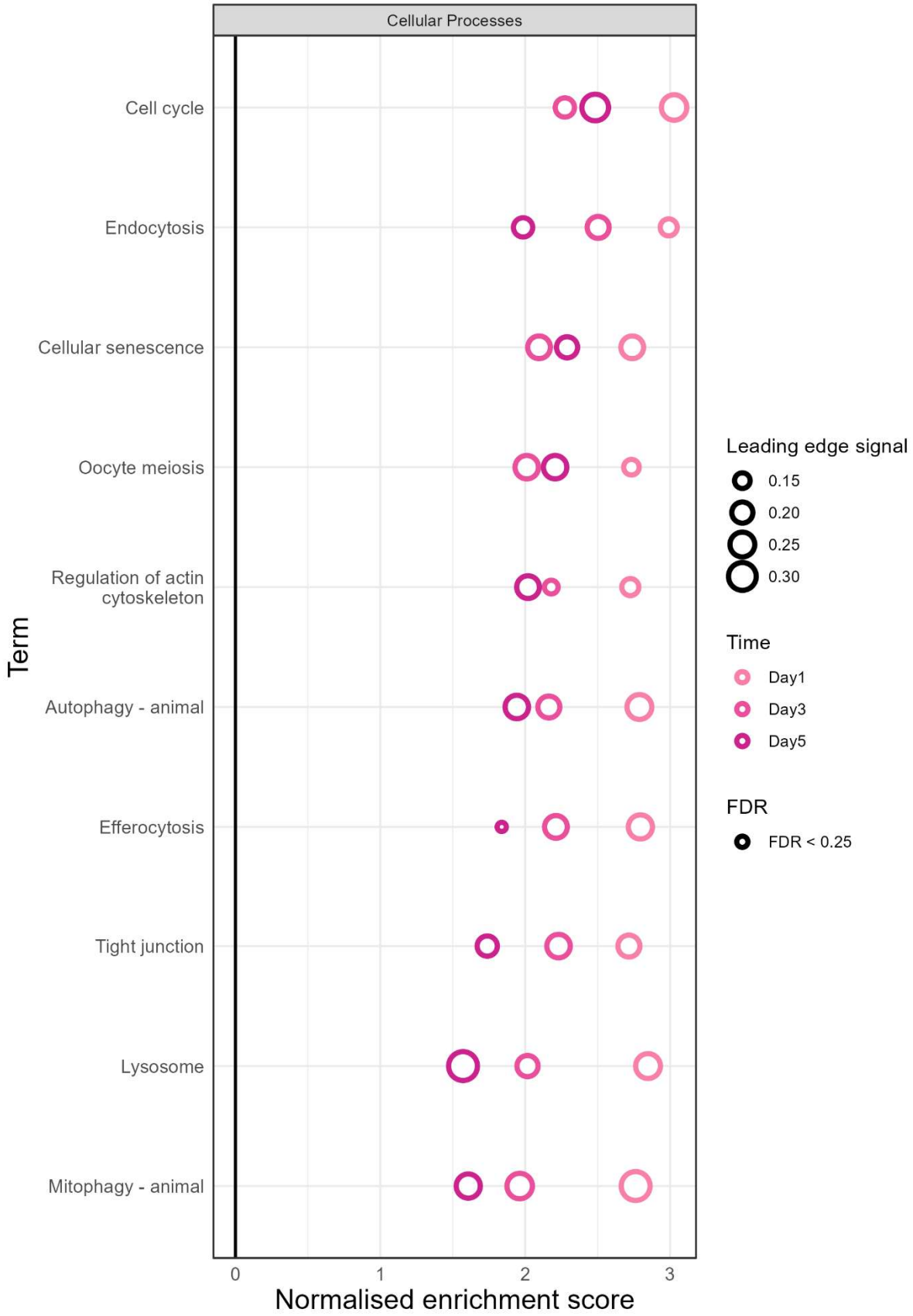
Supplementary Figure 16 ***Functional enrichment analysis of clustered protein profiles of DAPs (KEGG).*** Over-representation analysis was performed on each cluster, using the total proteome as the background proteins. Point size and colour represent -log10(p-value) and FDR, respectively. KEGG subcategory terms are shown.

Chapter 4

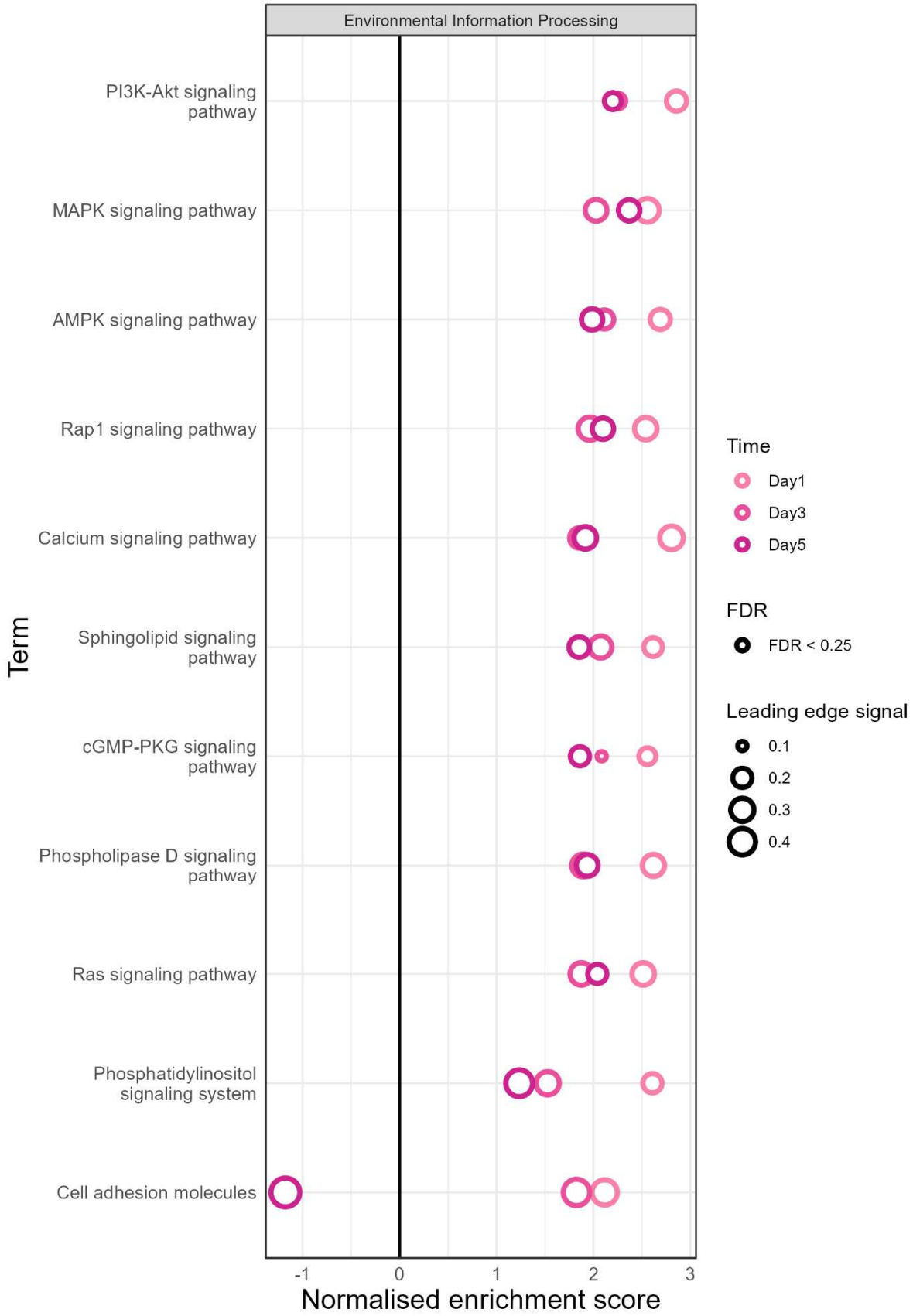


Supplementary Figure 17 ***Gene set enrichment analysis (GSEA) of proteome-normalised ubiquitome data from HoxB8 neutrophil differentiation (GOCC).*** Genes were pre-ranked based on corresponding ubiquitin-modified peptide log2(fold change) compared to day 0. Stroke colour represents the time point, size indicates leading edge signal, and permutation-based FDR is indicated by fill colour. Gene ontology (GO) cellular component (CC) terms are shown.

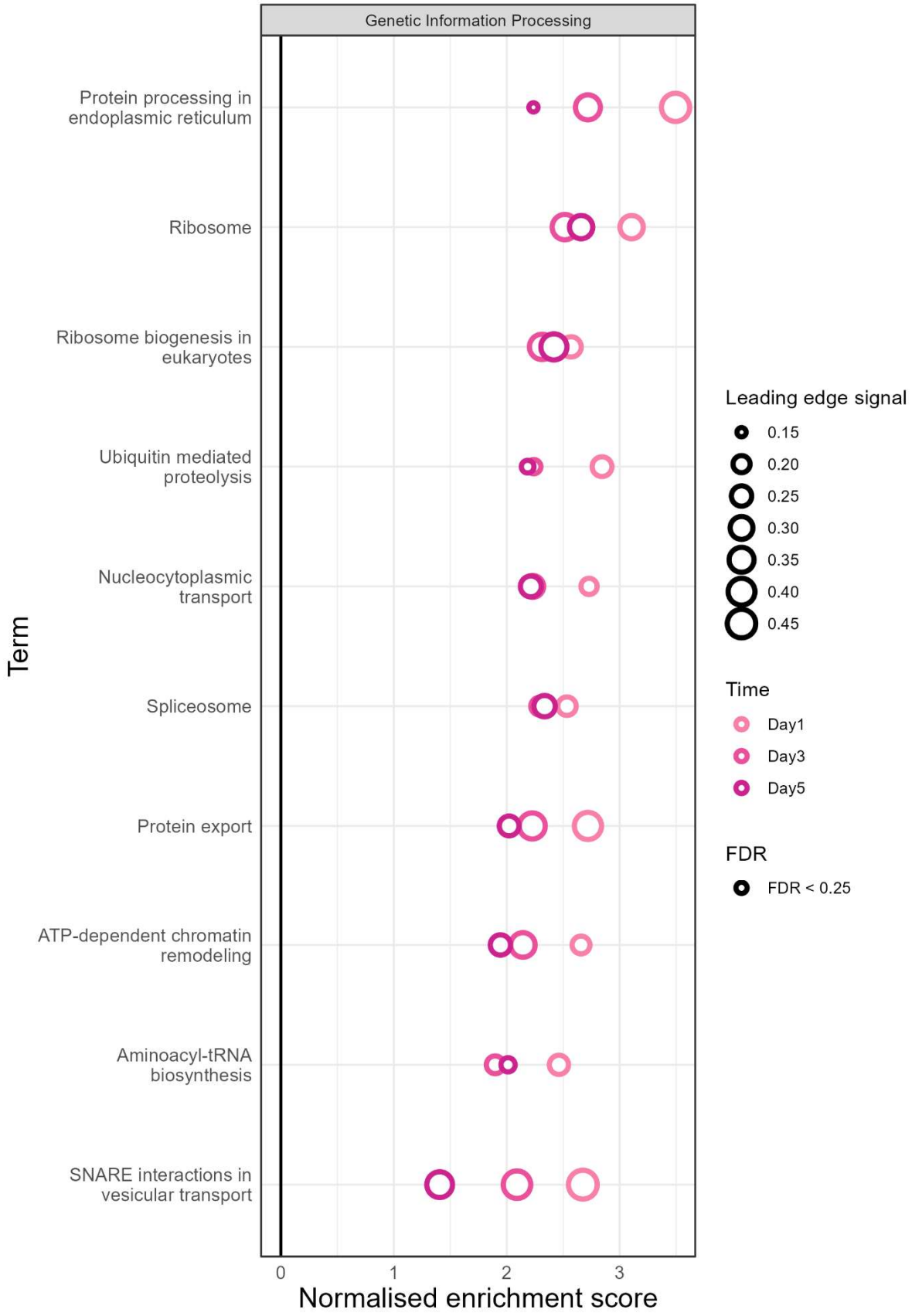
Supplementary Figure 18 ***Gene set enrichment analysis (GSEA) of proteome-normalised ubiquitome data from HoxB8 neutrophil differentiation (GOMF).*** Genes were pre-ranked based on corresponding ubiquitin-modified peptide log2(fold change) compared to day 0. Stroke colour represents the time point, size indicates leading edge signal, and permutation-based FDR is indicated by fill colour. Gene ontology (GO) molecular function (MF) terms are shown.



Supplementary Figure 19 ***Gene set enrichment analysis (GSEA) of proteome-normalised ubiquitome data from HoxB8 neutrophil differentiation (KEGG, Cellular Processes).*** Genes were pre-ranked based on corresponding ubiquitin-modified peptide log2(fold change) compared to day 0. Stroke colour represents the time point, size indicates leading edge signal, and permutation-based FDR is indicated by fill colour. The KEGG pathway subcategory 'Cellular Processes' terms are shown.

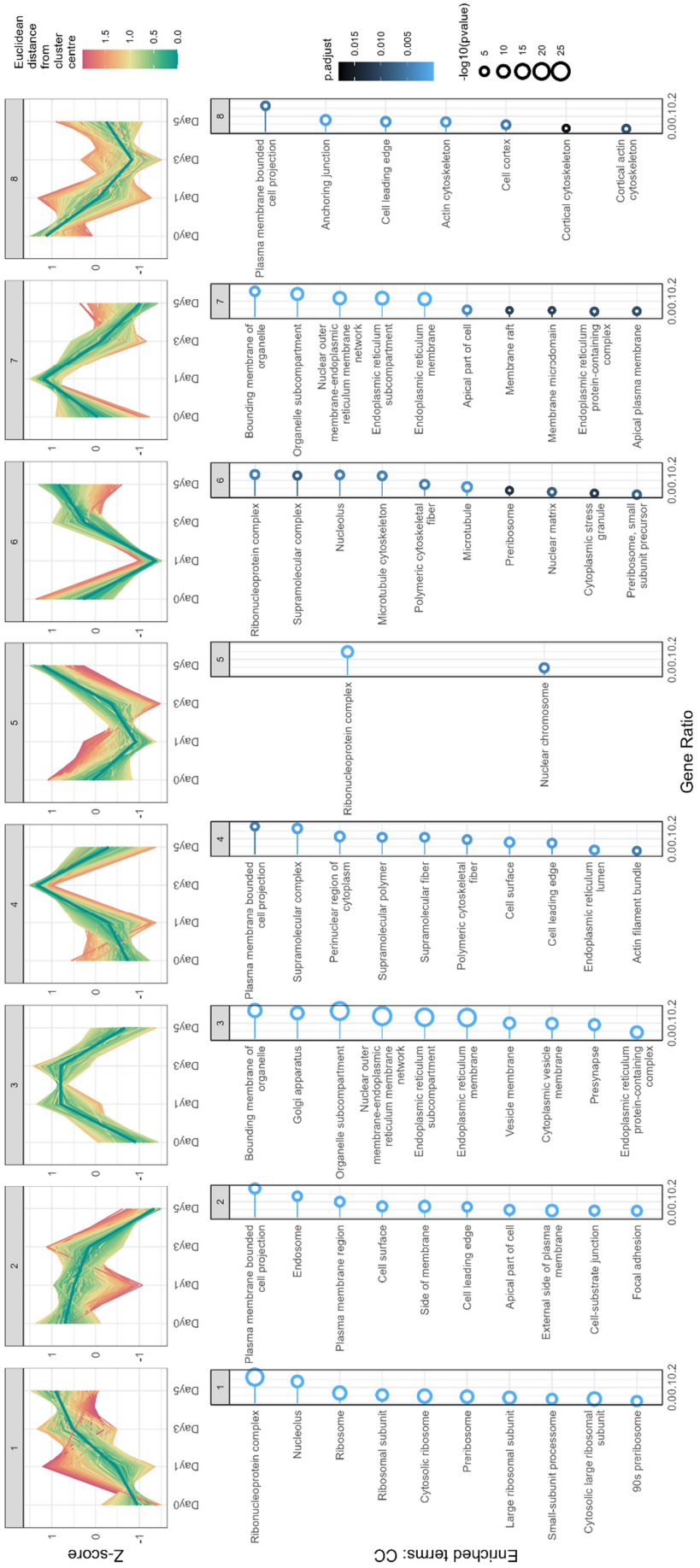


Supplementary Figure 20 ***Gene set enrichment analysis (GSEA) of proteome-normalised ubiquitome data from HoxB8 neutrophil differentiation (KEGG, Environmental Information Processing).*** Genes were pre-ranked based on corresponding ubiquitin-modified peptide log2(fold change) compared to day 0. Stroke colour represents the time point, size indicates leading edge signal, and permutation-based FDR is indicated by fill colour. The KEGG pathway subcategory 'Environmental Information Processing' terms are shown.

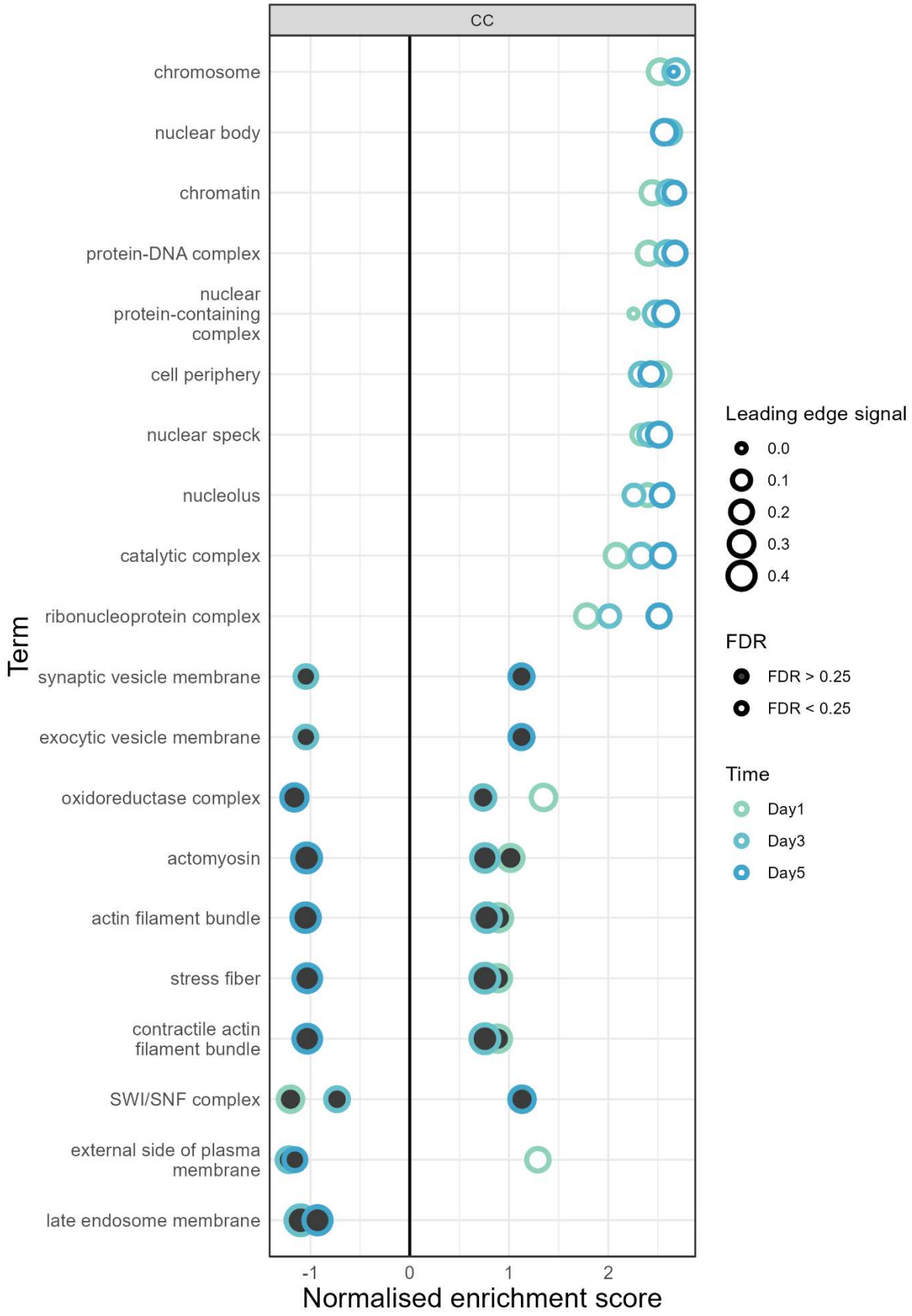


Supplementary Figure 21 ***Gene set enrichment analysis (GSEA) of proteome-normalised ubiquitome data from HoxB8 neutrophil differentiation (KEGG, Genetic Information Processing).*** Genes were pre-ranked based on corresponding ubiquitin-modified peptide log2(fold change) compared to day 0. Stroke colour represents the time point, size indicates leading edge signal, and permutation-based FDR is indicated by fill colour. The KEGG pathway subcategory 'Genetic Information Processing' terms are shown.

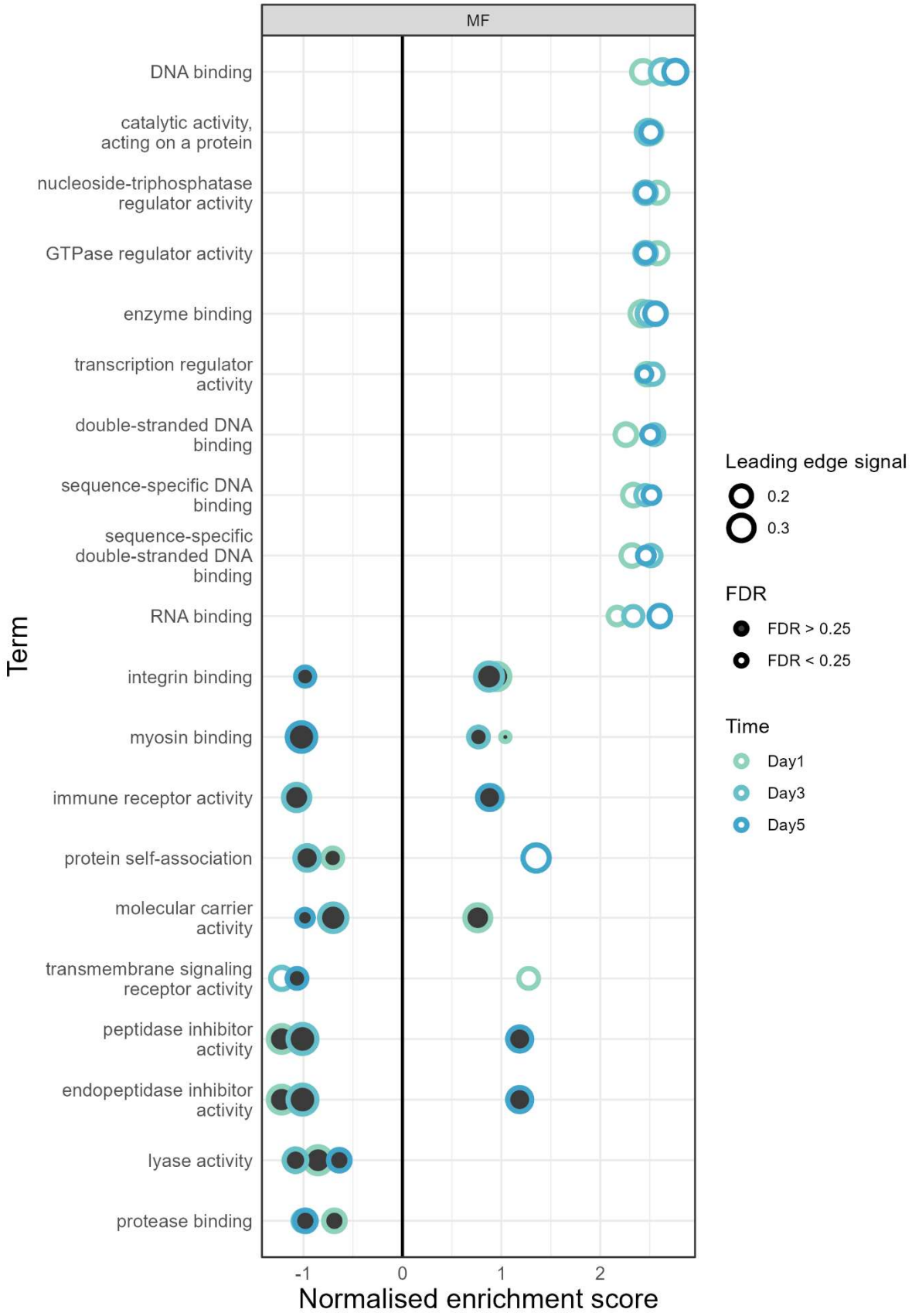
Supplementary Figure 22 ***Gene set enrichment analysis (GSEA) of proteome-normalised ubiquitome data from HoxB8 neutrophil differentiation (KEGG, Metabolism).*** Genes were pre-ranked based on corresponding ubiquitin-modified peptide log2(fold change) compared to day 0. Stroke colour represents the time point, size indicates leading edge signal, and permutation-based FDR is indicated by fill colour. The KEGG pathway subcategory 'Metabolism' terms are shown.



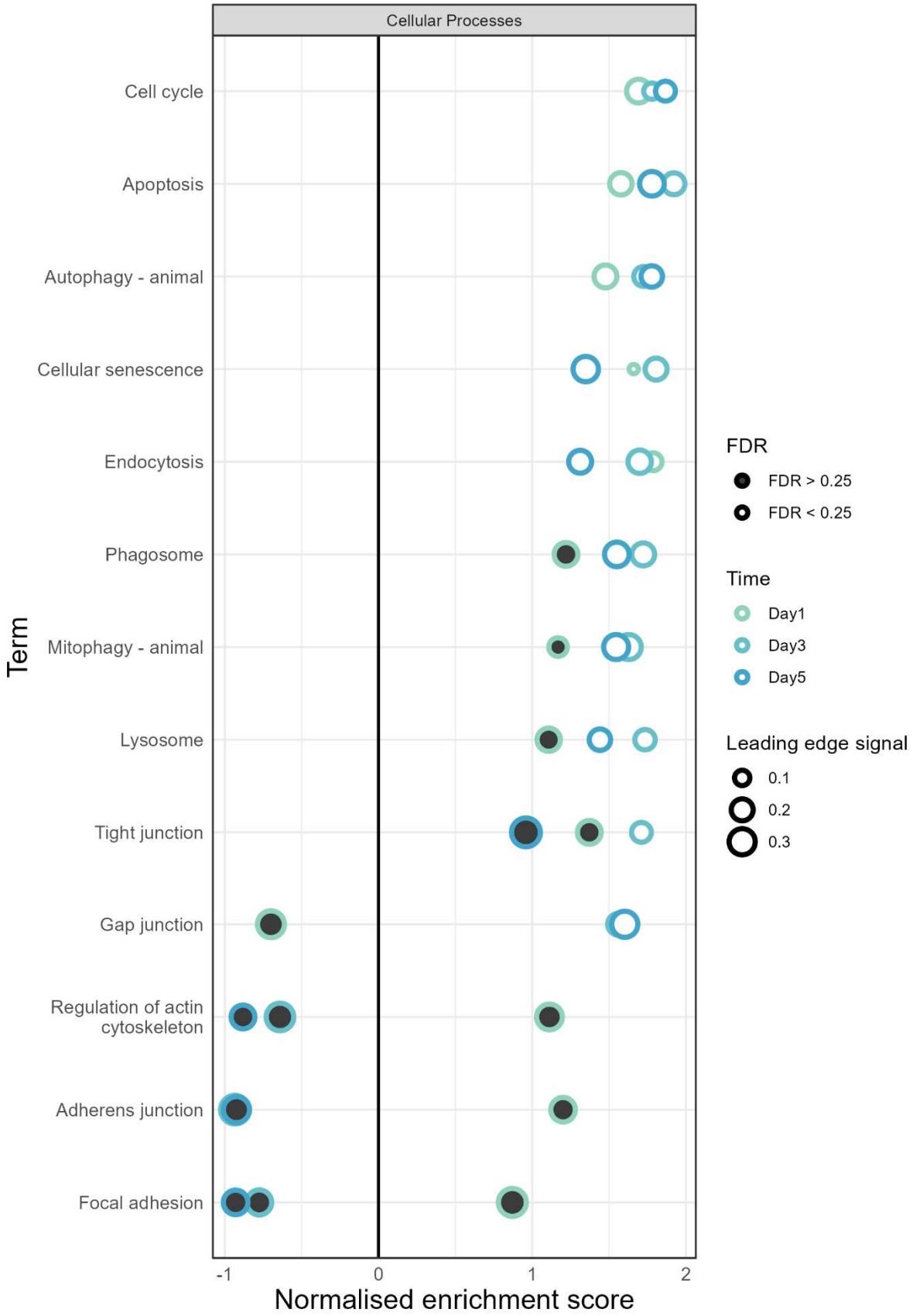
Supplementary Figure 23 ***Functional enrichment analysis of clustered proteome-normalised ubiquitin-modified peptide profiles that are differentially abundant during HoxB8 neutrophil differentiation (GOCC).*** Over-representation analysis was performed on each cluster, using the proteins identified in the ubiquitome and total proteome as the background proteins. Point size and colour represent -log10(p-value) and FDR, respectively. Gene ontology (GO) cellular component (CC) terms are shown.



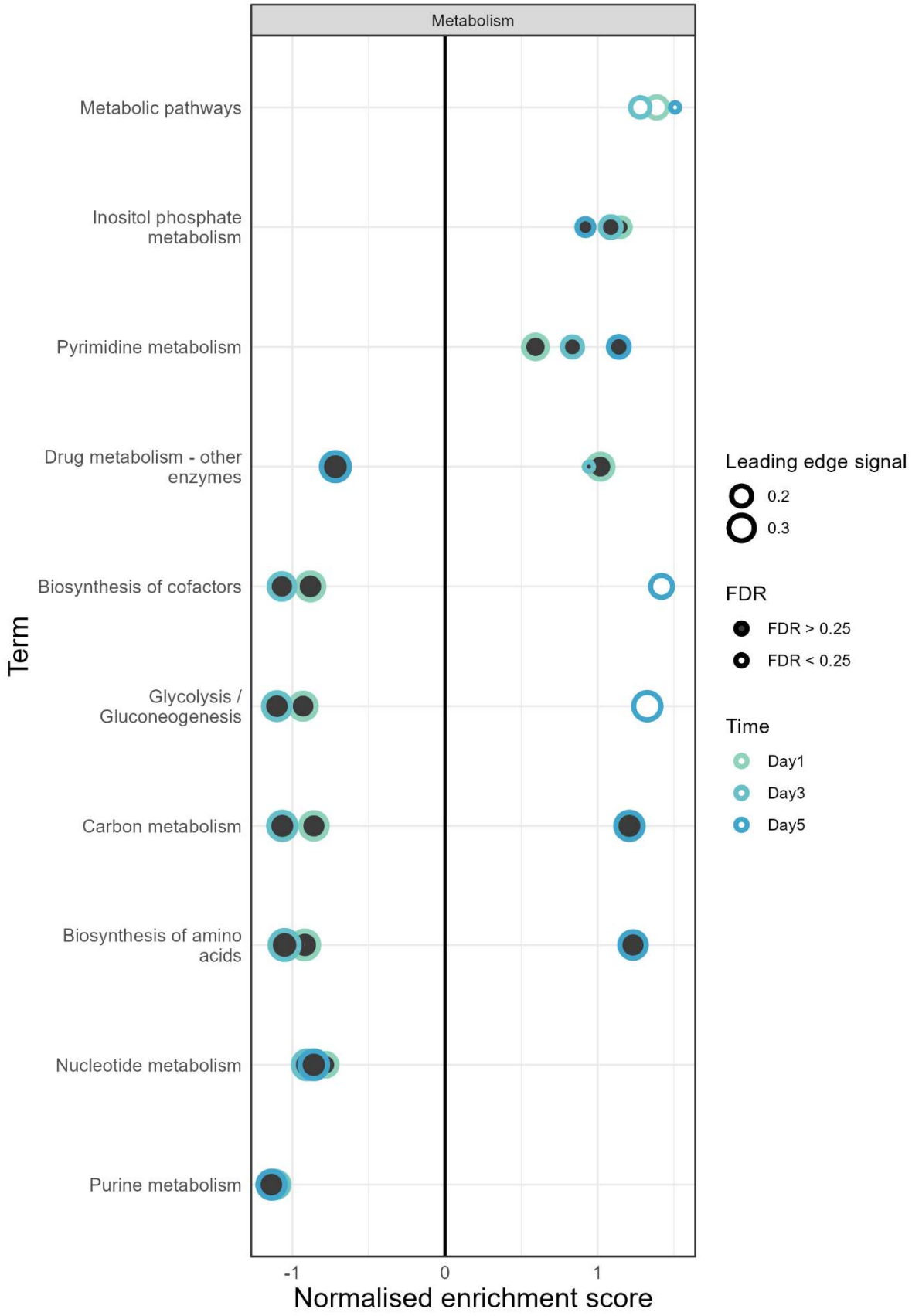
Supplementary Figure 24 ***Gene set enrichment analysis (GSEA) of proteome-normalised phosphoproteome data from HoxB8 neutrophil differentiation (GOCC).*** Genes were pre-ranked based on corresponding phosphopeptide log2(fold change) compared to day 0. Stroke colour represents the time point, size indicates leading edge signal, and permutation-based FDR is indicated by fill colour. Gene ontology (GO) cellular component (CC) terms are shown.



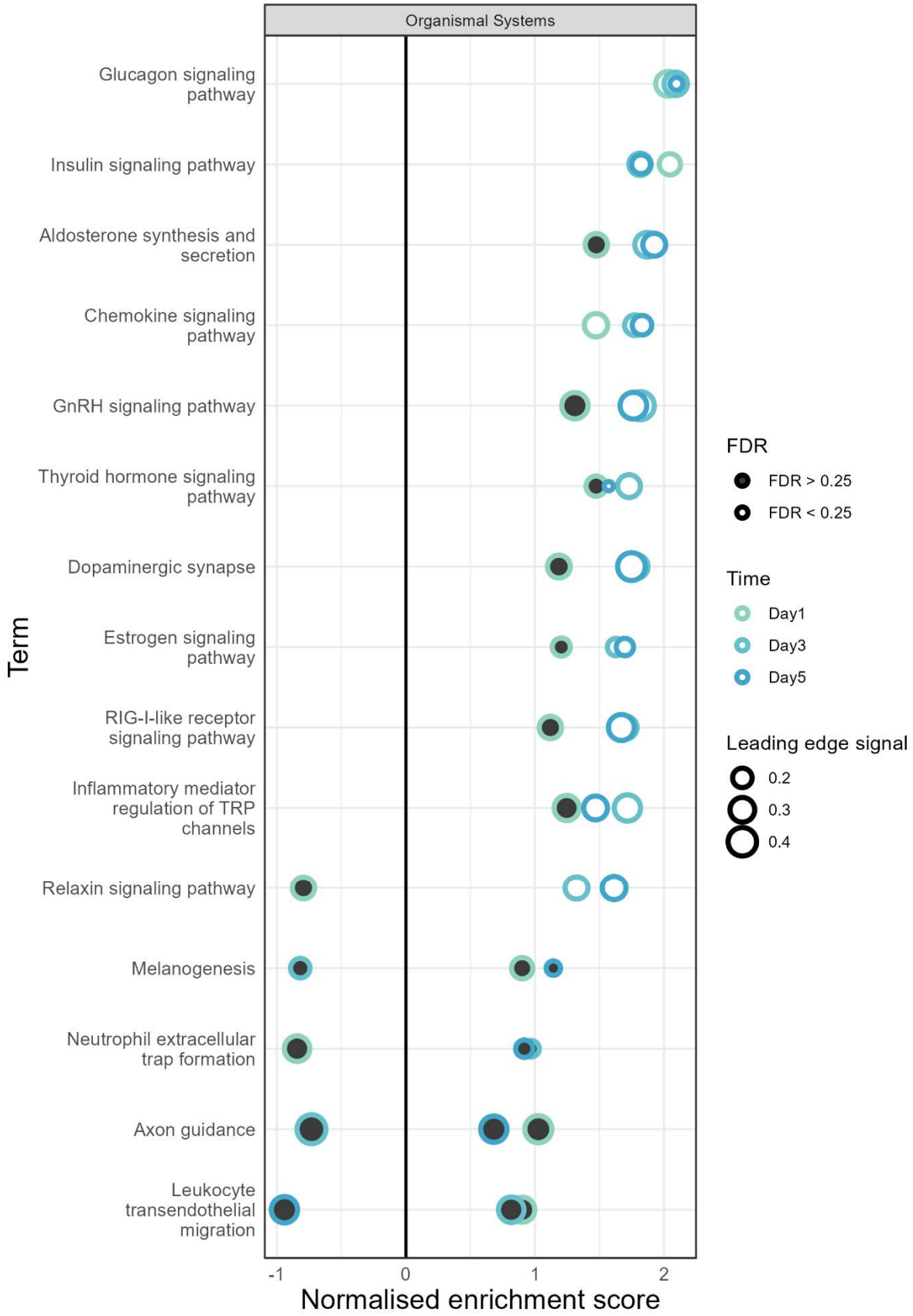
Supplementary Figure 25 ***Gene set enrichment analysis (GSEA) of proteome-normalised phosphoproteome data from HoxB8 neutrophil differentiation (GOMF).*** Genes were pre-ranked based on corresponding phosphopeptide log2(fold change) compared to day 0. Stroke colour represents the time point, size indicates leading edge signal, and permutation-based FDR is indicated by fill colour. Gene ontology (GO) molecular function (MF) terms are shown.



Supplementary Figure 26 ***Gene set enrichment analysis (GSEA) of proteome-normalised phosphoproteome data from HoxB8 neutrophil differentiation (KEGG, Cellular Processes).*** Genes were pre-ranked based on corresponding phosphopeptide log2(fold change) compared to day 0. Stroke colour represents the time point, size indicates leading edge signal, and permutation-based FDR is indicated by fill colour. KEGG subcategory 'Cellular Processes' terms are shown.

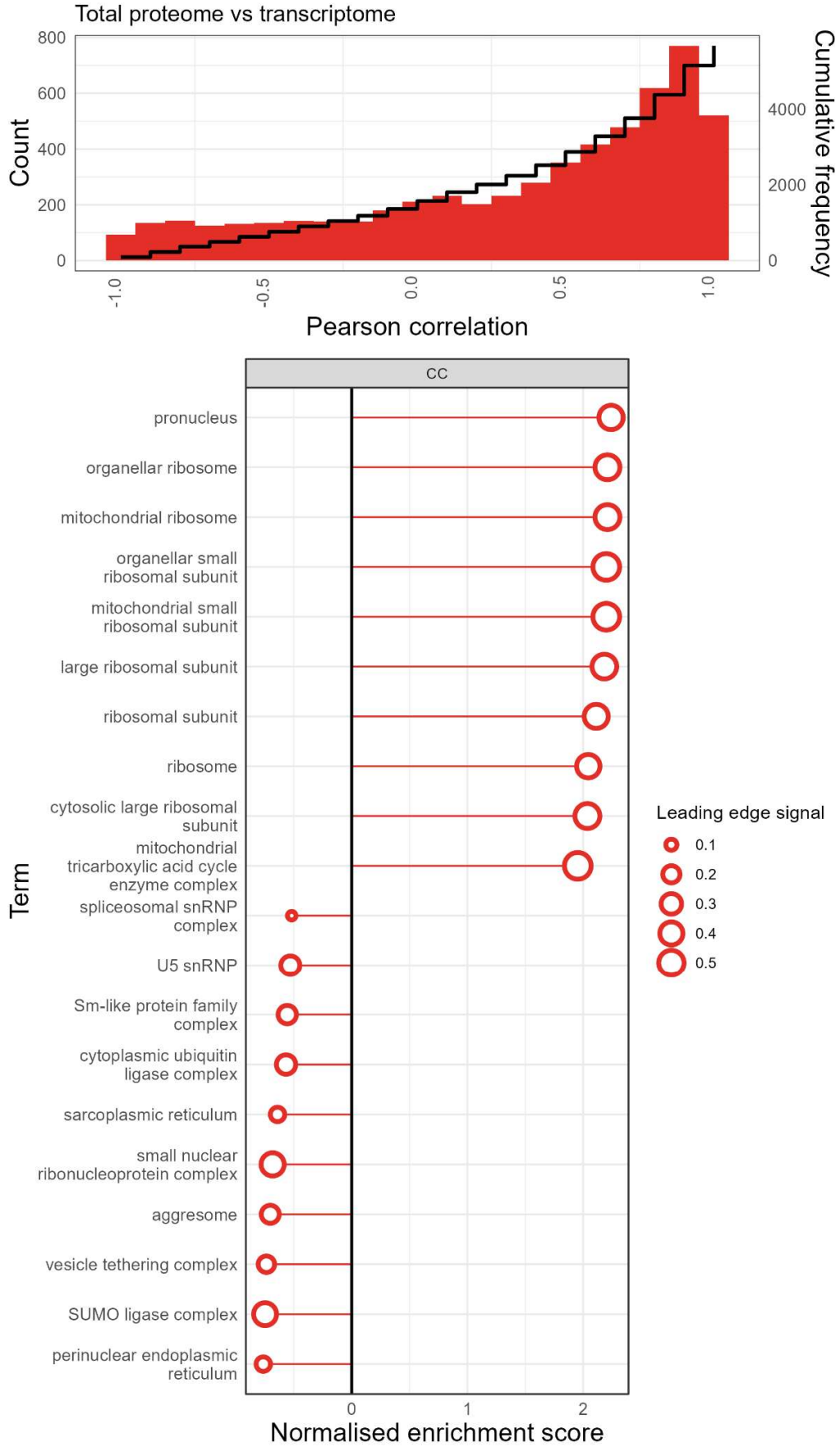


Supplementary Figure 27 ***Gene set enrichment analysis (GSEA) of proteome-normalised phosphoproteome data from HoxB8 neutrophil differentiation (KEGG, Metabolism).*** Genes were pre-ranked based on corresponding phosphopeptide log2(fold change) compared to day 0. Stroke colour represents the time point, size indicates leading edge signal, and permutation-based FDR is indicated by fill colour. KEGG subcategory 'Metabolism' terms are shown.

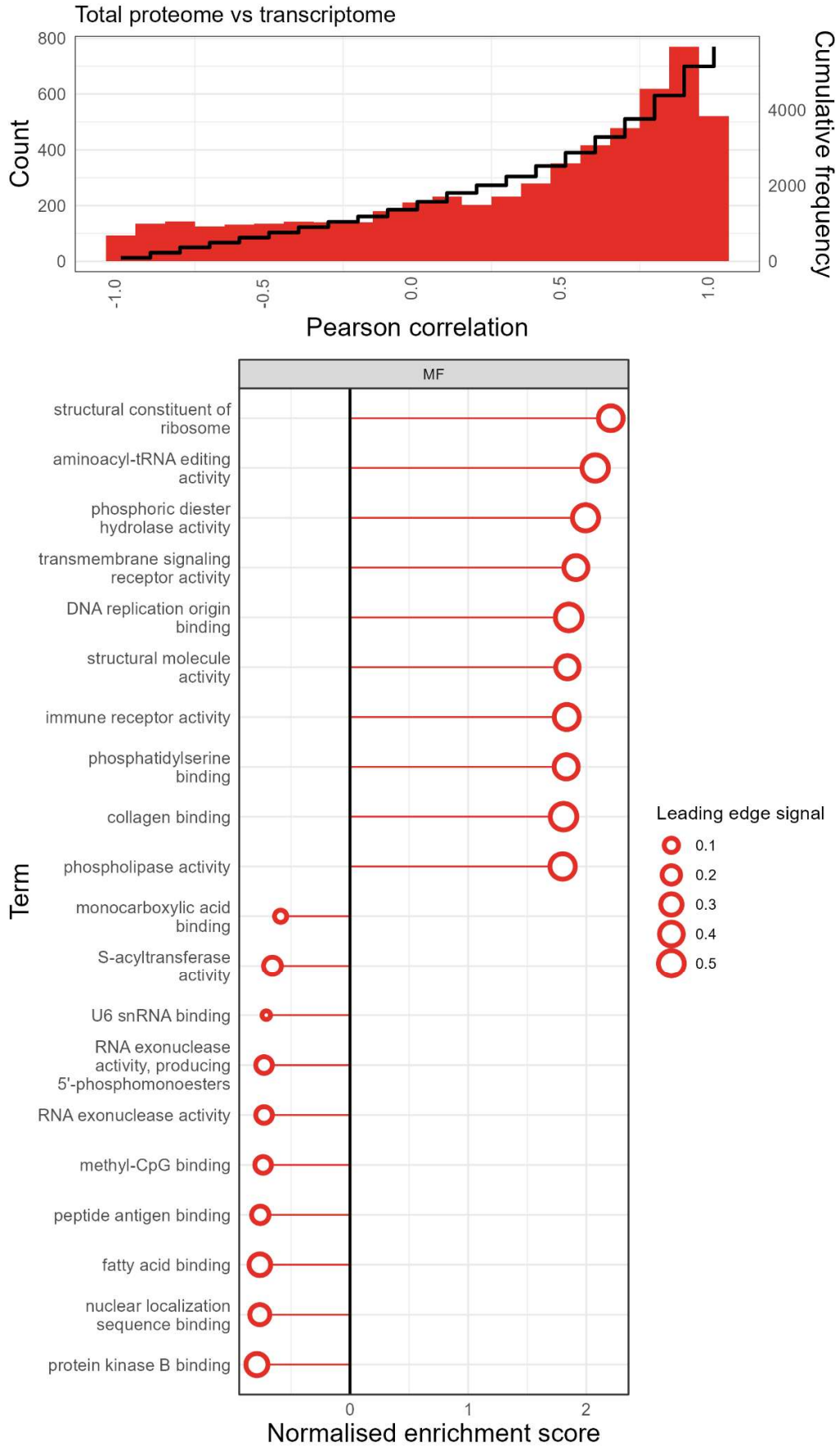


Supplementary Figure 28 ***Gene set enrichment analysis (GSEA) of proteome-normalised phosphoproteome data from HoxB8 neutrophil differentiation (KEGG, Organismal Systems).*** Genes were pre-ranked based on corresponding phosphopeptide log2(fold change) compared to day 0. Stroke colour represents the time point, size indicates leading edge signal, and permutation-based FDR is indicated by fill colour. KEGG subcategory 'Organismal Systems' terms are shown.

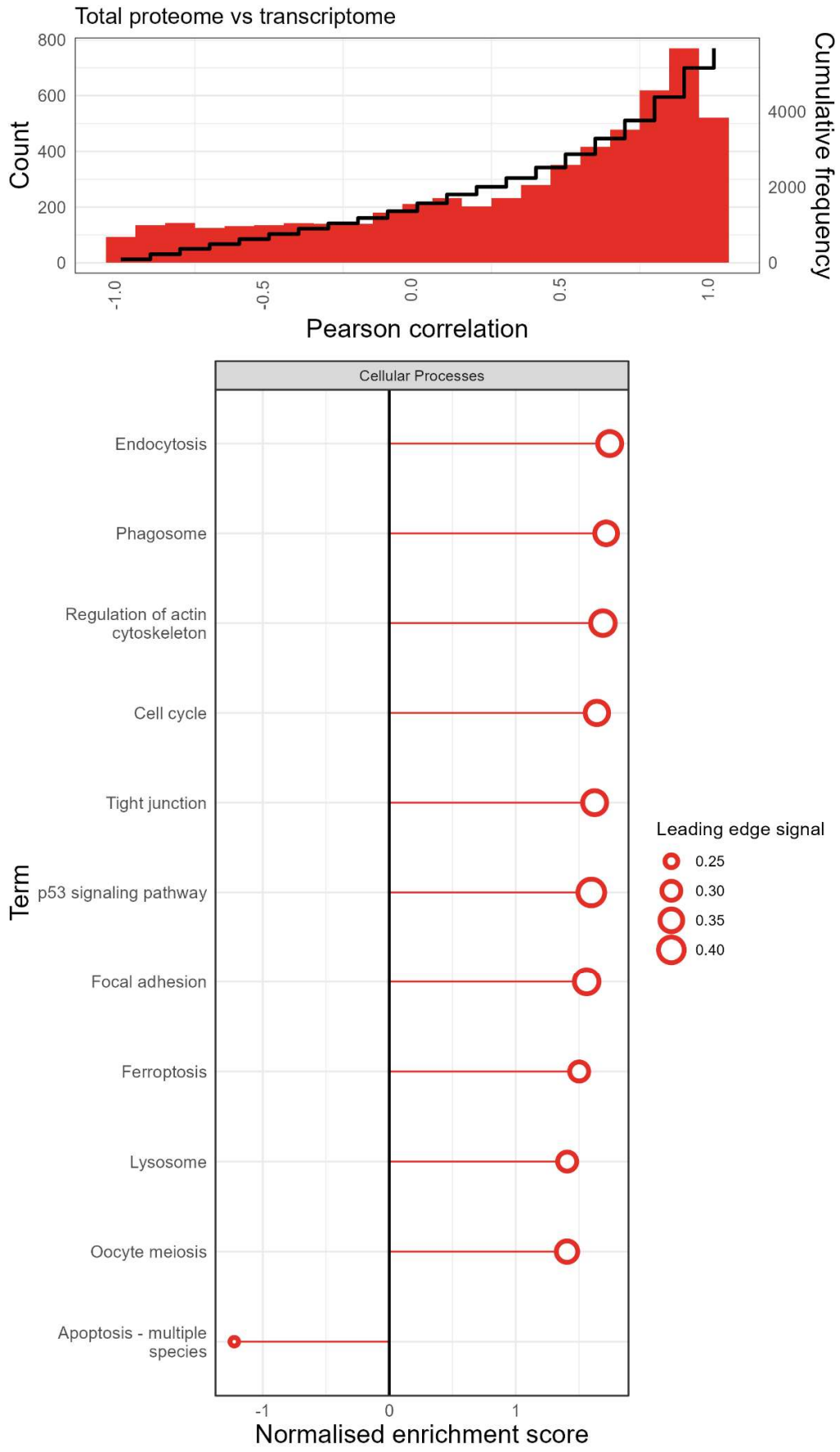
Chapter 5



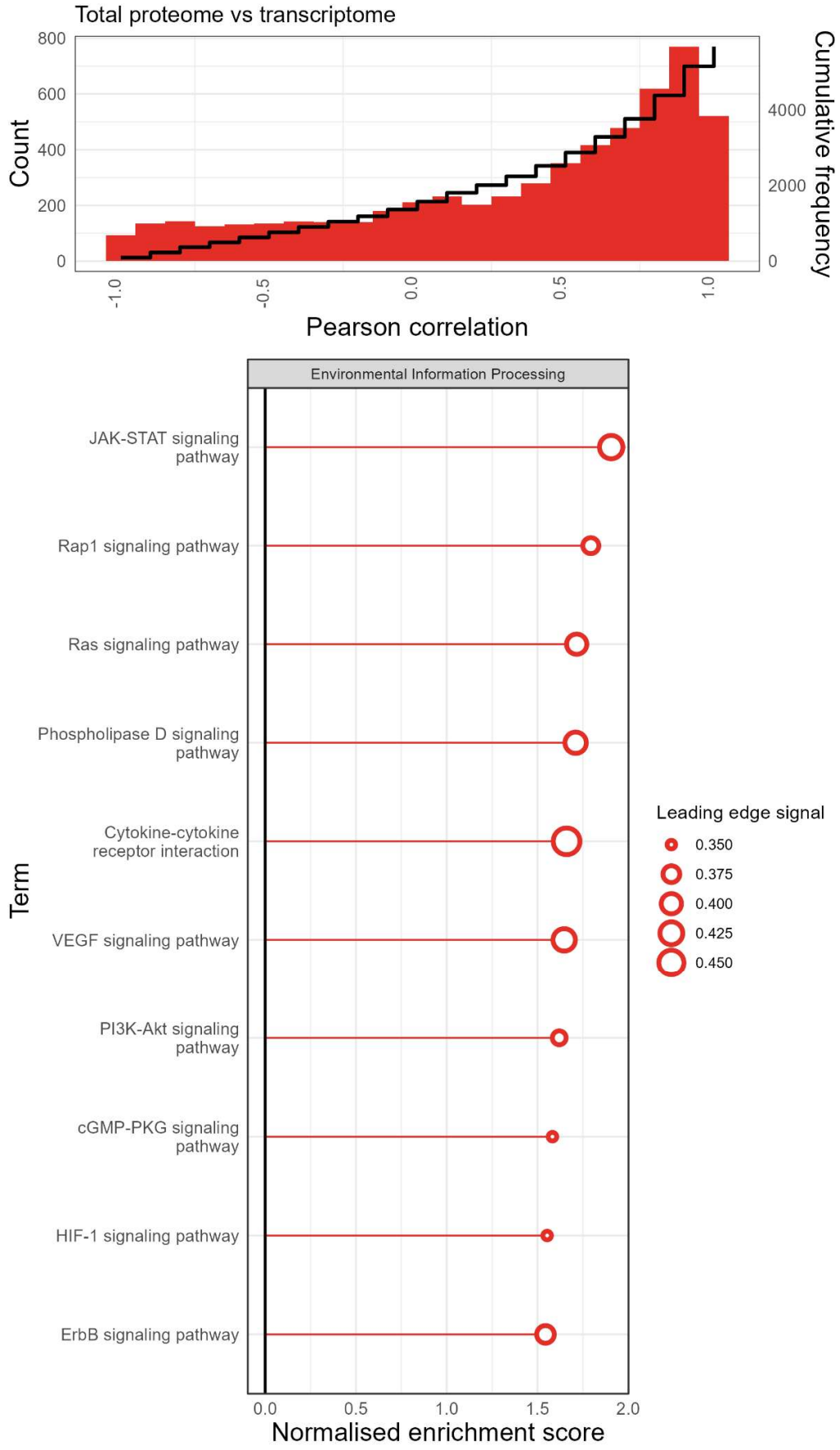
Supplementary Figure 29 ***Gene set enrichment analysis (GSEA) of transcriptome-proteome data during HoxB8 neutrophil differentiation (GOBP).*** Genes were pre-ranked based on correlation between mRNA levels and corresponding protein levels, across all time points of HoxB8 differentiation. Gene ontology (GO) cellular component (CC) terms are shown.



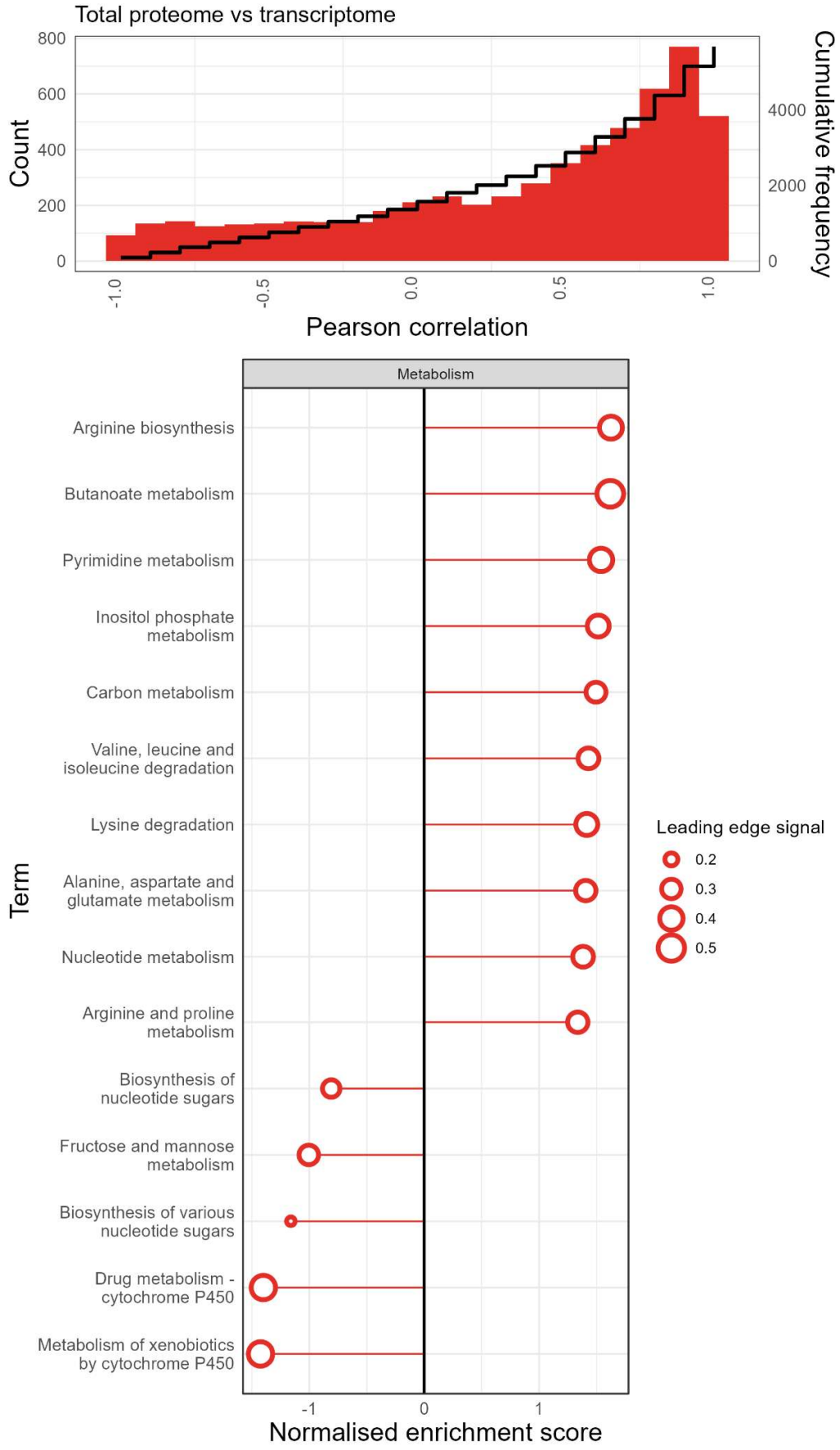
Supplementary Figure 30 ***Gene set enrichment analysis (GSEA) of transcriptome-proteome data during HoxB8 neutrophil differentiation (GOBP).*** Genes were pre-ranked based on correlation between mRNA levels and corresponding protein levels, across all time points of HoxB8 differentiation. Gene ontology (GO) molecular function (MF) terms are shown.



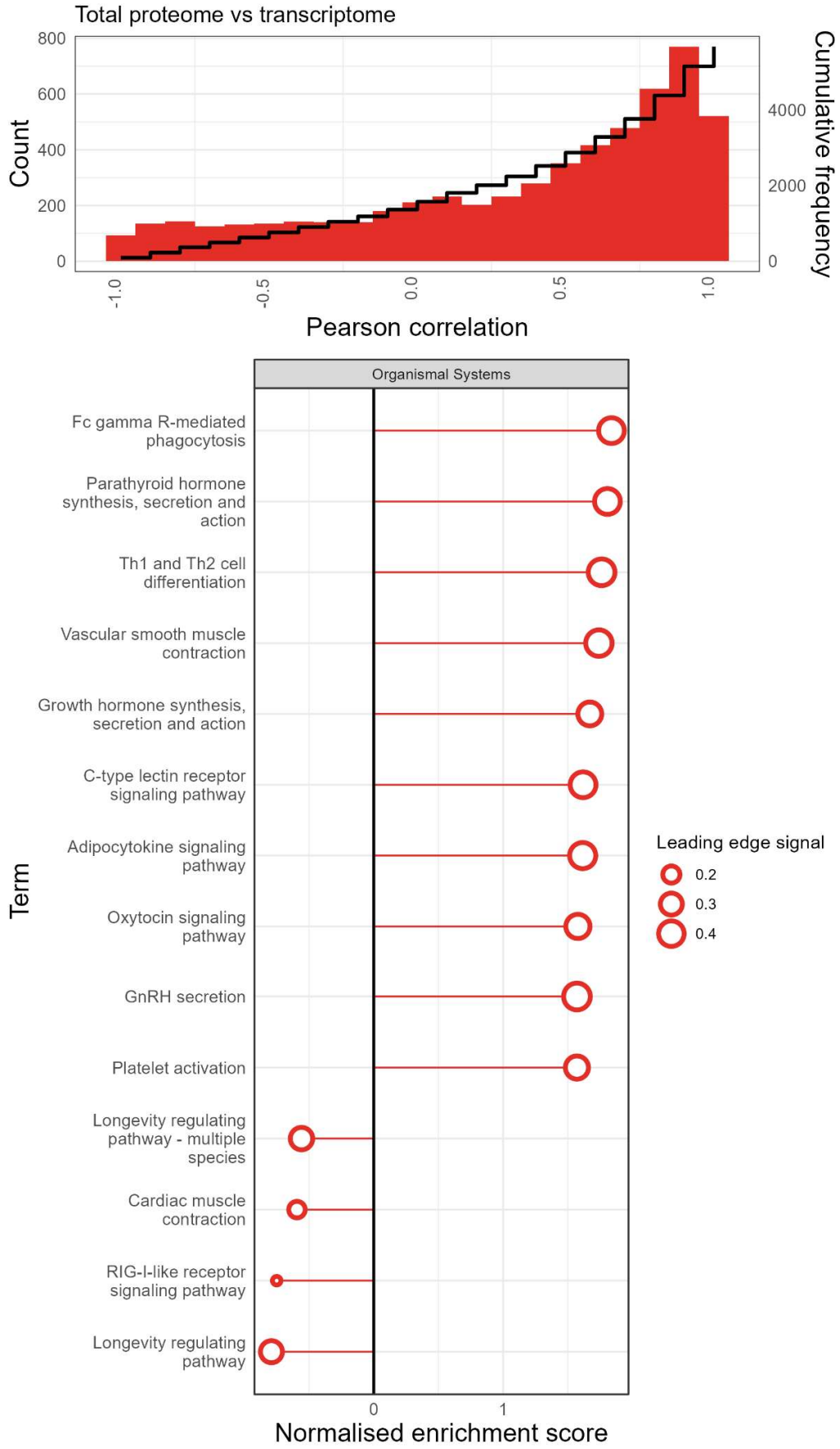
Supplementary Figure 31 ***Gene set enrichment analysis (GSEA) of transcriptome-proteome data during HoxB8 neutrophil differentiation (KEGG, Genetic Information Processing).*** Genes were pre-ranked based on correlation between mRNA levels and corresponding protein levels, across all time points of HoxB8 differentiation. The KEGG pathway subcategory 'Cellular Processes' terms are shown.



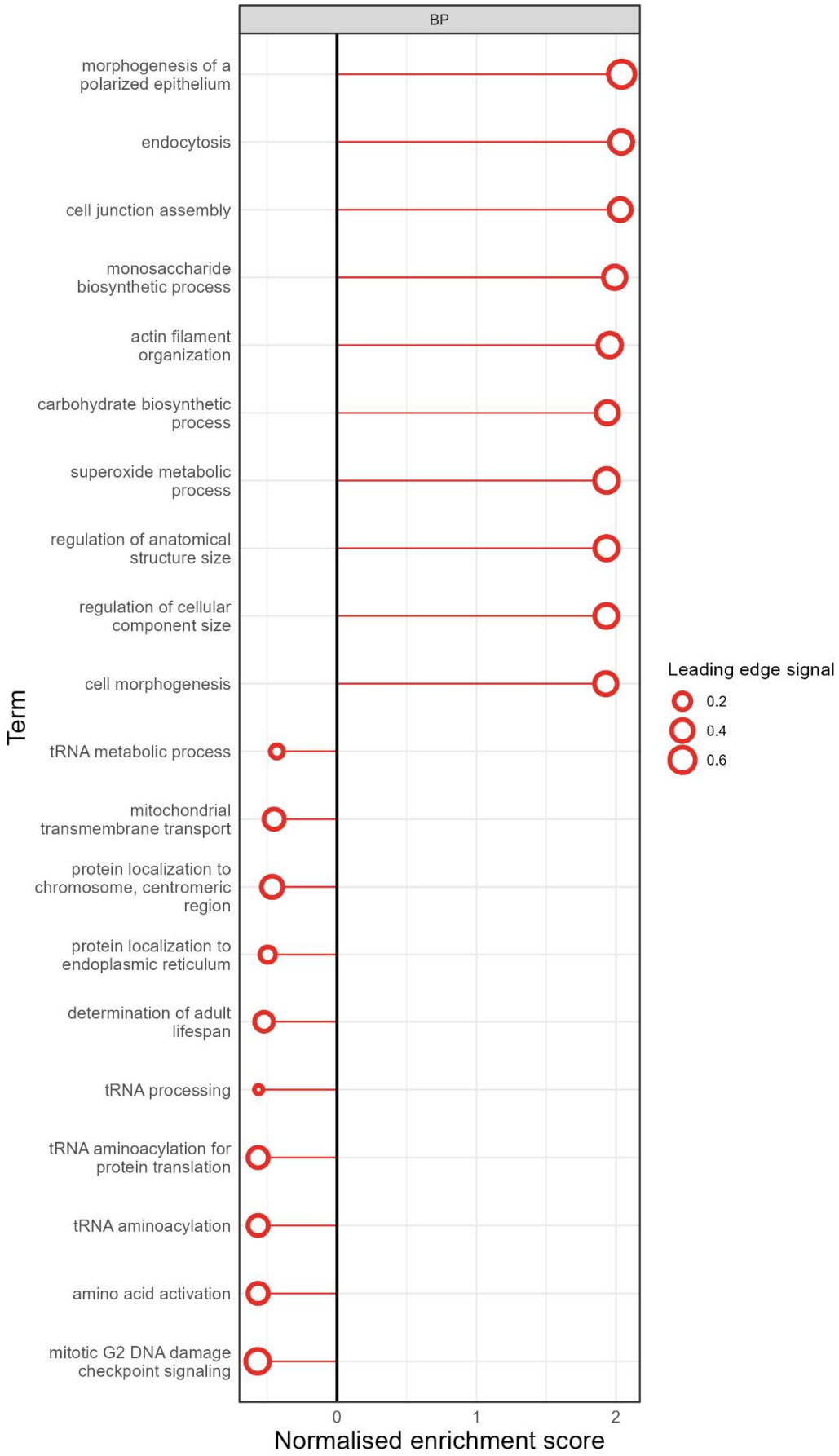
Supplementary Figure 32 ***Gene set enrichment analysis (GSEA) of transcriptome-proteome data during HoxB8 neutrophil differentiation (KEGG, Genetic Information Processing).*** Genes were pre-ranked based on correlation between mRNA levels and corresponding protein levels, across all time points of HoxB8 differentiation. The KEGG pathway subcategory 'Environmental Information Processing' terms are shown.



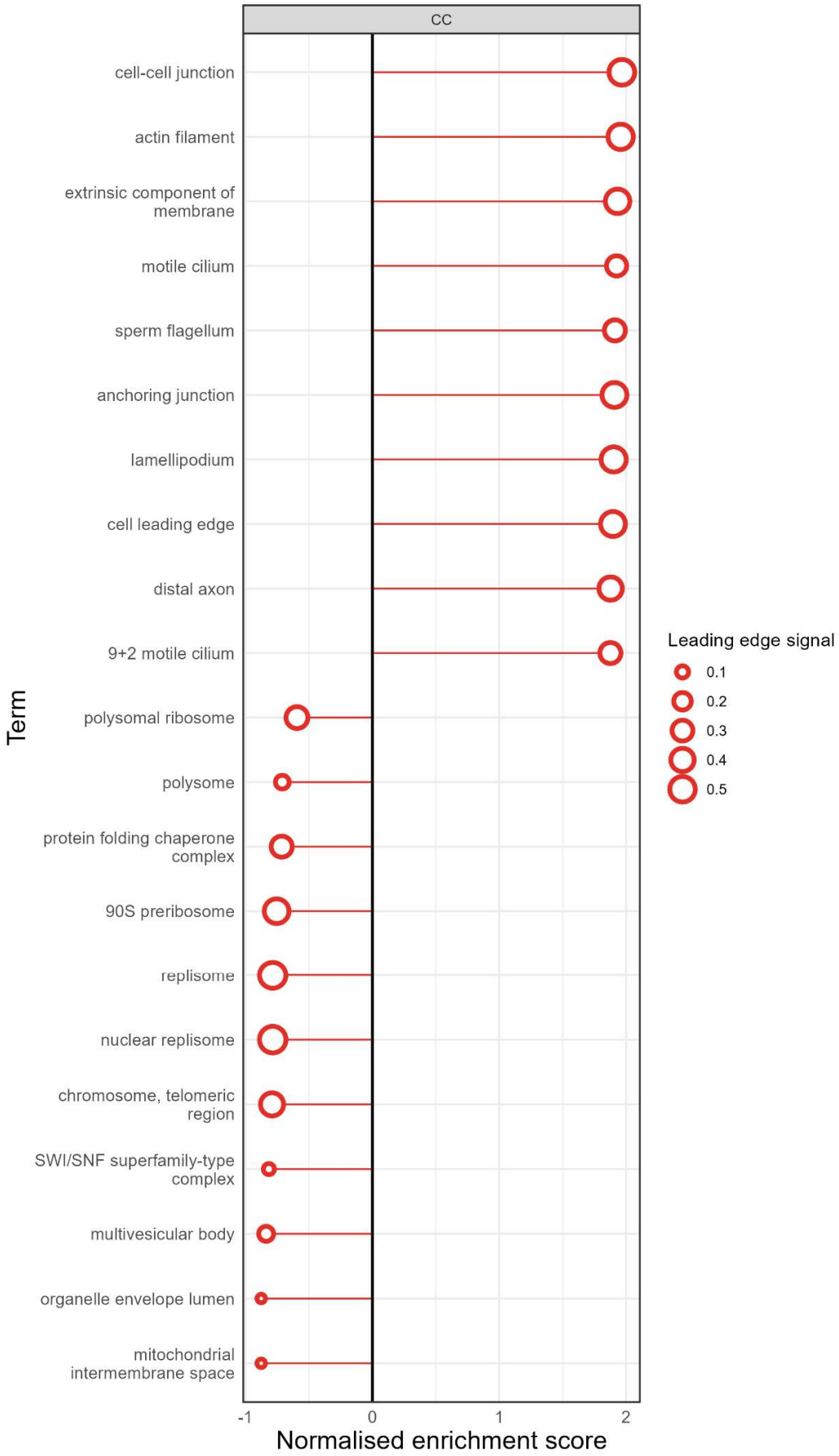
Supplementary Figure 33 ***Gene set enrichment analysis (GSEA) of transcriptome-proteome data during HoxB8 neutrophil differentiation (KEGG, Genetic Information Processing).*** Genes were pre-ranked based on correlation between mRNA levels and corresponding protein levels, across all time points of HoxB8 differentiation. The KEGG pathway subcategory 'Metabolism' terms are shown.



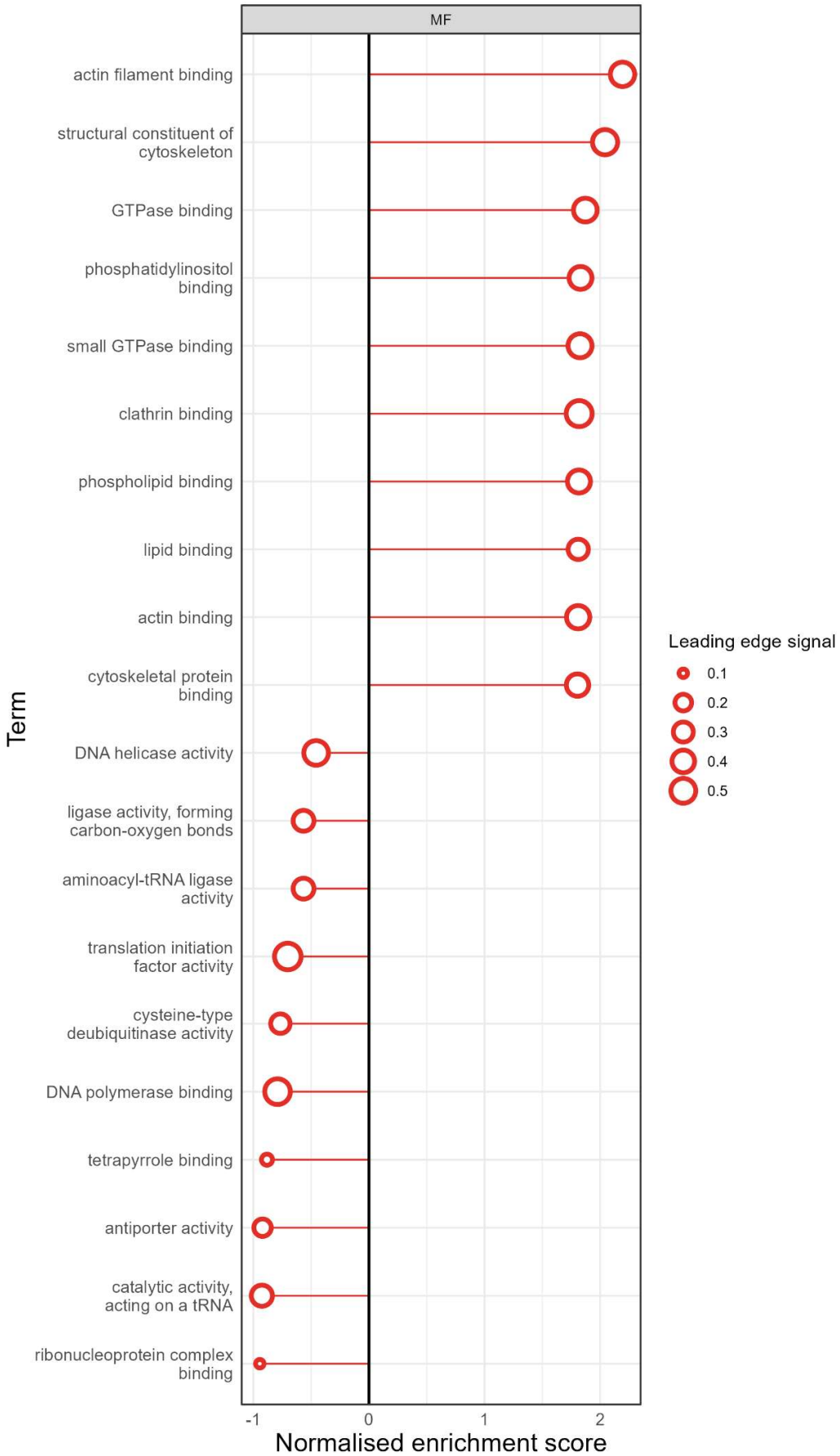
Supplementary Figure 34 ***Gene set enrichment analysis (GSEA) of transcriptome-proteome data during HoxB8 neutrophil differentiation (KEGG, Genetic Information Processing).*** Genes were pre-ranked based on correlation between mRNA levels and corresponding protein levels, across all time points of HoxB8 differentiation. The KEGG pathway subcategory 'Organismal Systems' terms are shown.



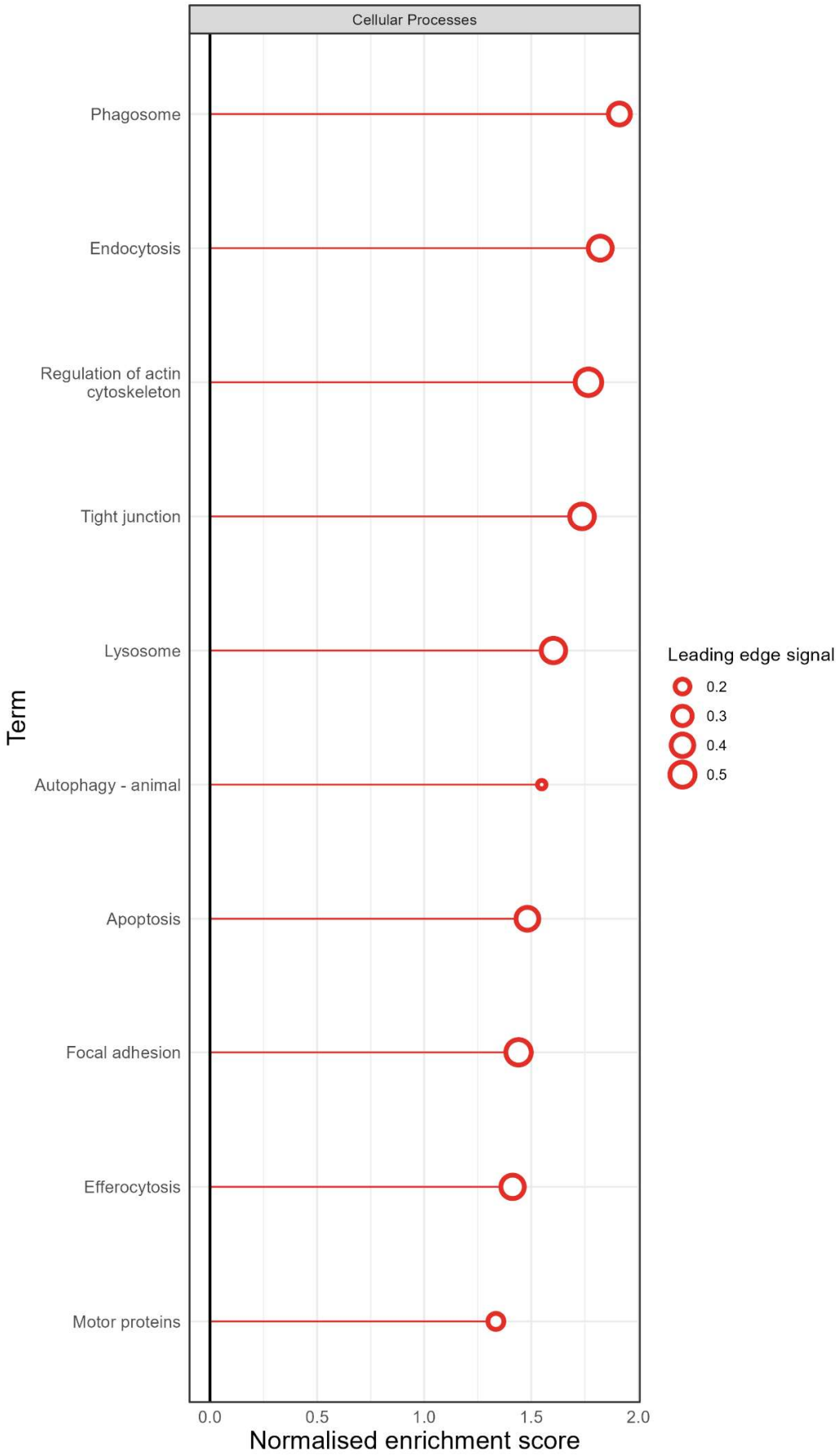
Supplementary Figure 35 ***Gene set enrichment analysis (GSEA) of transcriptome contributions to proteome levels during HoxB8 neutrophil differentiation, using a linear regression approach.*** Genes were pre-ranked based on statistically significant coefficients corresponding to transcriptome variables used to predict protein abundance. Gene ontology (GO) biological process (BP) terms are shown.



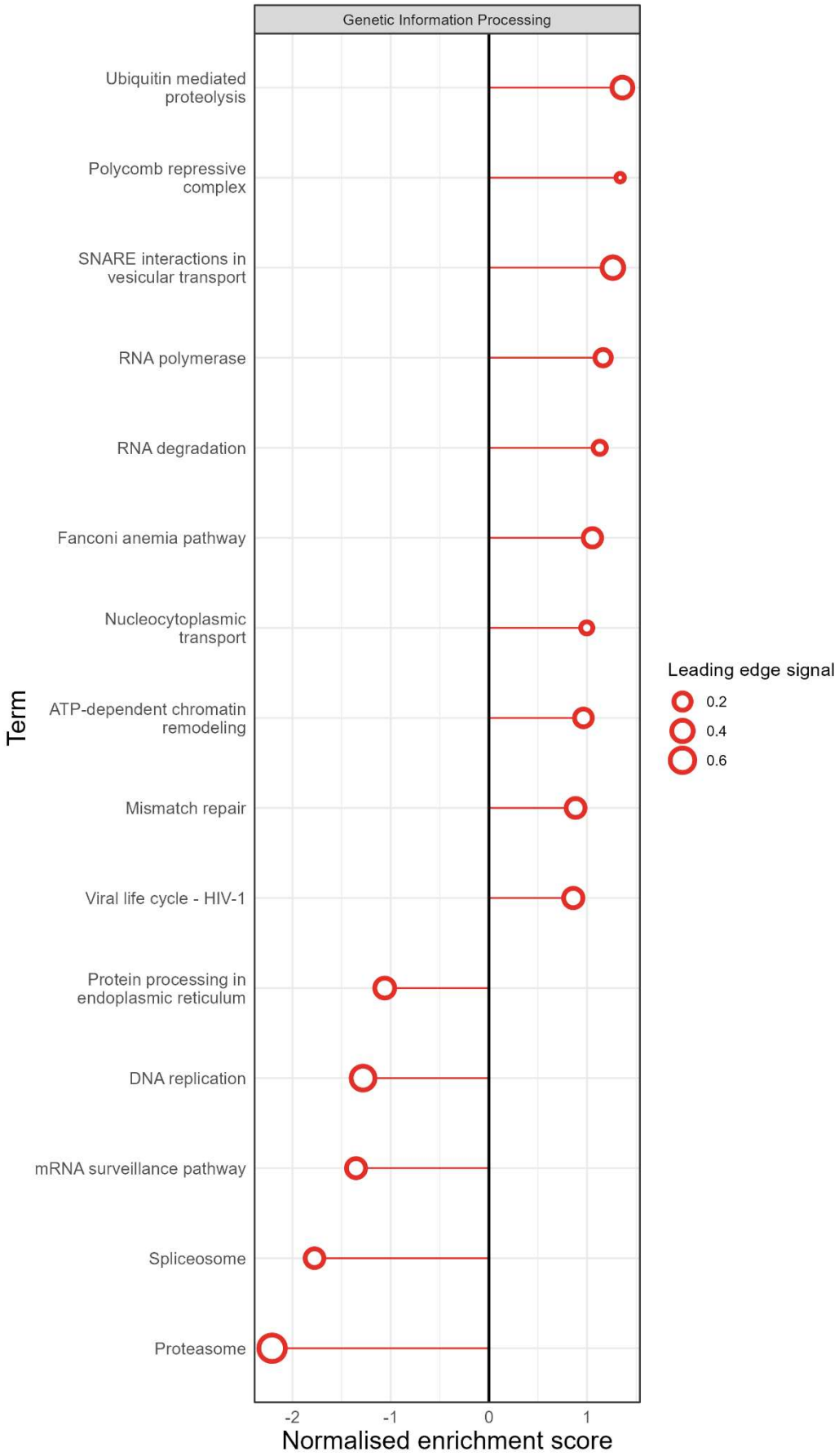
Supplementary Figure 36 ***Gene set enrichment analysis (GSEA) of transcriptome contributions to proteome levels during HoxB8 neutrophil differentiation, using a linear regression approach.*** Genes were pre-ranked based on statistically significant coefficients corresponding to transcriptome variables used to predict protein abundance. Gene ontology (GO) cellular component (CC) terms are shown.



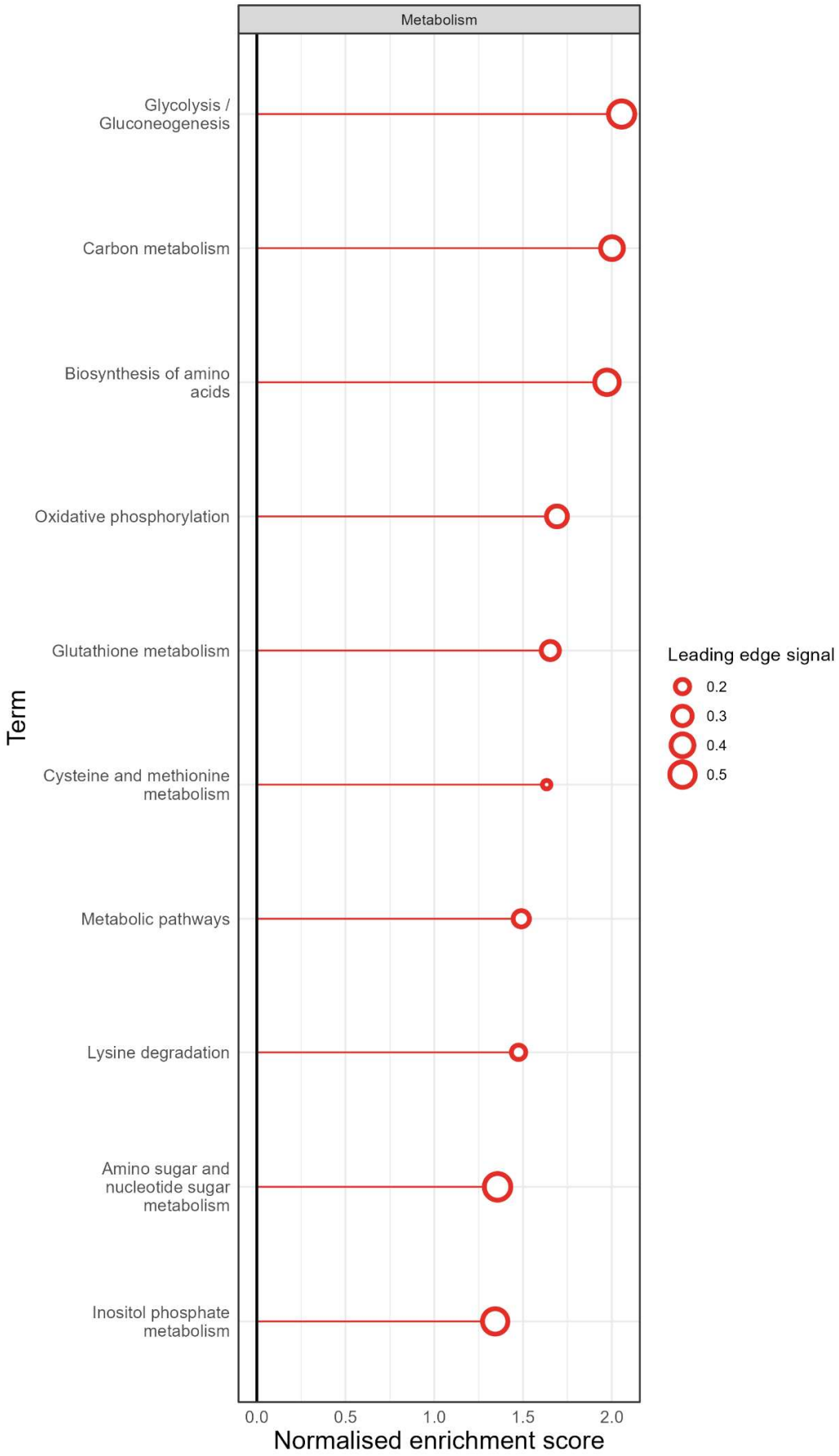
Supplementary Figure 37 ***Gene set enrichment analysis (GSEA) of transcriptome contributions to proteome levels during HoxB8 neutrophil differentiation, using a linear regression approach.*** Genes were pre-ranked based on statistically significant coefficients corresponding to transcriptome variables used to predict protein abundance. Gene ontology (GO) molecular function (MF) terms are shown.



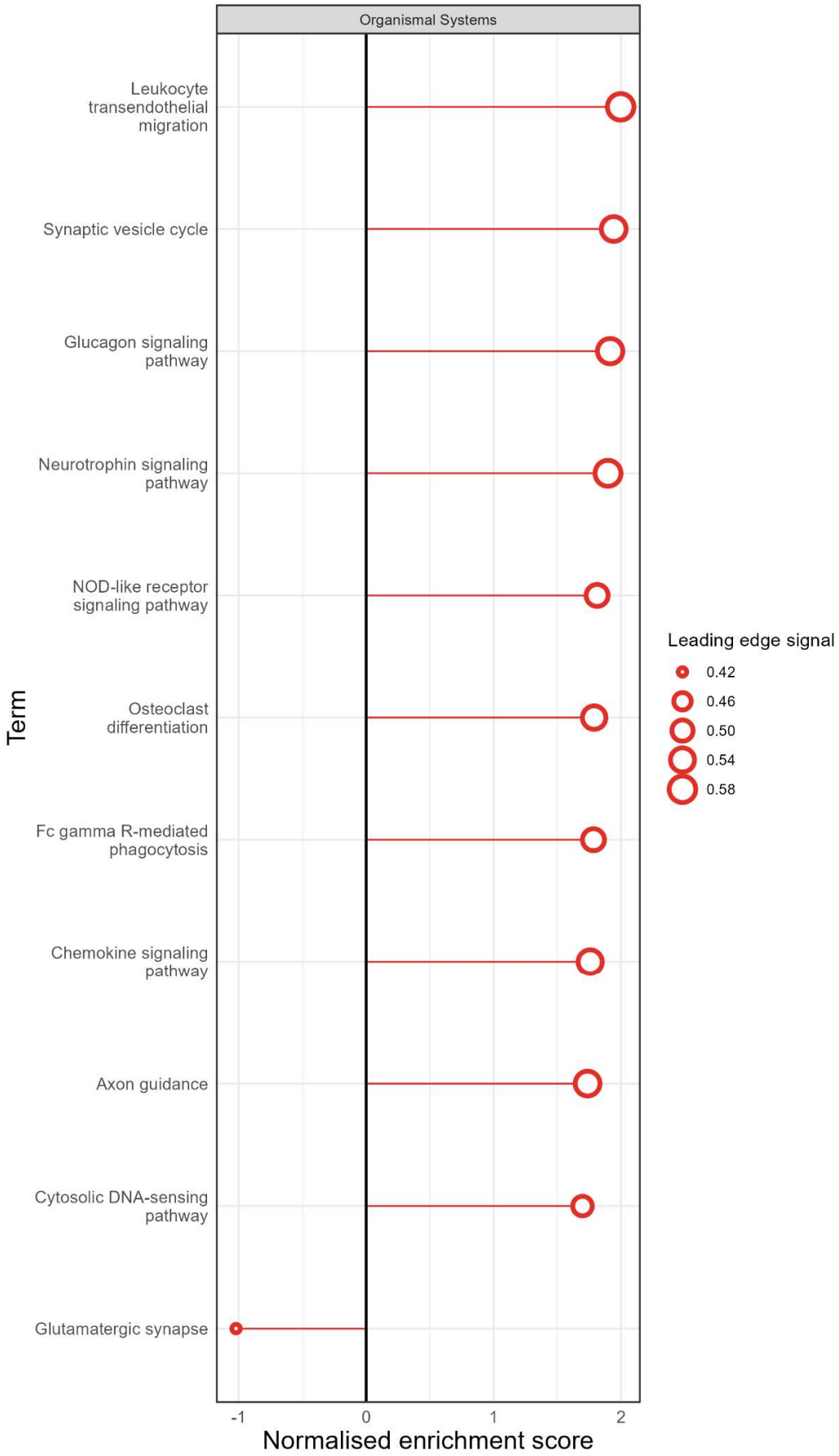
Supplementary Figure 38 ***Gene set enrichment analysis (GSEA) of transcriptome contributions to proteome levels during HoxB8 neutrophil differentiation, using a linear regression approach.*** Genes were pre-ranked based on statistically significant coefficients corresponding to transcriptome variables used to predict protein abundance. The KEGG pathway subcategory 'Cellular Processes' terms are shown.



Supplementary Figure 39 ***Gene set enrichment analysis (GSEA) of transcriptome contributions to proteome levels during HoxB8 neutrophil differentiation, using a linear regression approach.*** Genes were pre-ranked based on statistically significant coefficients corresponding to transcriptome variables used to predict protein abundance. The KEGG pathway subcategory 'Genetic Information Processing' terms are shown.

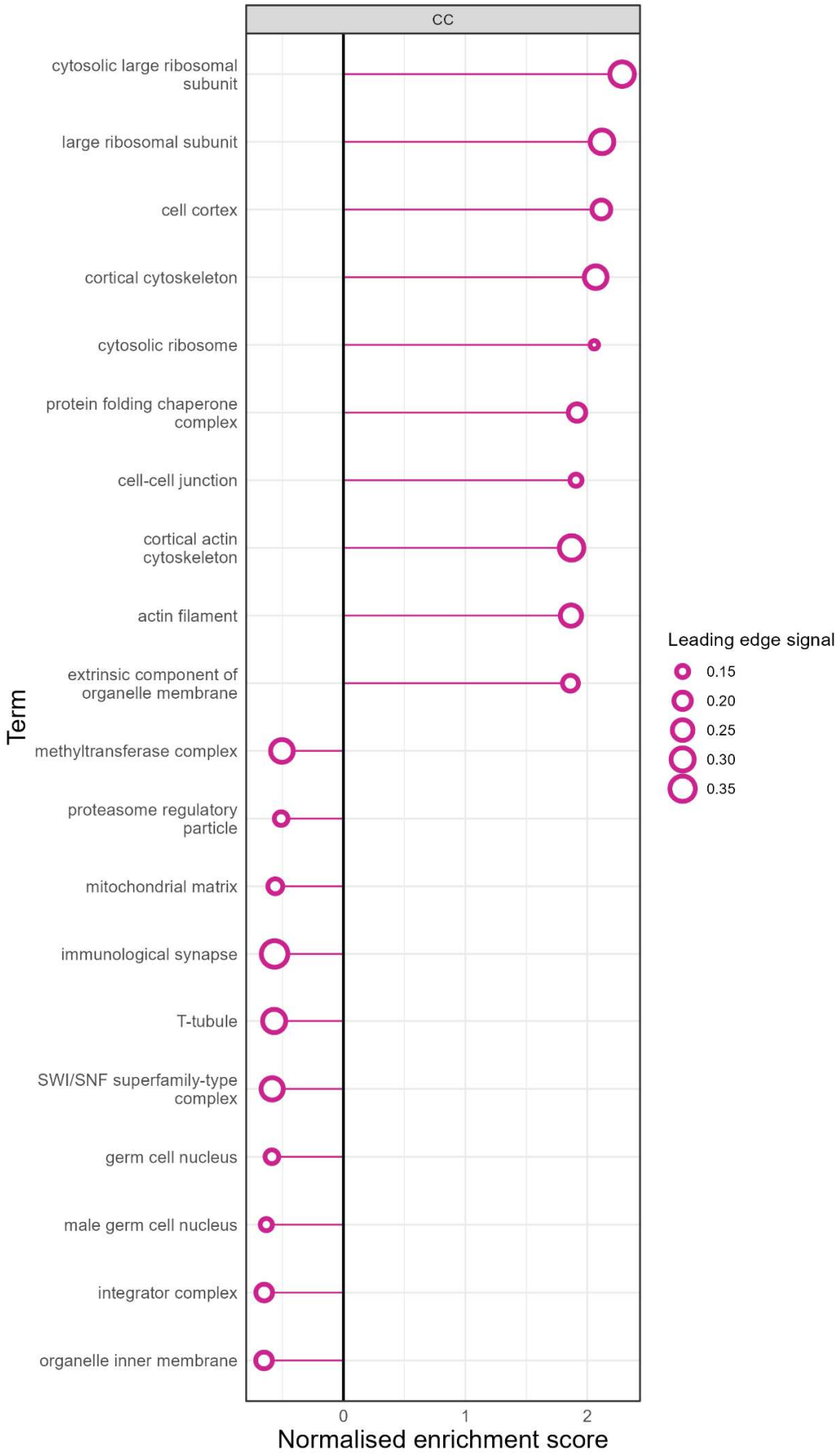


Supplementary Figure 40 ***Gene set enrichment analysis (GSEA) of transcriptome contributions to proteome levels during HoxB8 neutrophil differentiation, using a linear regression approach.*** Genes were pre-ranked based on statistically significant coefficients corresponding to transcriptome variables used to predict protein abundance. The KEGG pathway subcategory 'Metabolism' terms are shown.

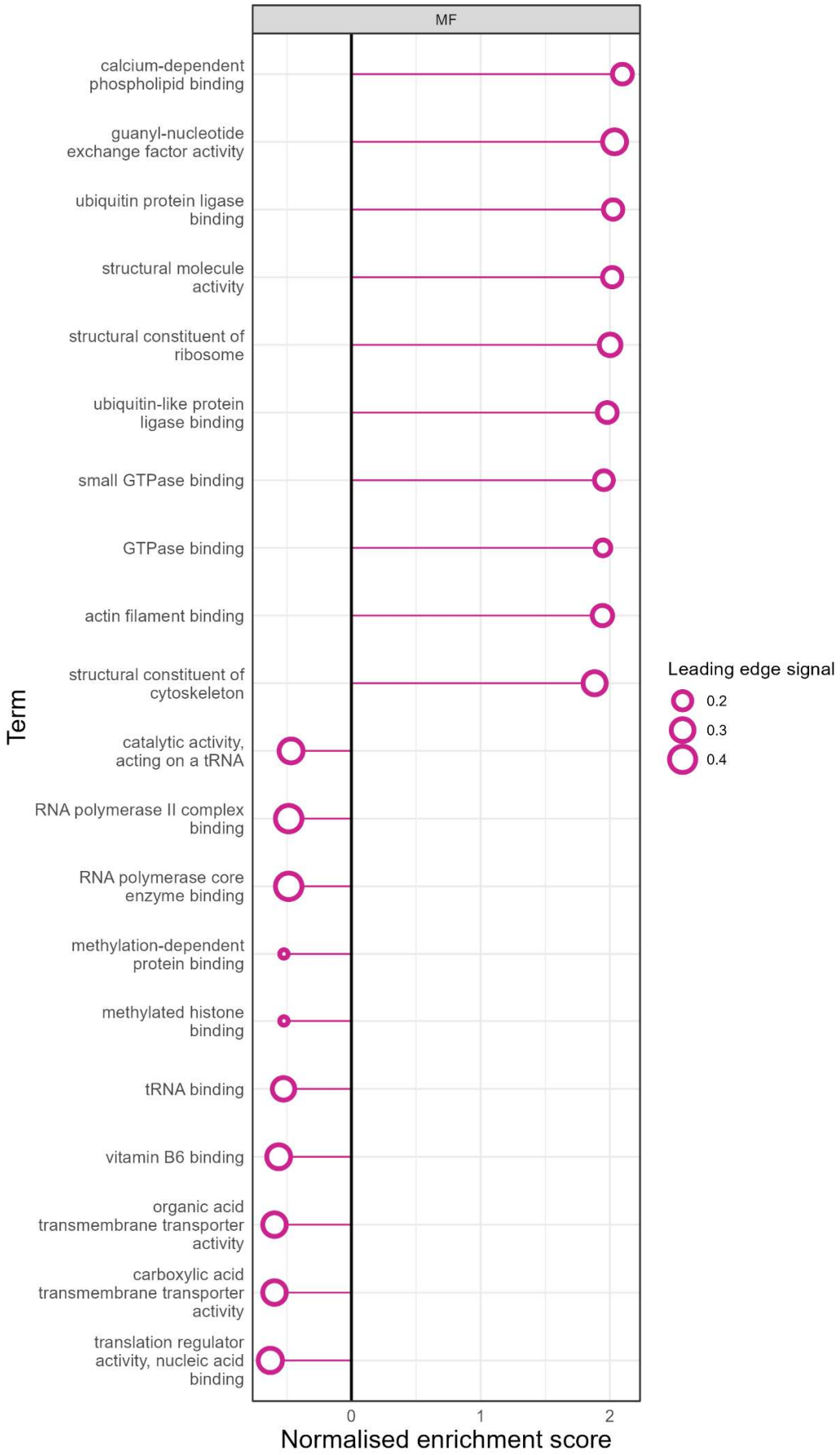


Supplementary Figure 41 ***Gene set enrichment analysis (GSEA) of transcriptome contributions to proteome levels during HoxB8 neutrophil differentiation, using a linear regression approach.*** Genes were pre-ranked based on statistically significant coefficients corresponding to transcriptome variables used to predict protein abundance. The KEGG pathway subcategory 'Organismal Systems' terms are shown.

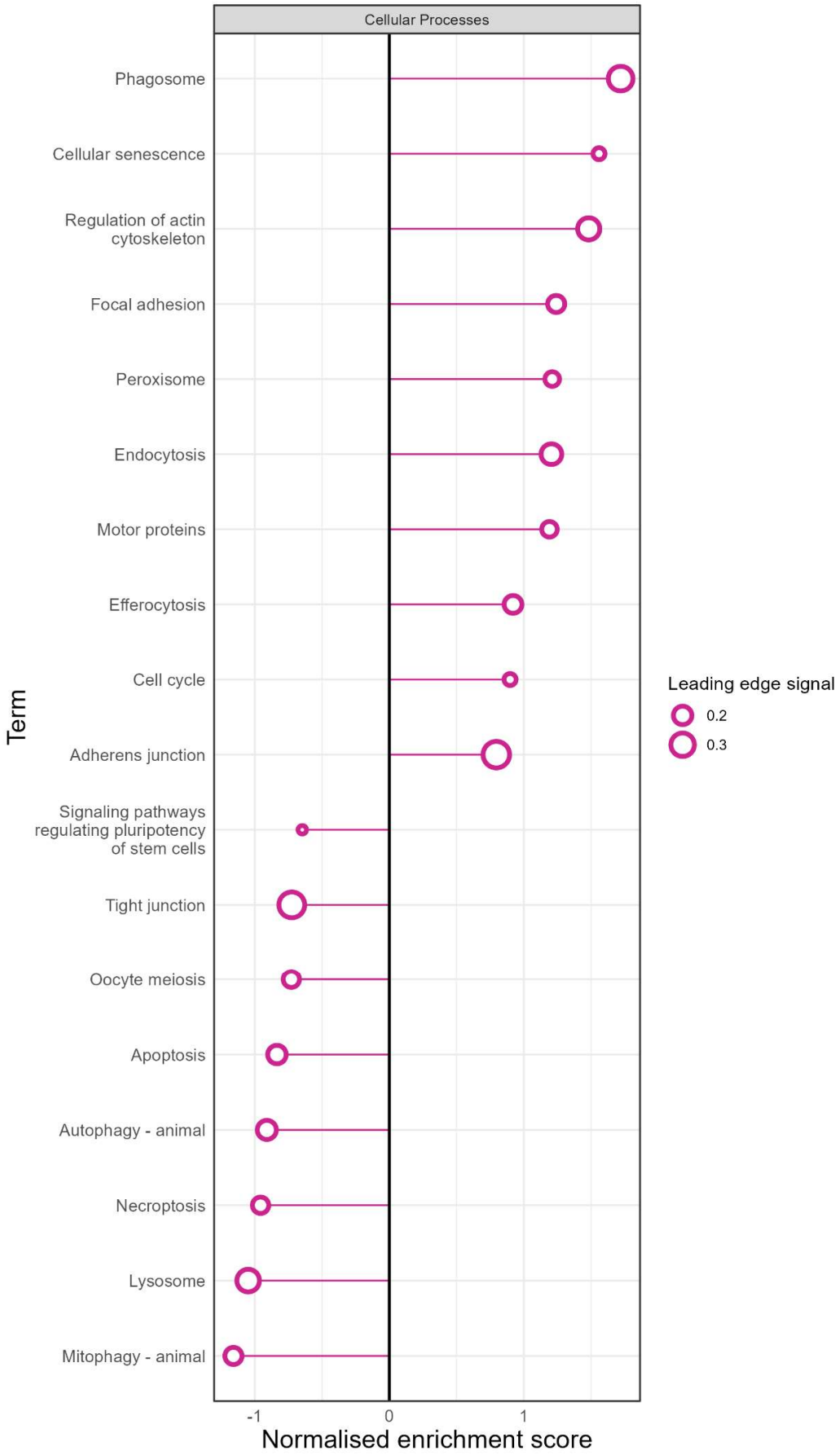
Supplementary Figure 42 ***Gene set enrichment analysis (GSEA) of ubiquitome contributions to proteome levels during HoxB8 neutrophil differentiation, using a linear regression approach.*** Genes were pre-ranked based on statistically significant coefficients corresponding to ubiquitome variables used to predict protein abundance. Gene ontology (GO) biological process (BP) terms are shown.



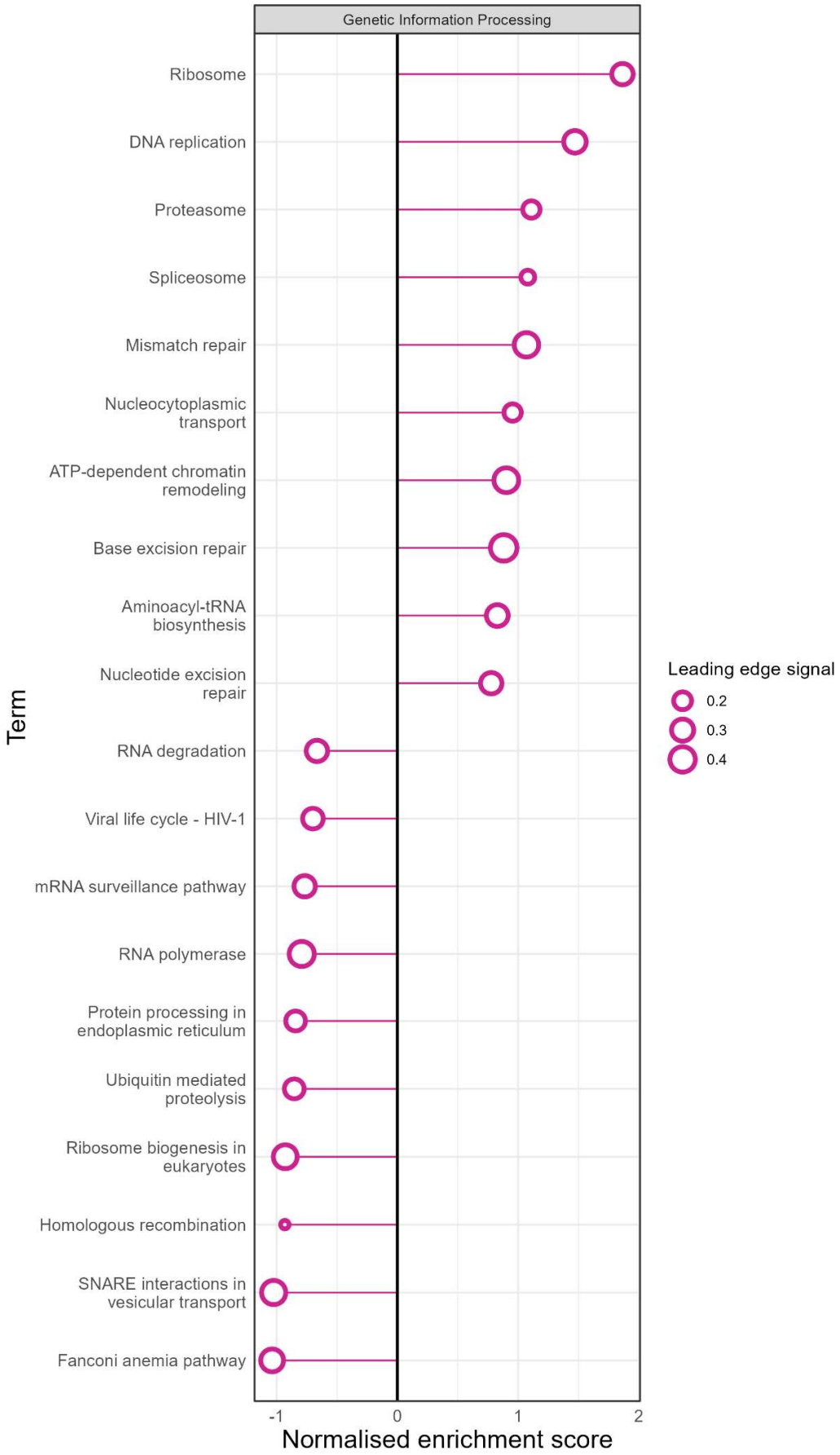
Supplementary Figure 43 ***Gene set enrichment analysis (GSEA) of ubiquitome contributions to proteome levels during HoxB8 neutrophil differentiation, using a linear regression approach.*** Genes were pre-ranked based on statistically significant coefficients corresponding to ubiquitome variables used to predict protein abundance. Gene ontology (GO) cellular component (CC) terms are shown.



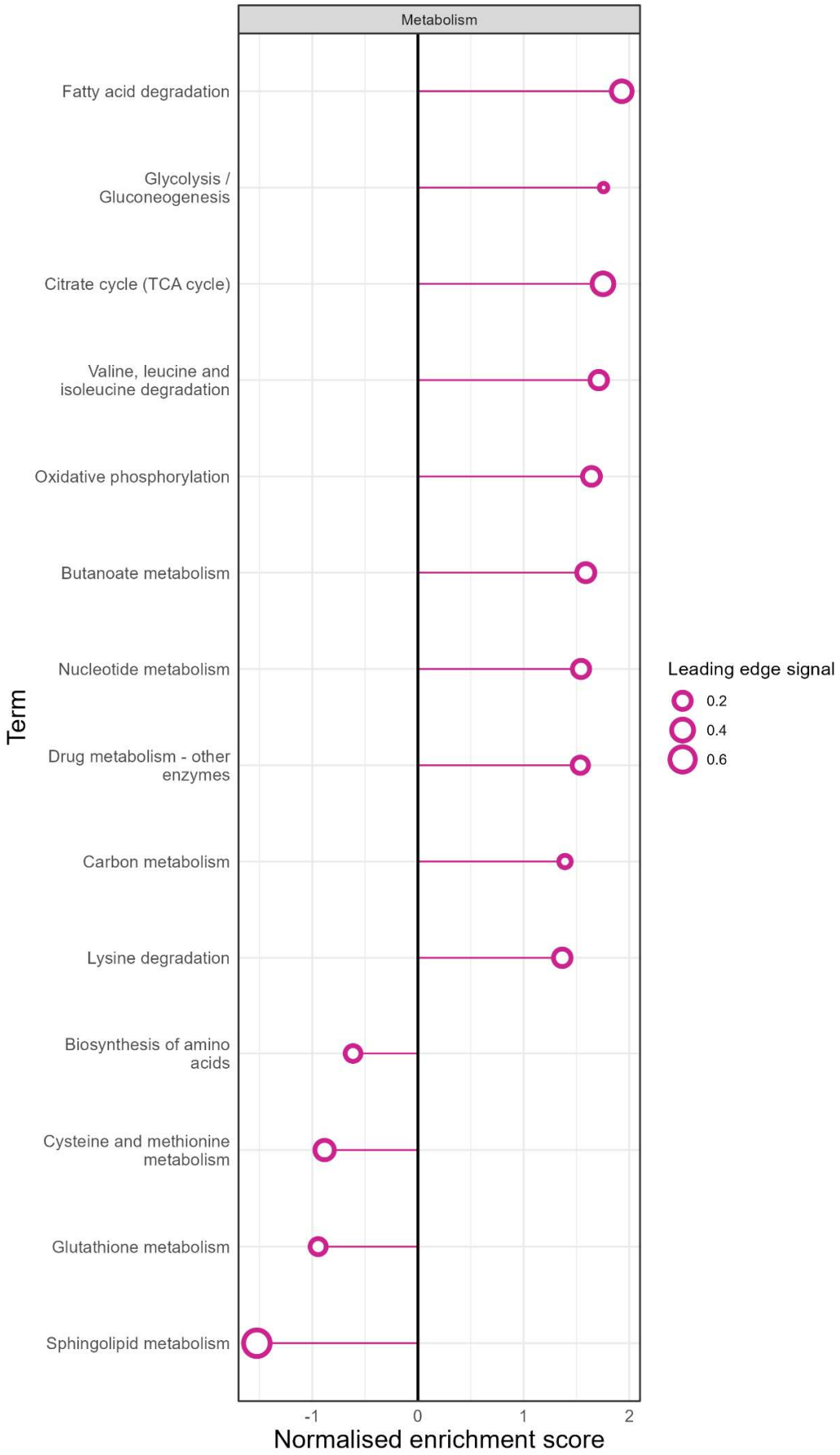
Supplementary Figure 44 ***Gene set enrichment analysis (GSEA) of ubiquitome contributions to proteome levels during HoxB8 neutrophil differentiation, using a linear regression approach.*** Genes were pre-ranked based on statistically significant coefficients corresponding to ubiquitome variables used to predict protein abundance. Gene ontology (GO) molecular function (MF) terms are shown.



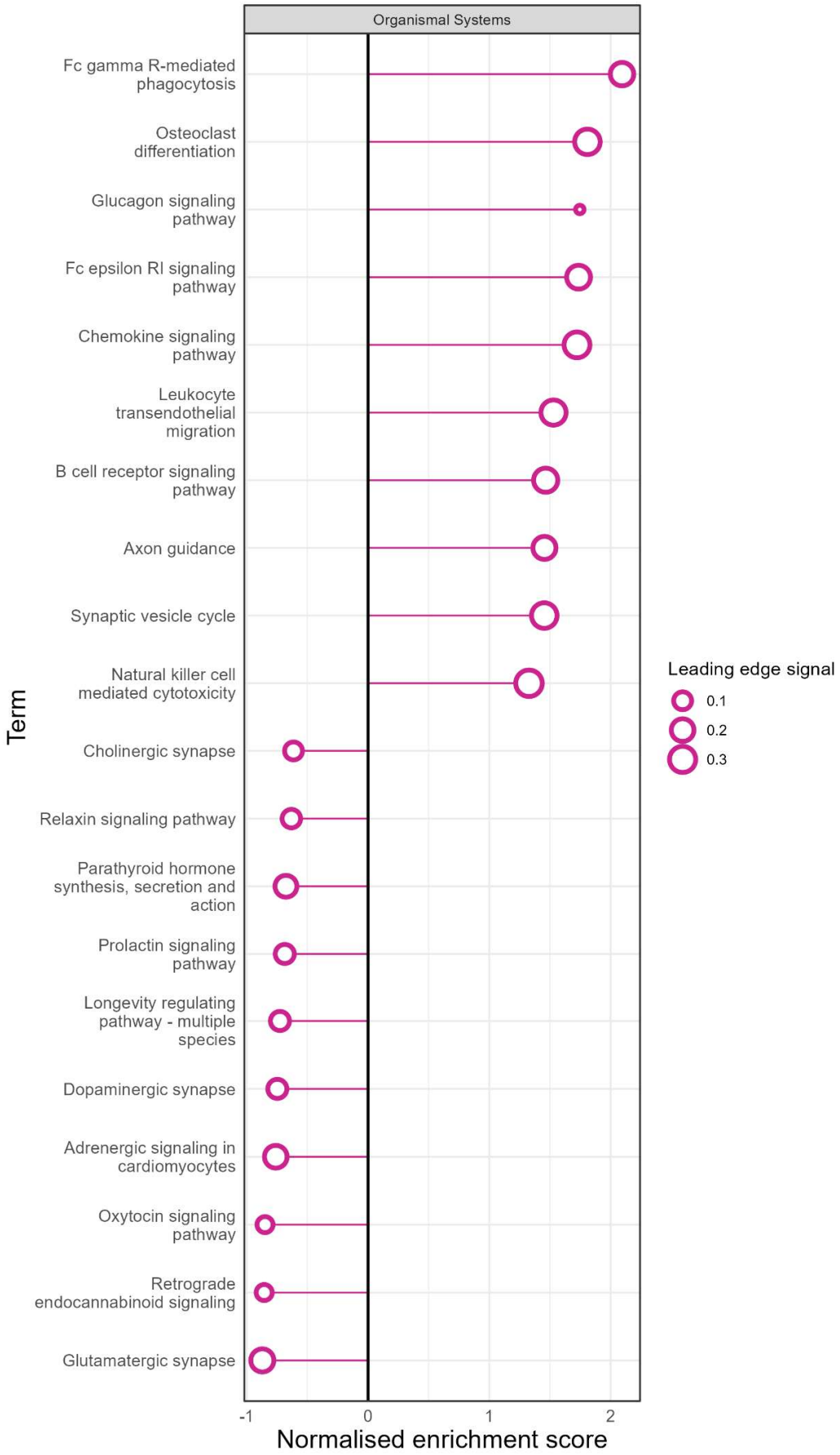
Supplementary Figure 45 ***Gene set enrichment analysis (GSEA) of ubiquitome contributions to proteome levels during HoxB8 neutrophil differentiation, using a linear regression approach.*** Genes were pre-ranked based on statistically significant coefficients corresponding to ubiquitome variables used to predict protein abundance. The KEGG pathway subcategory 'Cellular Processes' terms are shown.



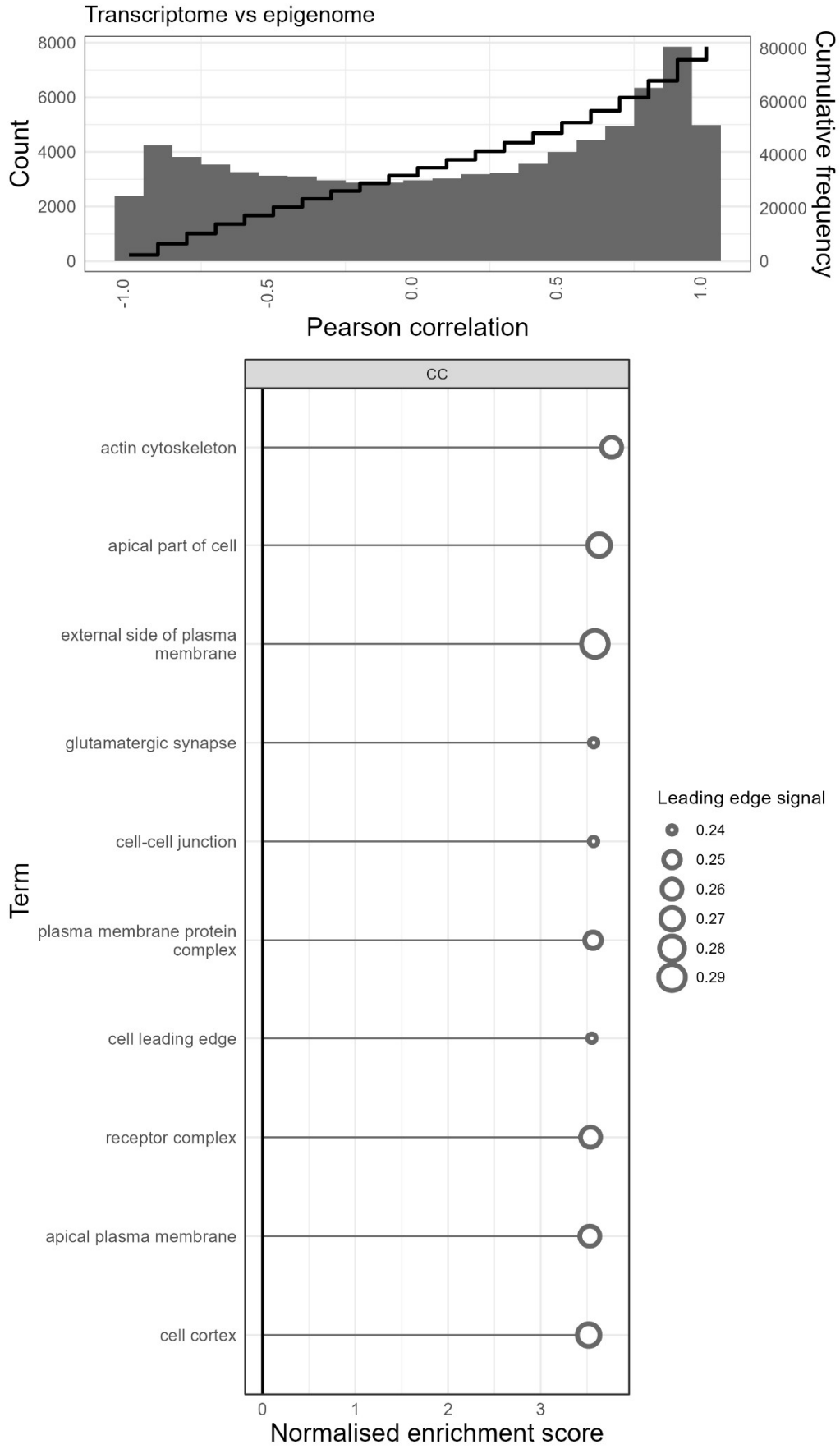
Supplementary Figure 46 ***Gene set enrichment analysis (GSEA) of ubiquitome contributions to proteome levels during HoxB8 neutrophil differentiation, using a linear regression approach.*** Genes were pre-ranked based on statistically significant coefficients corresponding to ubiquitome variables used to predict protein abundance. The KEGG pathway subcategory 'Genetic Information Processing' terms are shown.



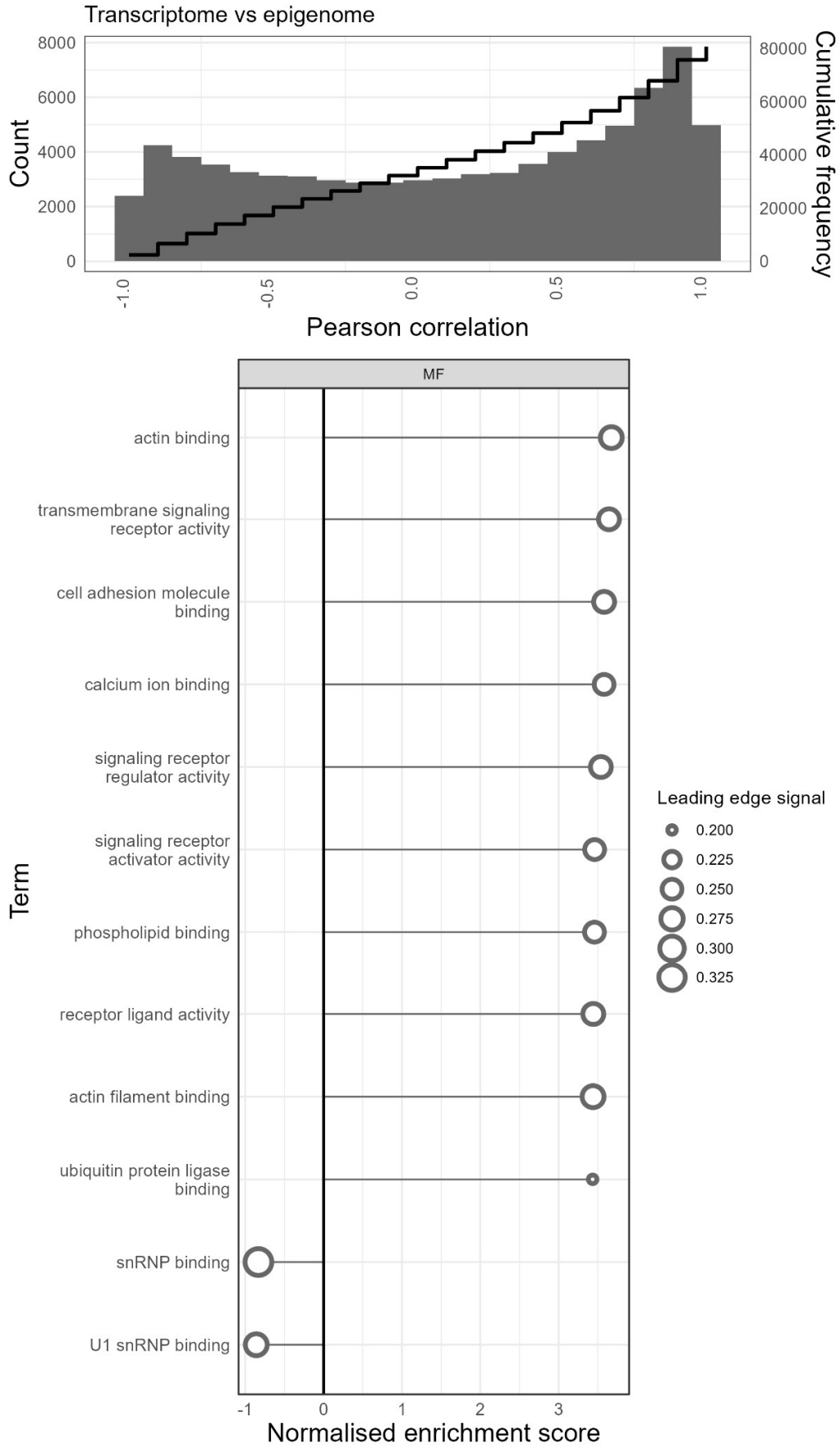
Supplementary Figure 47 ***Gene set enrichment analysis (GSEA) of ubiquitome contributions to proteome levels during HoxB8 neutrophil differentiation, using a linear regression approach.*** Genes were pre-ranked based on statistically significant coefficients corresponding to ubiquitome variables used to predict protein abundance. The KEGG pathway subcategory 'Metabolism' terms are shown.



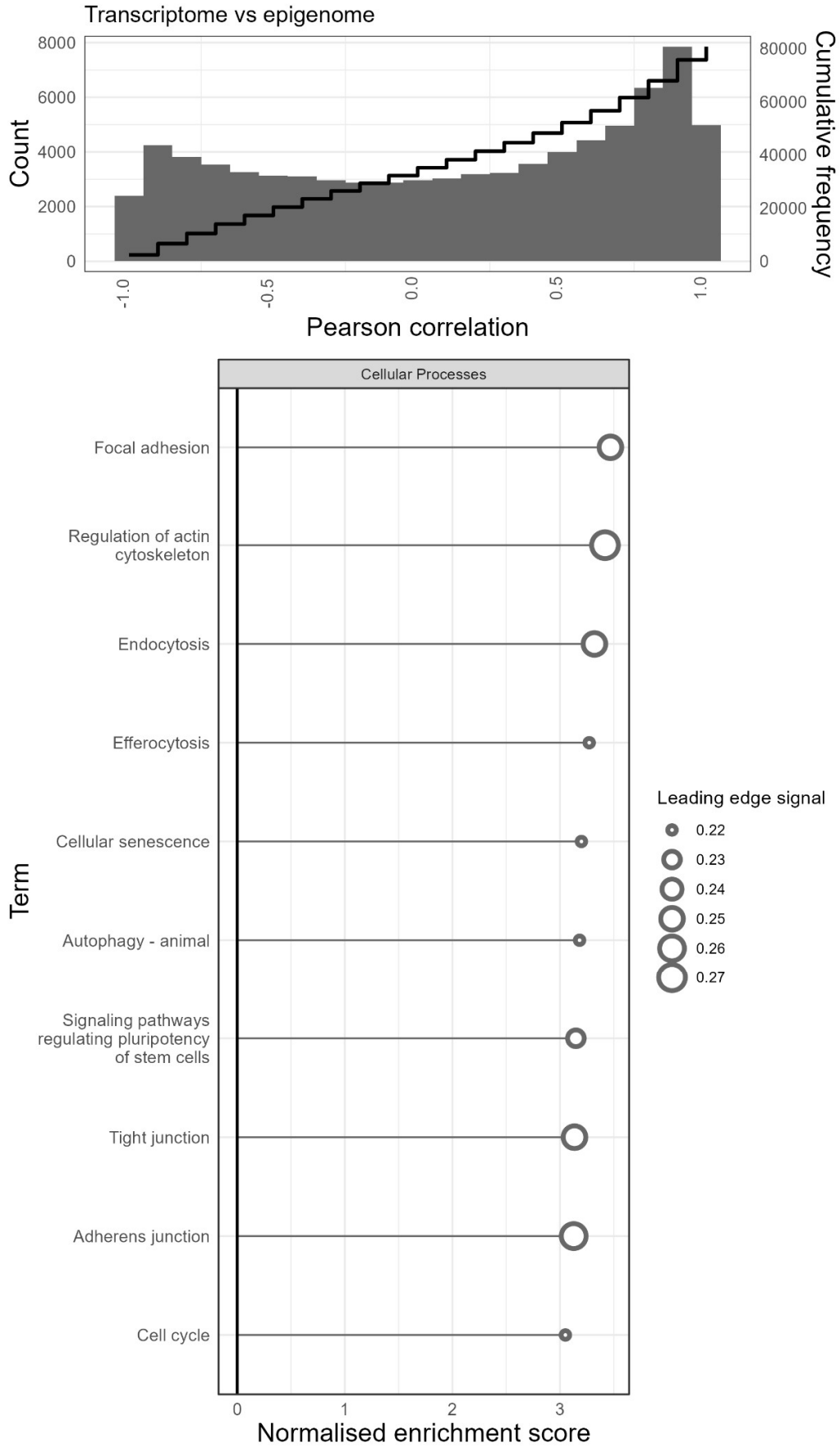
Supplementary Figure 48 ***Gene set enrichment analysis (GSEA) of ubiquitome contributions to proteome levels during HoxB8 neutrophil differentiation, using a linear regression approach.*** Genes were pre-ranked based on statistically significant coefficients corresponding to ubiquitome variables used to predict protein abundance. The KEGG pathway subcategory 'Organismal Systems' terms are shown.



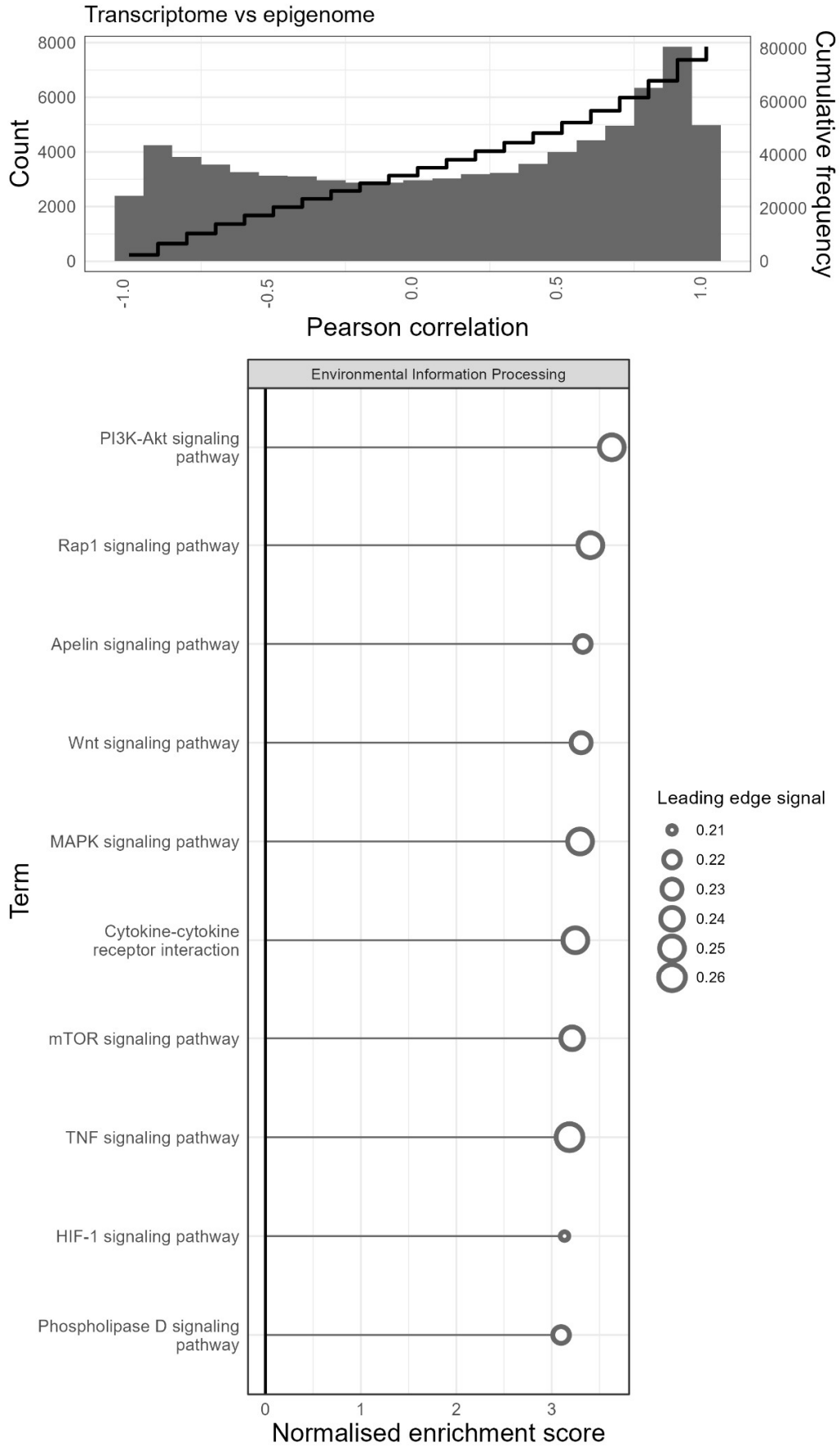
Supplementary Figure 49 ***Gene set enrichment analysis (GSEA) of epigenome-transcriptome data during HoxB8 neutrophil differentiation (GOBP).*** Genes were pre-ranked based on correlation between epigenome levels and corresponding mRNA levels, across all time points of HoxB8 differentiation. Gene ontology (GO) cellular component (CC) terms are shown.



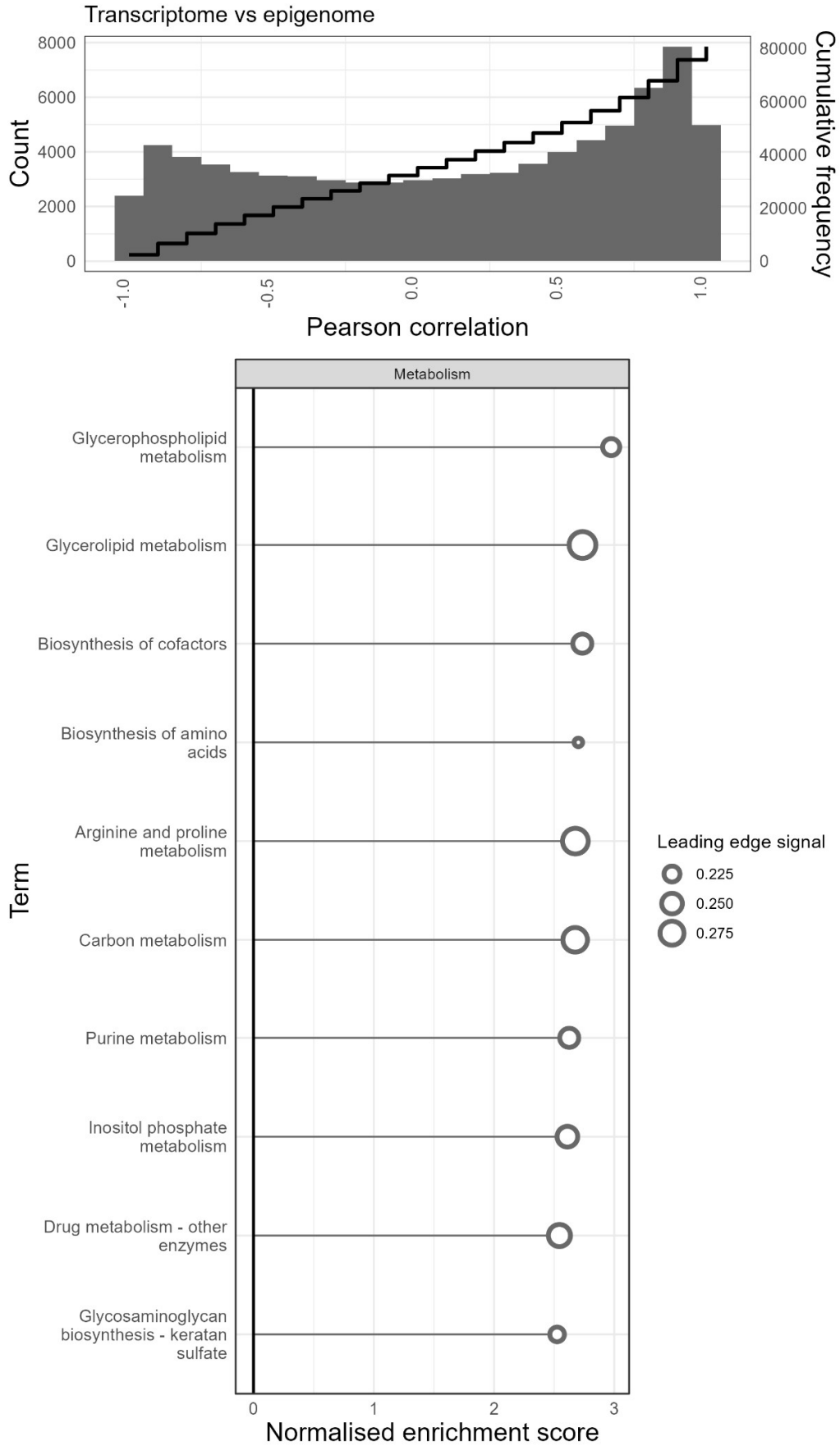
Supplementary Figure 50 ***Gene set enrichment analysis (GSEA) of epigenome-transcriptome data during HoxB8 neutrophil differentiation (GOBP).*** Genes were pre-ranked based on correlation between epigenome levels and corresponding mRNA levels, across all time points of HoxB8 differentiation. Gene ontology (GO) molecular function (MF) terms are shown.



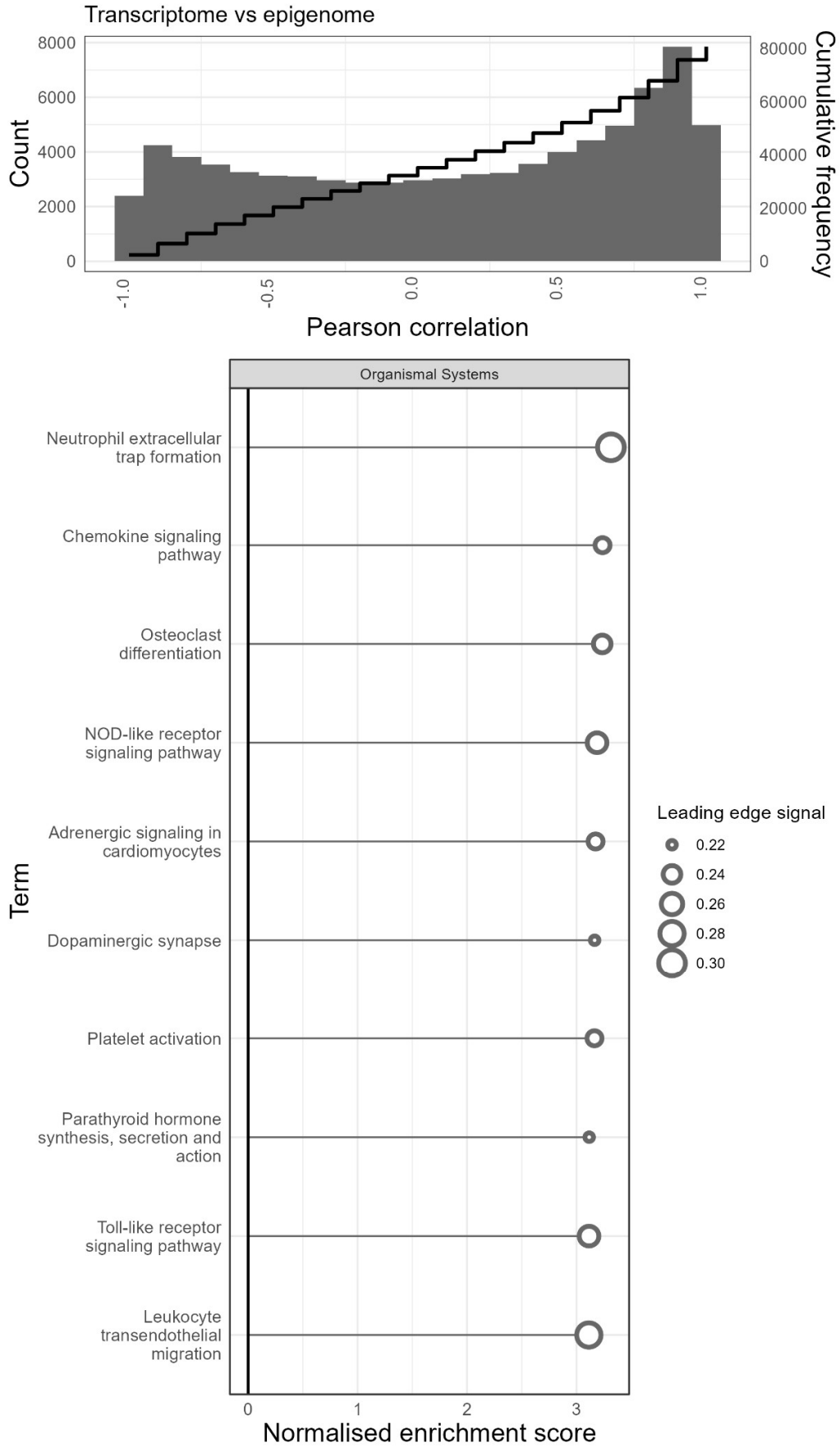
Supplementary Figure 51 ***Gene set enrichment analysis (GSEA) of epigenome-transcriptome data during HoxB8 neutrophil differentiation (KEGG, Genetic Information Processing).*** Genes were pre-ranked based on correlation between epigenome levels and corresponding mRNA levels, across all time points of HoxB8 differentiation. The KEGG pathway subcategory 'Cellular Process' terms are shown.



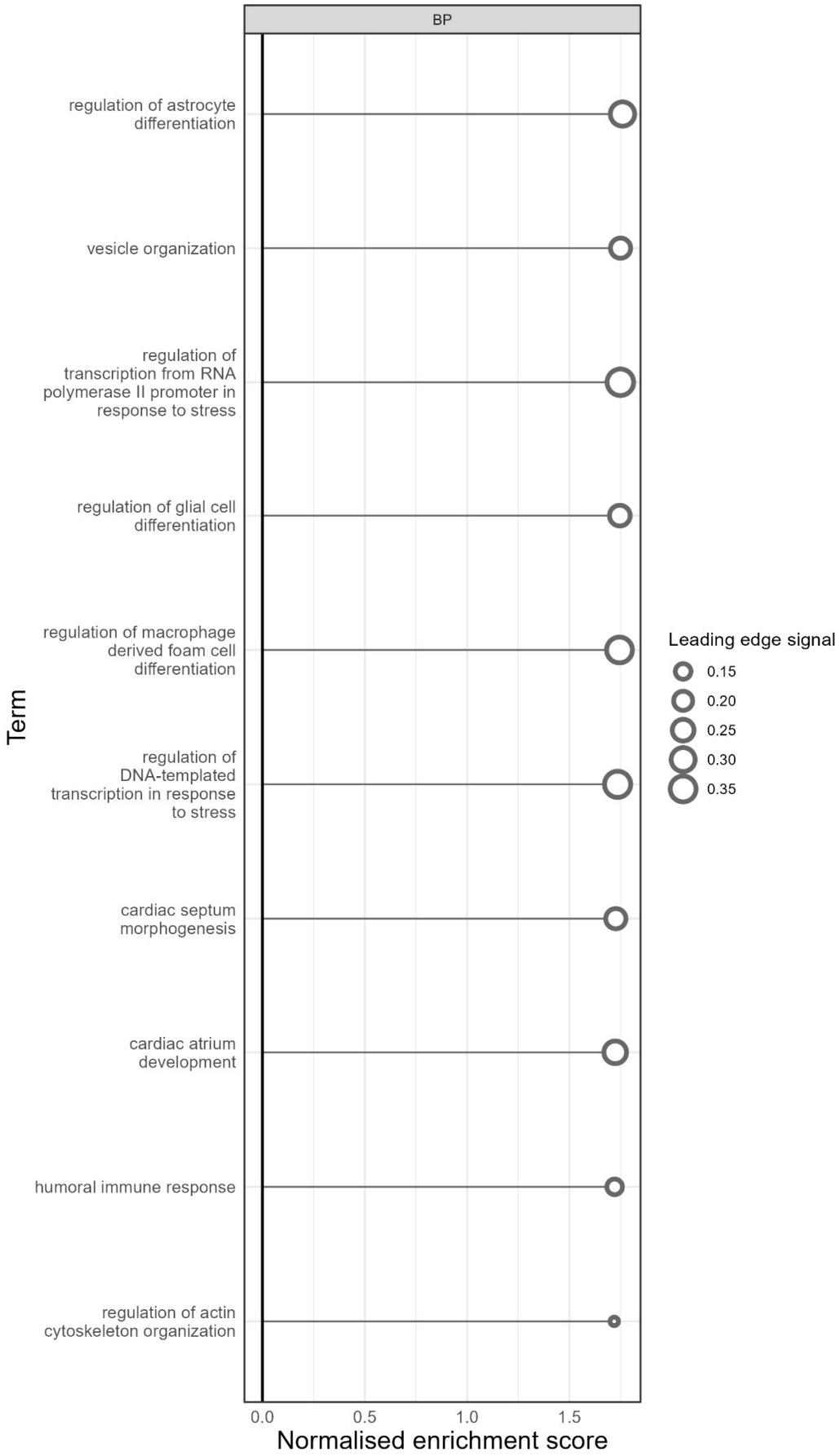
Supplementary Figure 52 ***Gene set enrichment analysis (GSEA) of epigenome-transcriptome data during HoxB8 neutrophil differentiation (KEGG, Genetic Information Processing).*** Genes were pre-ranked based on correlation between epigenome levels and corresponding mRNA levels, across all time points of HoxB8 differentiation. The KEGG pathway subcategory 'Environmental Information Processing' terms are shown.



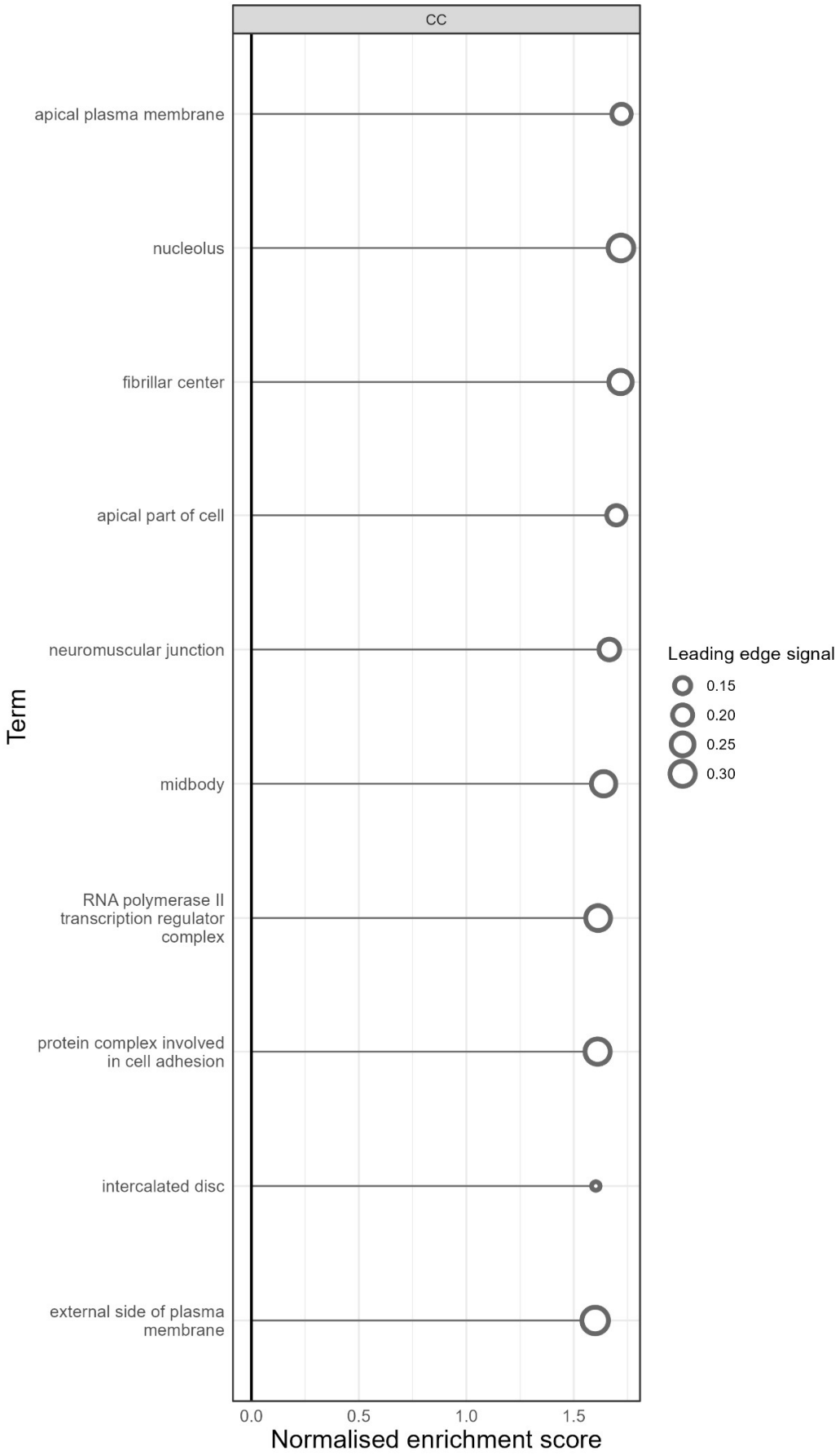
Supplementary Figure 53 ***Gene set enrichment analysis (GSEA) of epigenome-transcriptome data during HoxB8 neutrophil differentiation (KEGG, Genetic Information Processing).*** Genes were pre-ranked based on correlation between epigenome levels and corresponding mRNA levels, across all time points of HoxB8 differentiation. The KEGG pathway subcategory 'Metabolism' terms are shown.



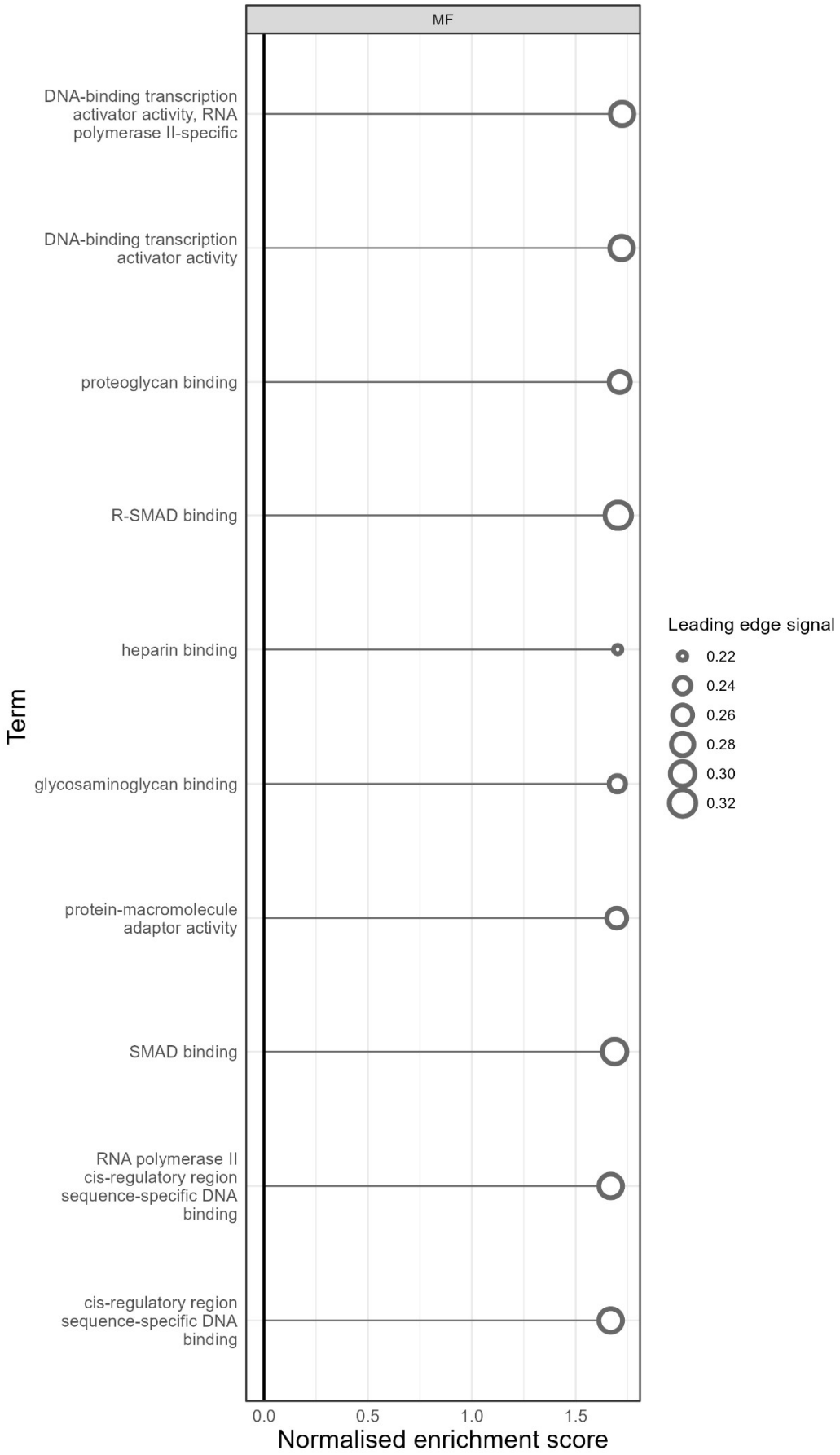
Supplementary Figure 54 ***Gene set enrichment analysis (GSEA) of epigenome-transcriptome data during HoxB8 neutrophil differentiation (KEGG, Genetic Information Processing).*** Genes were pre-ranked based on correlation between epigenome levels and corresponding mRNA levels, across all time points of HoxB8 differentiation. The KEGG pathway subcategory 'Organismal Systems' terms are shown.



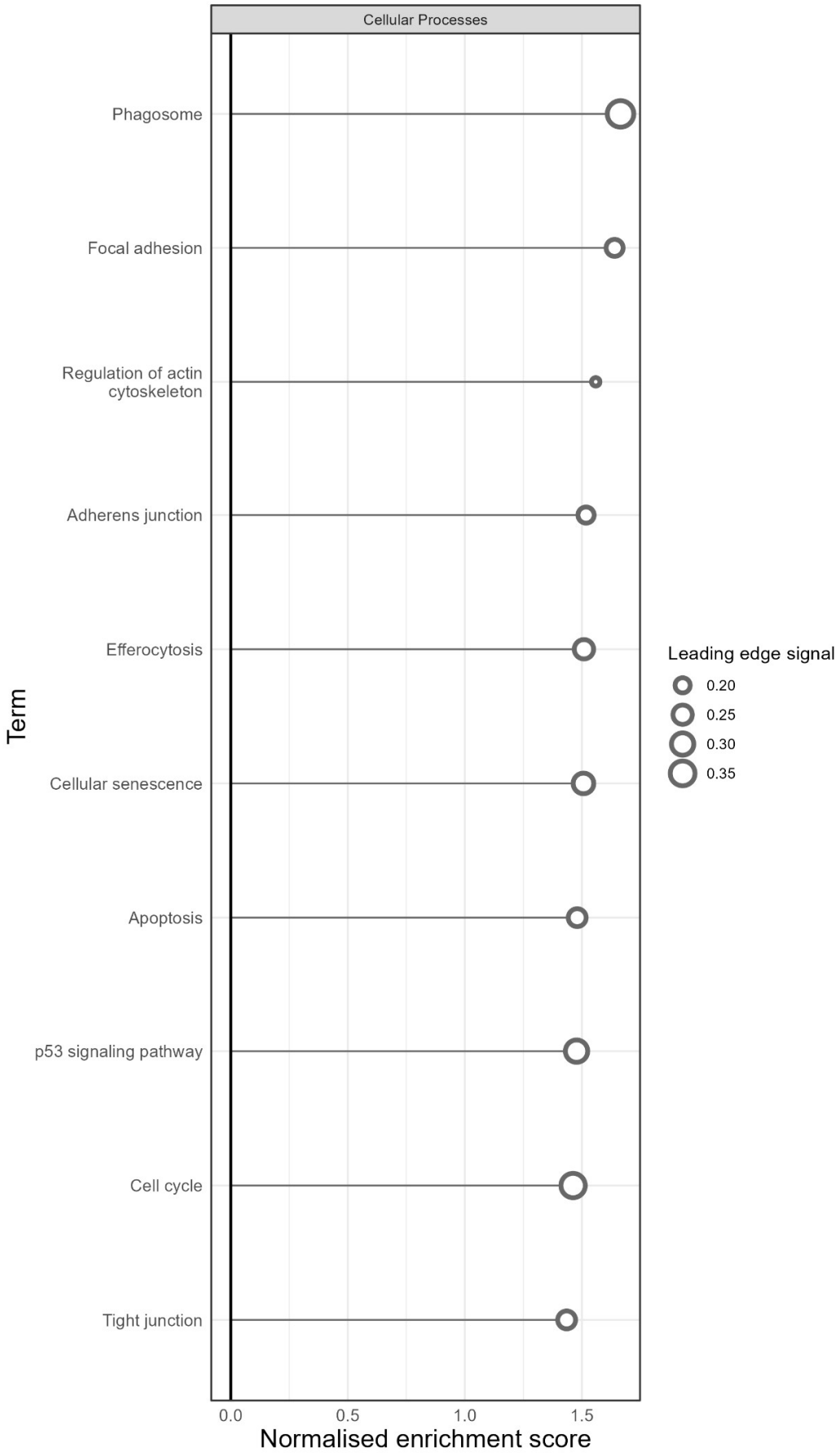
Supplementary Figure 55 ***Gene set enrichment analysis (GSEA) of epigenome contributions to mRNA levels during HoxB8 neutrophil differentiation, using a linear regression approach.*** Genes were pre-ranked based on statistically significant coefficients corresponding to epigenome variables used to predict mRNA levels. Gene ontology (GO) biological process (BP) terms are shown.



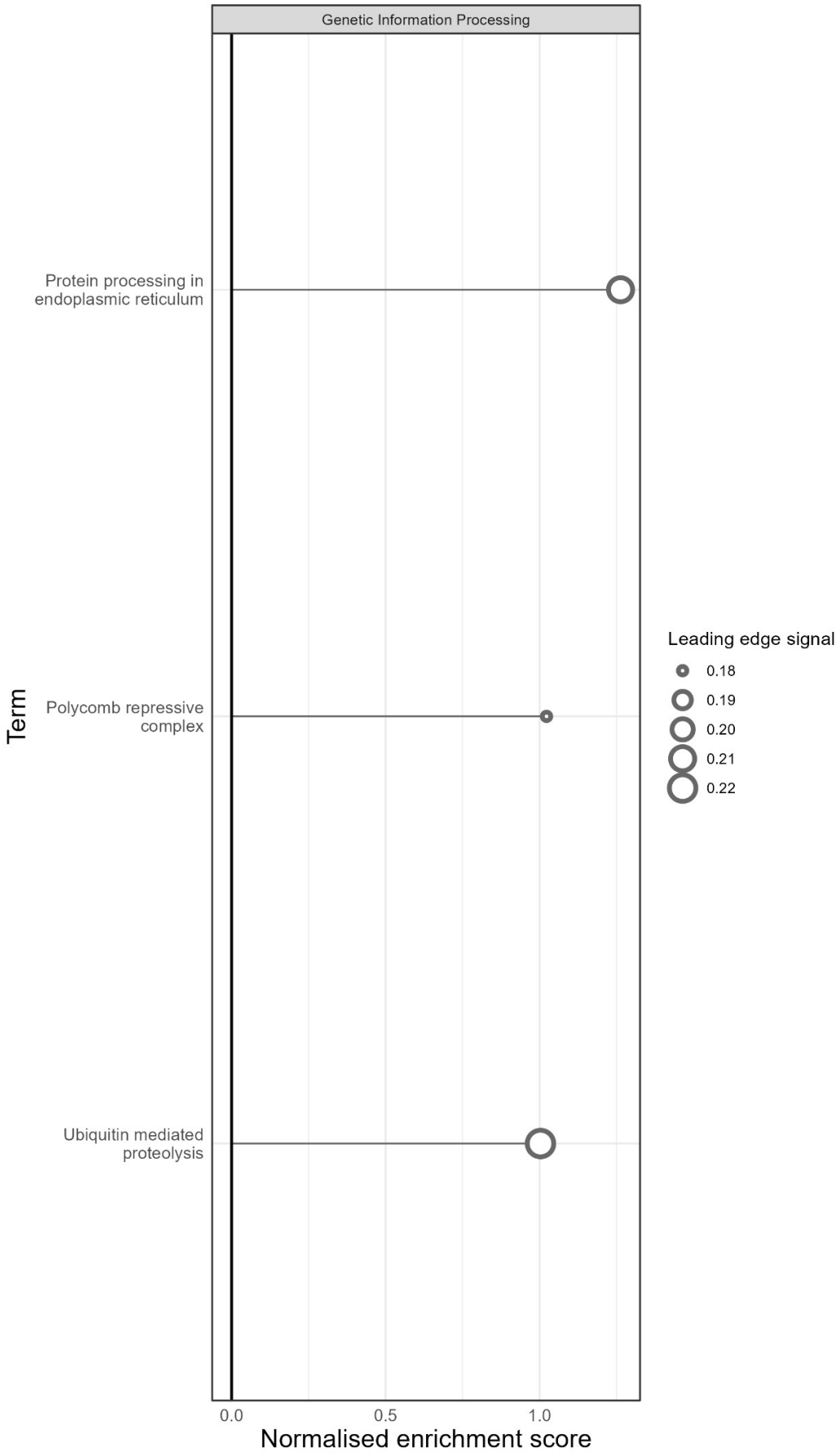
Supplementary Figure 56 ***Gene set enrichment analysis (GSEA) of epigenome contributions to mRNA levels during HoxB8 neutrophil differentiation, using a linear regression approach.*** Genes were pre-ranked based on statistically significant coefficients corresponding to epigenome variables used to predict mRNA levels. Gene ontology (GO) cellular component (CC) terms are shown.



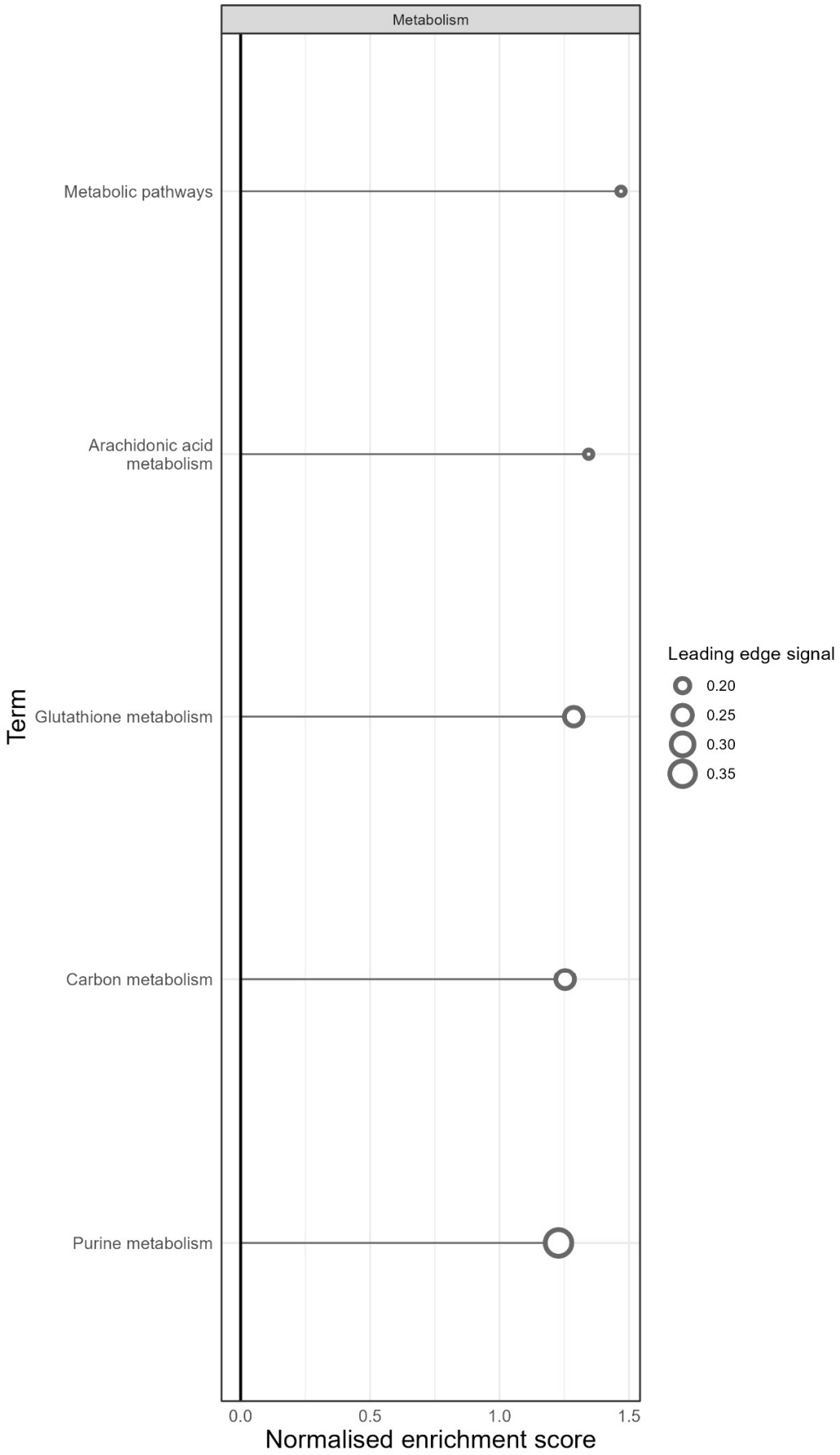
Supplementary Figure 57 ***Gene set enrichment analysis (GSEA) of epigenome contributions to mRNA levels during HoxB8 neutrophil differentiation, using a linear regression approach.*** Genes were pre-ranked based on statistically significant coefficients corresponding to epigenome variables used to predict mRNA levels. Gene ontology (GO) molecular functions (MF) terms are shown.



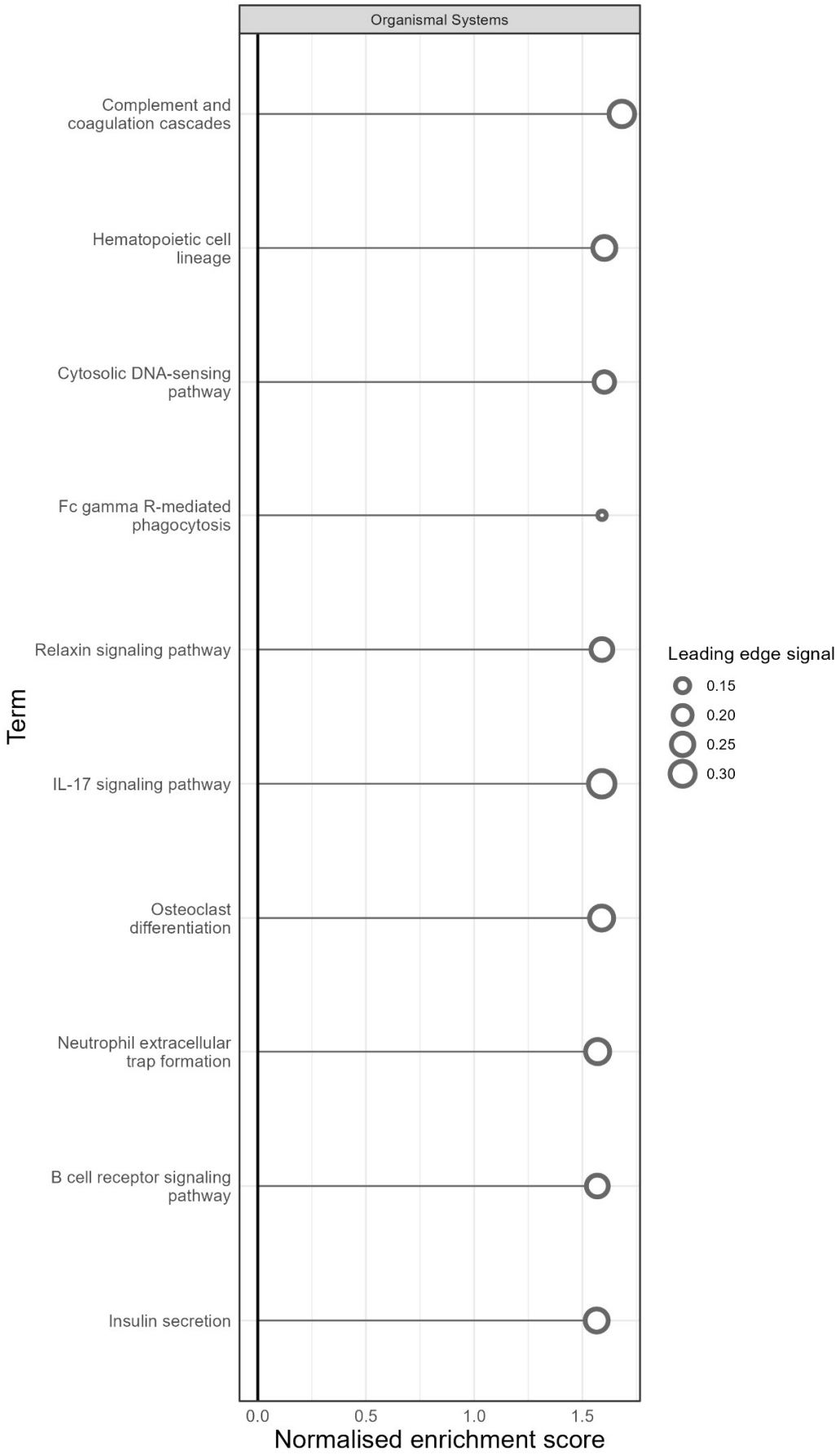
Supplementary Figure 58 ***Gene set enrichment analysis (GSEA) of epigenome contributions to mRNA levels during HoxB8 neutrophil differentiation, using a linear regression approach.*** Genes were pre-ranked based on statistically significant coefficients corresponding to epigenome variables used to predict mRNA levels. The KEGG pathway subcategory 'Cellular Processes' terms are shown.



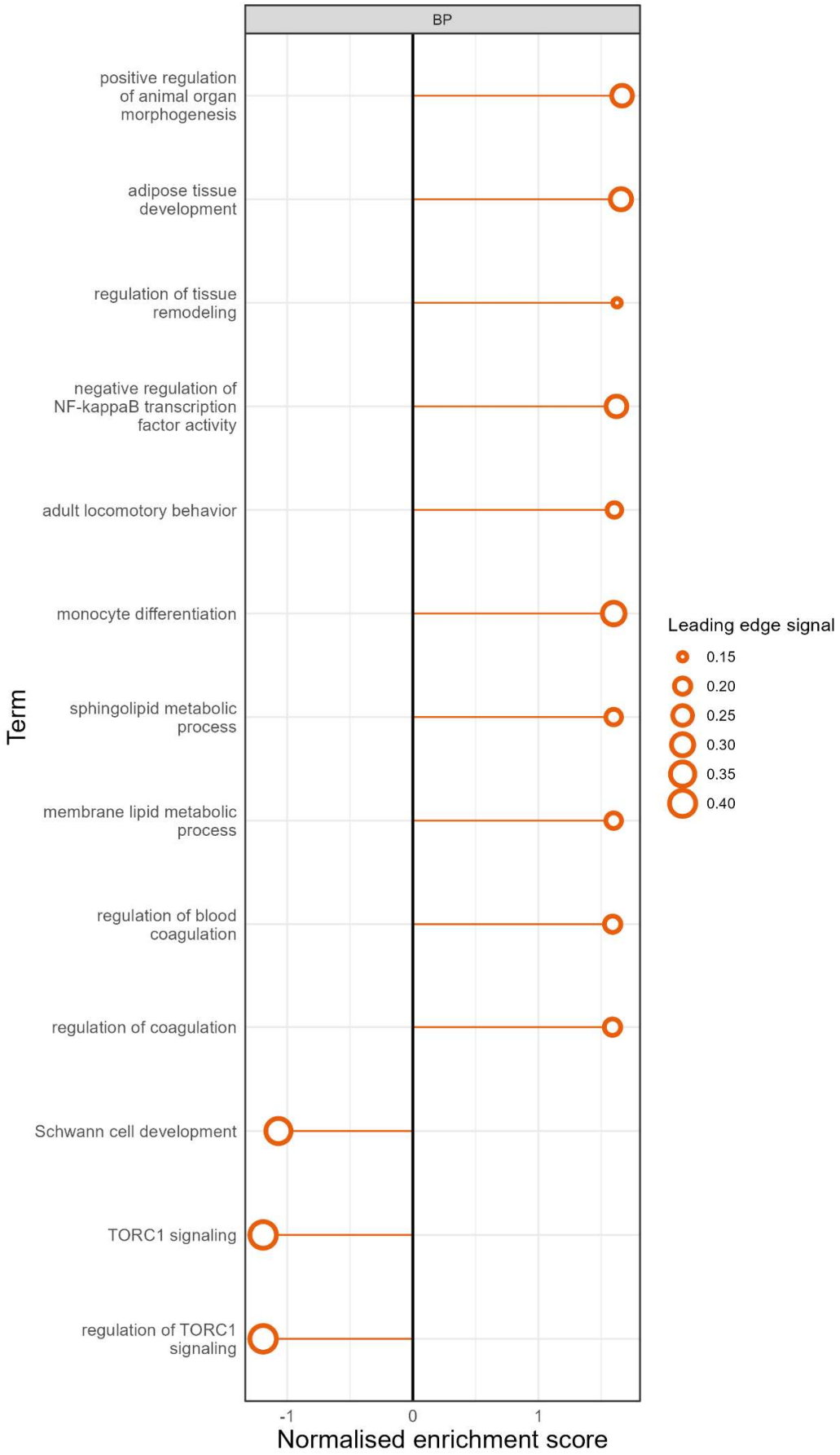
Supplementary Figure 59 ***Gene set enrichment analysis (GSEA) of epigenome contributions to mRNA levels during HoxB8 neutrophil differentiation, using a linear regression approach.*** Genes were pre-ranked based on statistically significant coefficients corresponding to epigenome variables used to predict mRNA levels. The KEGG pathway subcategory 'Genetic Information Processing' terms are shown.



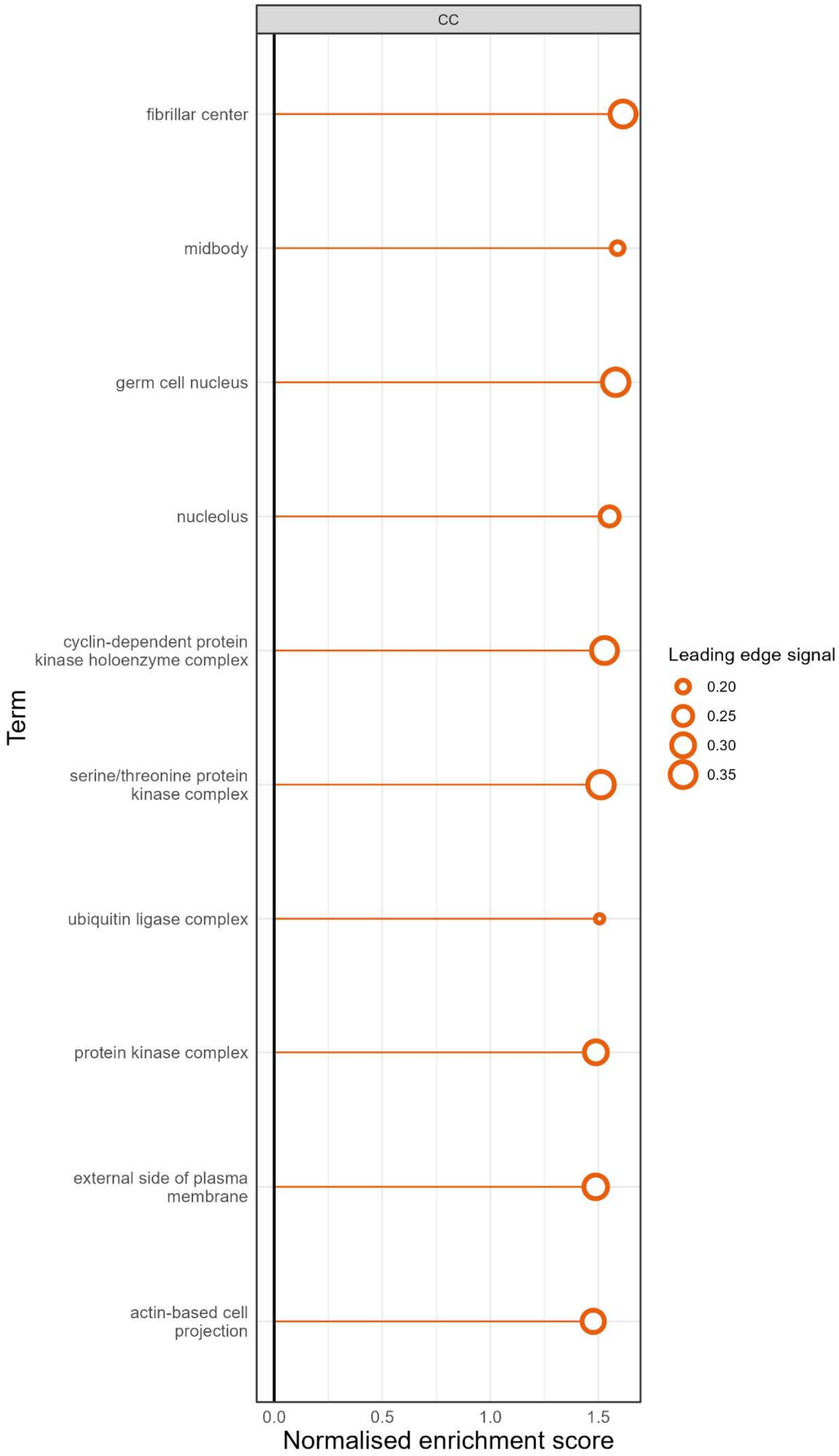
Supplementary Figure 60 ***Gene set enrichment analysis (GSEA) of epigenome contributions to mRNA levels during HoxB8 neutrophil differentiation, using a linear regression approach.*** Genes were pre-ranked based on statistically significant coefficients corresponding to epigenome variables used to predict mRNA levels. The KEGG pathway subcategory 'Metabolism' terms are shown.



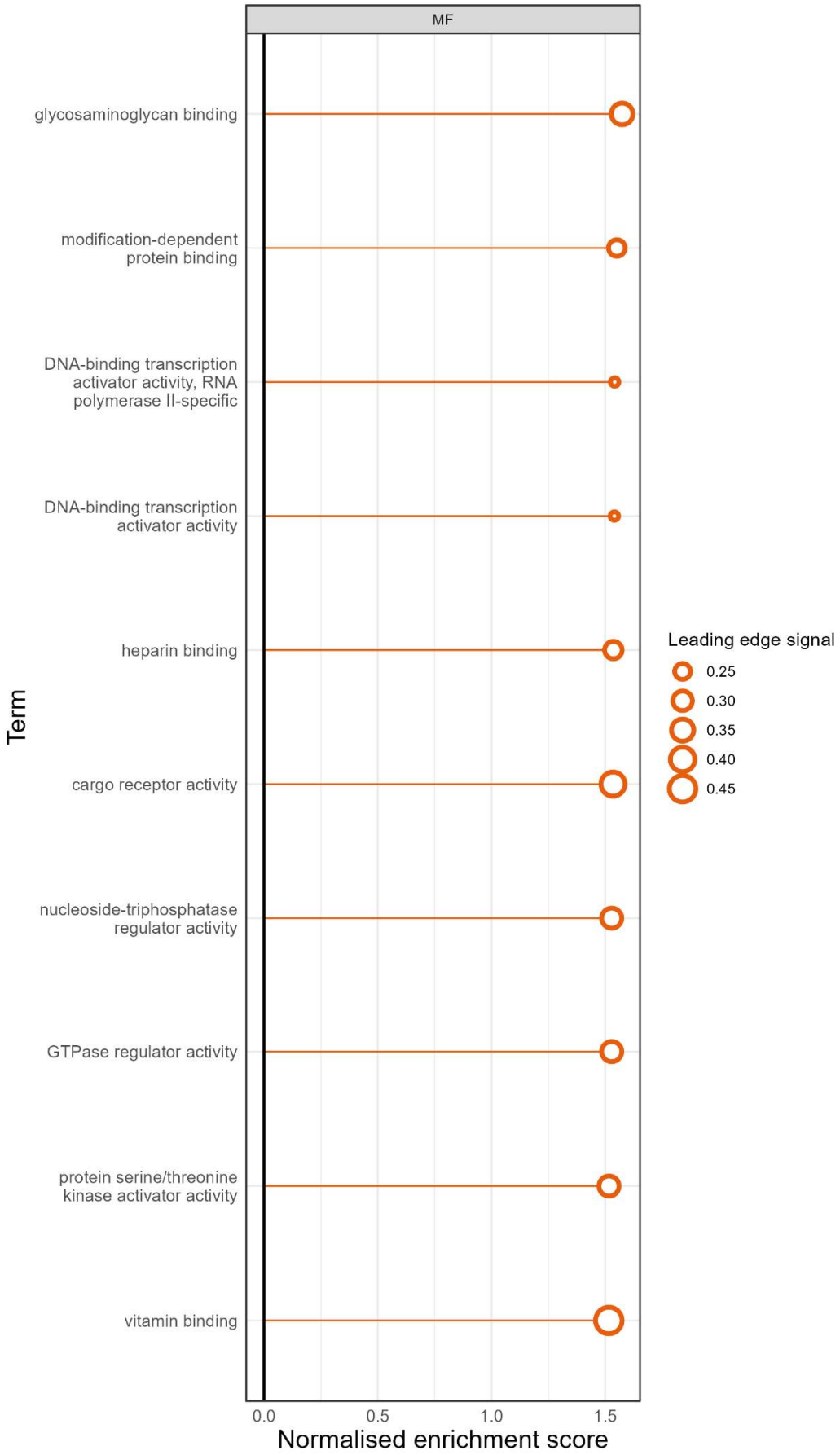
Supplementary Figure 61 ***Gene set enrichment analysis (GSEA) of epigenome contributions to mRNA levels during HoxB8 neutrophil differentiation, using a linear regression approach.*** Genes were pre-ranked based on statistically significant coefficients corresponding to epigenome variables used to predict mRNA levels. The KEGG pathway subcategory 'Organismal Systems' terms are shown.



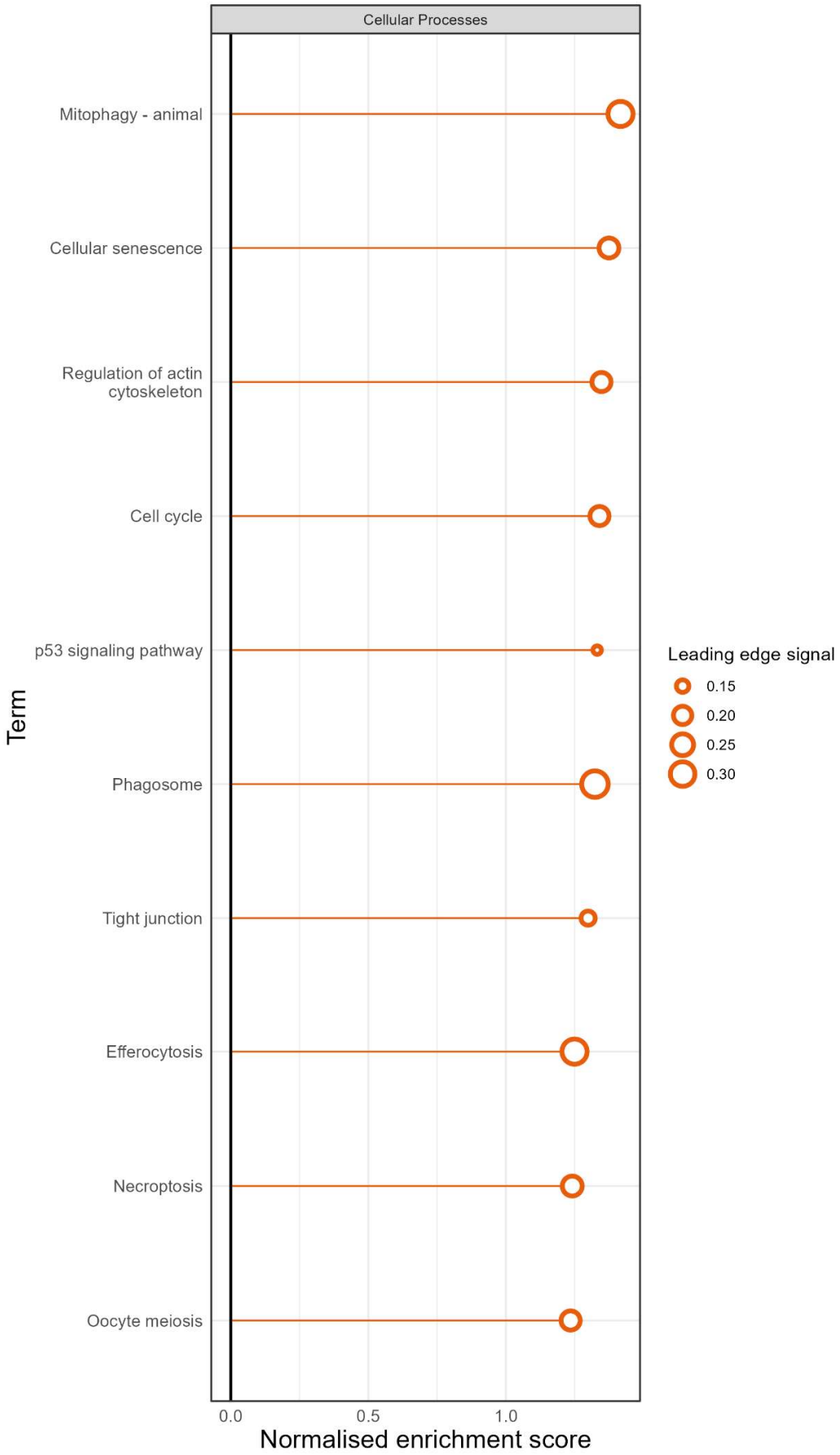
Supplementary Figure 62 ***Gene set enrichment analysis (GSEA) of TF abundance contributions to mRNA levels during HoxB8 neutrophil differentiation, using a linear regression approach.*** Genes were pre-ranked based on statistically significant coefficients corresponding to TF abundance variables used to predict mRNA levels. Gene ontology (GO) biological process (BP) terms are shown.



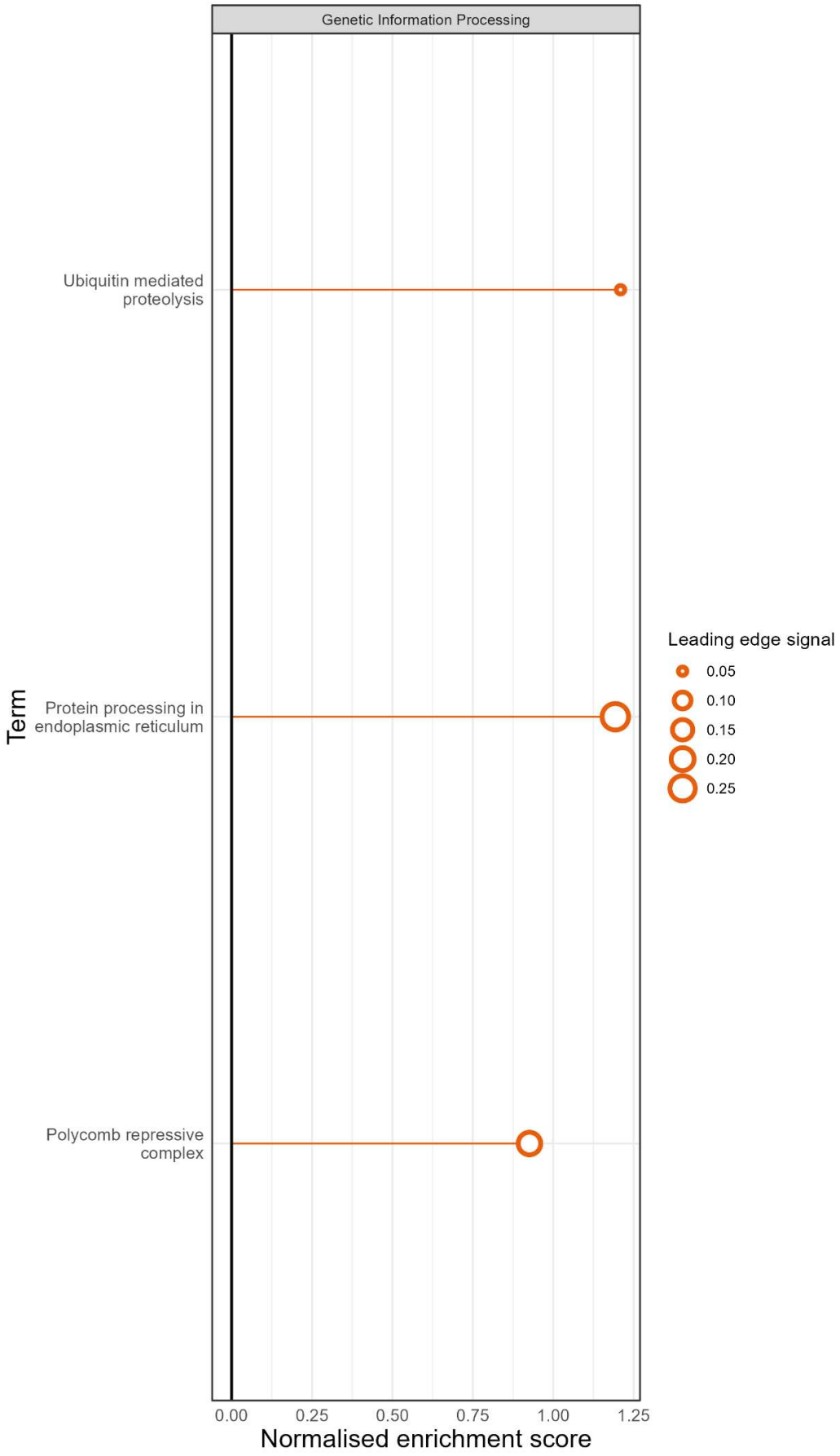
Supplementary Figure 63 ***Gene set enrichment analysis (GSEA) of TF abundance contributions to mRNA levels during HoxB8 neutrophil differentiation, using a linear regression approach.*** Genes were pre-ranked based on statistically significant coefficients corresponding to TF abundance variables used to predict mRNA levels. Gene ontology (GO) cellular component (CC) terms are shown.



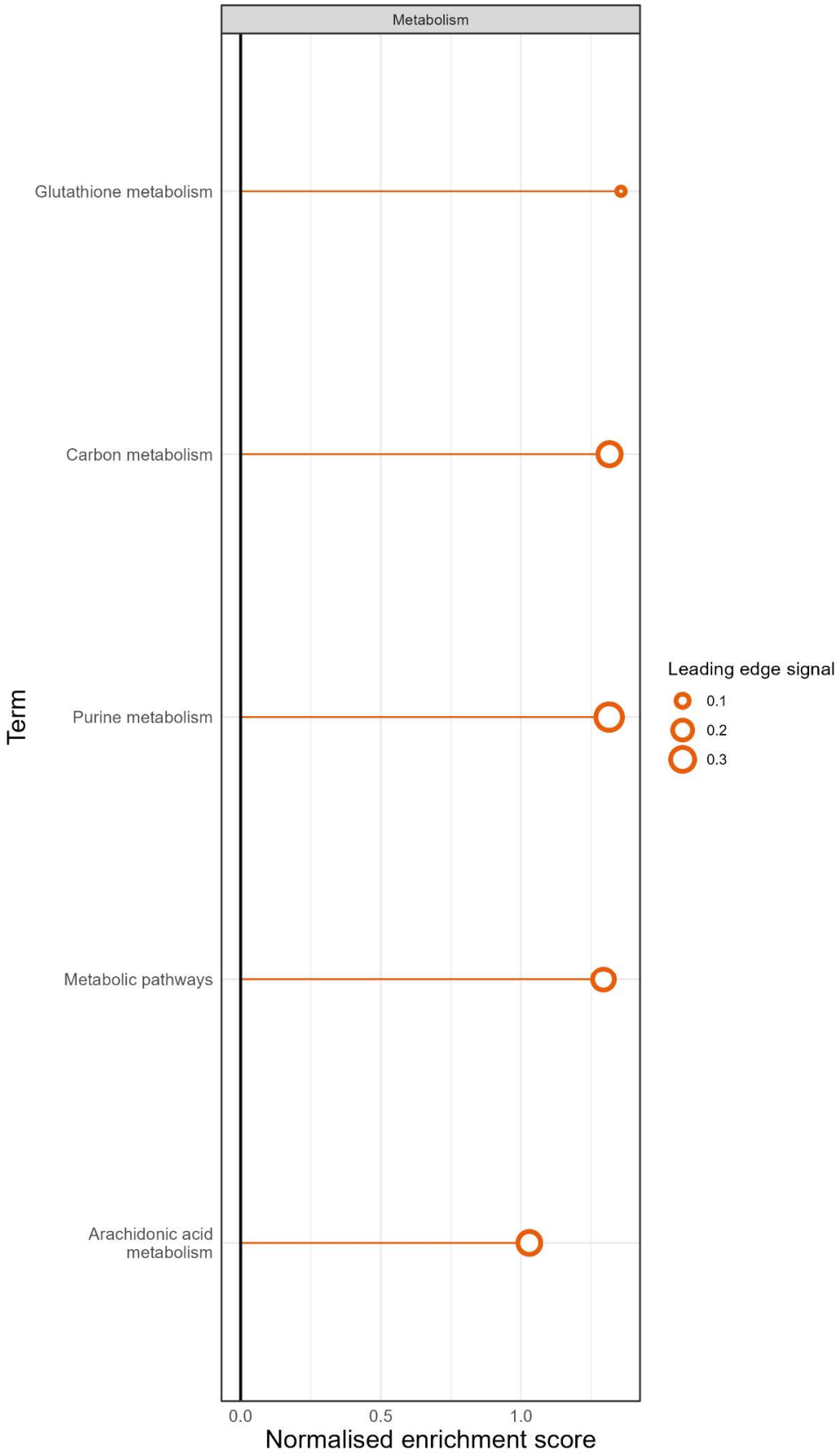
Supplementary Figure 64 ***Gene set enrichment analysis (GSEA) of TF abundance contributions to mRNA levels during HoxB8 neutrophil differentiation, using a linear regression approach.*** Genes were pre-ranked based on statistically significant coefficients corresponding to TF abundance variables used to predict mRNA levels. Gene ontology (GO) molecular function (MF) terms are shown.



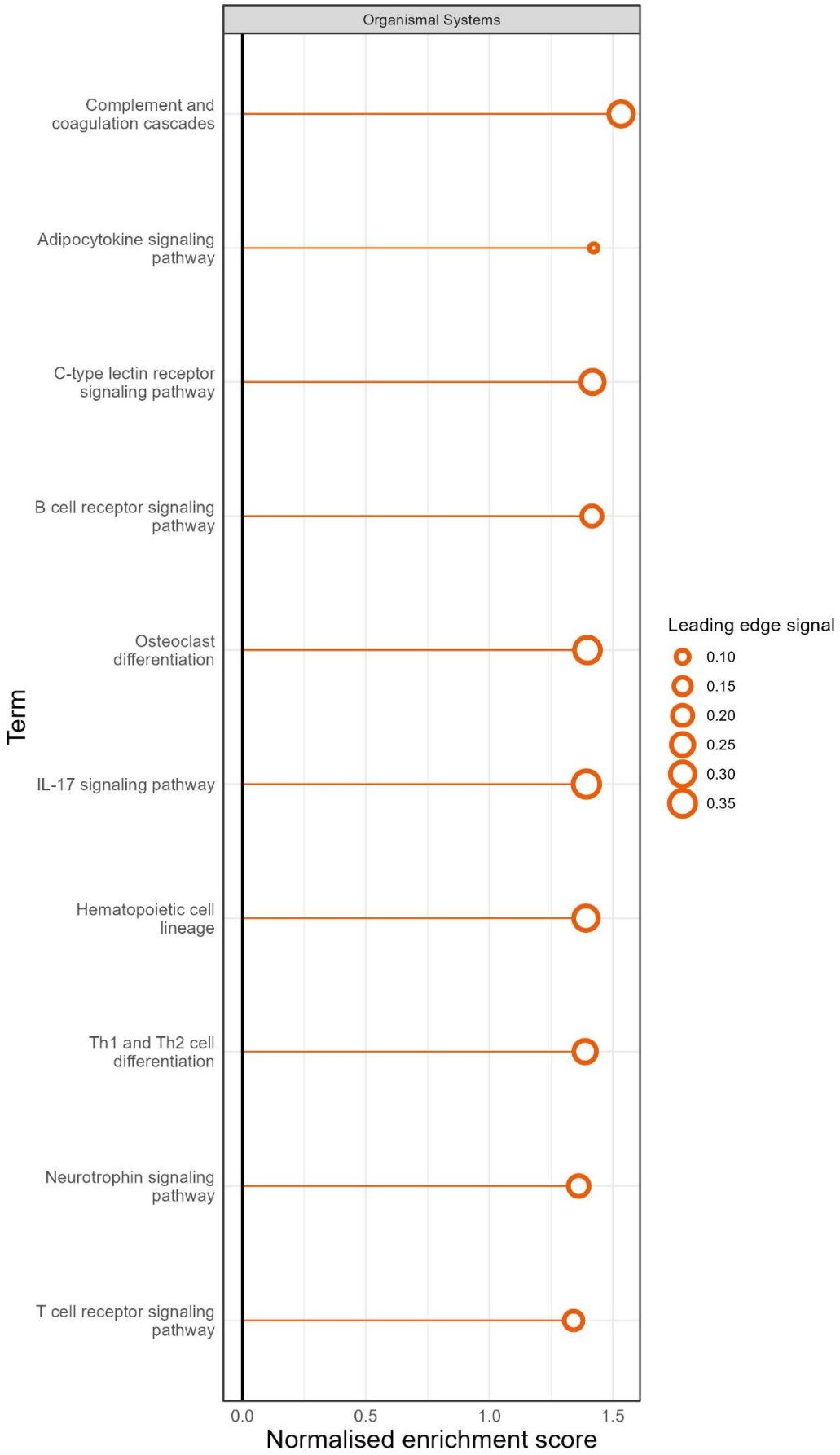
Supplementary Figure 65 ***Gene set enrichment analysis (GSEA) of TF abundance contributions to mRNA levels during HoxB8 neutrophil differentiation, using a linear regression approach.*** Genes were pre-ranked based on statistically significant coefficients corresponding to TF abundance variables used to predict mRNA levels. The KEGG pathway subcategory 'Cellular Processes' terms are shown.



Supplementary Figure 66 ***Gene set enrichment analysis (GSEA) of TF abundance contributions to mRNA levels during HoxB8 neutrophil differentiation, using a linear regression approach.*** Genes were pre-ranked based on statistically significant coefficients corresponding to TF abundance variables used to predict mRNA levels. The KEGG pathway subcategory 'Genetic Information Processing' terms are shown.



Supplementary Figure 67 ***Gene set enrichment analysis (GSEA) of TF abundance contributions to mRNA levels during HoxB8 neutrophil differentiation, using a linear regression approach.*** Genes were pre-ranked based on statistically significant coefficients corresponding to TF abundance variables used to predict mRNA levels. The KEGG pathway subcategory 'Metabolism' terms are shown.



Supplementary Figure 68 ***Gene set enrichment analysis (GSEA) of TF abundance contributions to mRNA levels during HoxB8 neutrophil differentiation, using a linear regression approach.*** Genes were pre-ranked based on statistically significant coefficients corresponding to TF abundance variables used to predict mRNA levels. The KEGG pathway subcategory 'Organismal Systems' terms are shown.