Optimization of phosphoproteomic workflows for challenging sample types and limited sample availability

Dissertation for promotion
to Doctor rerum naturalium
at the department of Biology of the
Johannes Gutenberg-University Mainz

Thomas Michna
born 06.12.1990 in Landau in der Pfalz

Mainz, 2023

First reviewer + first supervisor	Prof. Dr. Stefan Tenzer
Second reviewer	Prof. Dr. Eckhard Thines
Second supervisor	Dr. Stefan Jacob

Contents

Contents i

List of Figures iv

List of Tables xiii

Acknowledgements xvii

1 Introduction 1

1.1 Scope of this work . . . 2

1.1.1 Aims . . . 2

1.1.2 Rapid evolutionary events in the model organism *Magnaporthe oryzae* . . . 3

1.1.3 Phosphoproteomic profiling of human osteosarcoma cells . . . 5

1.1.4 Isolated Th17 cells from mice . . . 7

1.2 Techniques in bottom-up proteome analysis . . . 10

1.2.1 Introduction to the analysis of the proteome by LC-MS/MS . . . 10

1.2.2 Cell lysis and protein digest . . . 14

1.2.3 Phosphopeptide enrichment . . . 18

1.2.4 Peptide analysis by liquid chromatography and mass spectrometry . 21

1.2.5 Ion mobility spectrometry . . . 25

1.2.6 Visualization of mass spectrometry data . . . 27

1.2.7 Data dependent versus data independent acquisition mode . . . 28

1.3 Peptide identification, quantification and bioinformatics . . . 31

1.3.1 Processing of DDA data . . . 31

1.3.2 Processing of DIA data . . . 33

1.3.3 Quantification strategies . . . 35

1.3.4 Preprocessing, missing values and statistical testing . . . 36

1.3.5 Data visualization and result analysis . . . 42

2 Materials and methods **47**

2.1 Sample overview 48

2.2 Sample preparation 51

2.2.1 Cell lysis and protein digest 51

2.2.2 Phosphopeptide enrichment 53

2.2.3 ERLIC chromatography parameters 54

2.3 Peptide identification 55

2.3.1 LC-MS/MS of *M.oryzae* as resource for osmostress signaling research 55

2.3.2 LC-MS/MS of *M.oryzae* in DDA for comparison to DIA 56

2.3.3 LC-MS/MS of *M.oryzae* in DIA for comparison to DDA 56

2.3.4 LC-MS/MS of HOS / Th17 57

2.3.5 Data processing parameters 57

2.3.6 Availability of raw files and R code 59

3 Results and discussion **61**

3.1 Improved sample preparation and measurement 62

3.1.1 Optimized cell lysis 62

3.1.2 Advancements in phosphopeptide enrichment methods 66

3.1.3 Comparison of DDA vs. DIA approach for phosphopeptide identification 78

3.2 Rapid evolutionary events in *M.oryzae* 83

3.2.1 Proteome results 83

3.2.2 Proteomic response in wild type upon KCl stress 92

3.2.3 Phosphopeptide results 102

3.2.4 Phosphosignaling in wild type upon KCl stress 112

3.2.5 Temporal changes in wild type upon KCl stress 114

3.2.6 Altered protein and phosphopeptide response of wild type versus adapted Hog1 deletion mutants 119

3.2.7 Proteome and phosphoproteome analysis of not adapted Hog1 deletion mutants 121

3.3 Phosphoproteomic profiling of HOS cell culture and clinical samples 125

3.3.1 Statistical analysis of the phosphoproteome 125

3.3.2 Successful validation of Ceritinib effects from low amount phosphoproteomics . . . 131
3.3.3 Characerization of isobaric phosphopeptide separation by IMS . . . 132
3.4 Downscaling of mouse Th17 phosphoproteomics . . . 135
3.4.1 Statistical analysis of the phosphoproteome . . . 135
3.4.2 Overlap with standard phosphoproteomics workflows . . . 136

4 Summary **141**

5 References **145**

6 Supplementary data **169**

7 Curriculum vitae **187**

8 List of publications **191**

9 Statutory declaration **193**

List of Figures

1.1 Overview of different mass spectrometer types. A) Single quadrupol MS. B) Time-of-flight MS. C) Orbitrap MS. D) Triple quadrupol MS. E) Quadrupol coupled to time-of-flight MS (qTOF). F) Tandem ion mobility separation coupled to qTOF. Images adapted from [78, 79] 26

1.2 Operating principle of the Thermo Fisher Exploris 480 exemplary for Orbitrap Mass Analyzers as shown in [78]. The ionized molecules can be separated by the integrated quadrupol first, before fragmentation. After back-transfer of the ions from the collision cell, the ions enter an electric field inside the orbitrap, that causes a circular and oscillating movement of the ions along and around the inner core rod. By measuring the frequency of the voltage differences between two isolated outer shells caused by the oscillation of the ions, the m/z ratio can be determined after Fourier-Transformation and calibration. 26

1.3 Scheme of a trapped ion mobility spectrometer (tims) adapted from [82]. Ionized molecules enter from left and are pushed by gas flow into the tims device. There, a counter directed electric field traps ions at the position where the counter directed force by the electric field equals the forward directed force on the ions caused by the pushing gas flow. The molecule geometry in the gas phase determines the area that is affected by the gas flow. Thus, larger molecule geometry, *i.e.* cross collisional section (CCS), causes greater forces acting on the ions, the ion mobility is increased. In consequence, a gradient electric field along the tims device is applied to trap ions at different positions along the tube, separated by their ion mobility. 27

1.4 Examples of a typical proteomics peptide chromatography run with A) TIC of technically successful raw data B) spray instability during data acquisition . . 28

1.5 Example for a base peak chromatogram of a typical proteomics peptide chromatography. The most intense signal in each MS1 is displayed, in proteomics experiments these are typically peptides. 29

1.6 Principles of data acquisition. A) In data dependent acquisition (DDA) mode, single precursors are selected and fragmented for identification. This mode yields comparably clean spectra, but with limited capacity. B) In data independent acquisition (DIA) mode, predefined m/z ranges are fragmented simultaneously. Thus, the resulting MS spectra are more complex compared to DDA, but all peptides within this m/z range can possibly be detected. Figure adapted from [83] 30

1.7 Overview of possible data visualization strategies A) correlation plot B) volcano plot C) principal component analysis D) clustered heatmap E) Protein-protein interaction network F) Gene ontology clustering by Cytoscape/ClueGO 46

3.1 Overview of different lysis buffer strategies for A) mouse brain as control and B) *Magnaporthe oryzae*. The yield was calculated by measurement of protein content in the lysate by Pierce 660nm Assay compared to the crude sample weight that was used for lysis. While the least effecitve lysis strategy for the mouse brain samples still yields around 6 % protein amount of the used crude sample, the most effective lysis strategy for the fungal samples never exceeds 2 % protein amount of the sample weight. 64

3.2 A) Effects of heat treatment to unspecific protease activity (MO: *Magnaporthe oryzae* in green / SB: *Saccharomyces bayanus* in orange / MB: Mouse brain from *Mus musculus* in blue). Fungal and murine samples were lysed according to the named treatment, tryptically digested and peptides mesaured by LC-MS. The mass spectra were serched with unspecific cleavage allowed. A high number of unspecifically cleaved peptides indicates a high protease activity before sample preparation. Based on the observed results, heat treatment serves as effective measure to inhibit unspecific protease activity B) The effect of the ratio between sample weight and lysis volume on protein yield of *Magnaporthe oryzae*. Samples were weight with very variable amounts, due to the inhomogeneous nature of the grinded mycellium. A very weak correlation could be observed, with no clear conclusion possible. An arbitraty ratio of sample weight to lysis buffer volume of 1:4 has been used for further experiments. . . 65

3.3 Overview of phosphopeptide enrichment of indicated amount of mouse brain peptides performed with magnetic beads measured on an Orbitrap Exploris 480 in DDA mode. A) Performance of different functional material and starting amount compared to TiO_2. B) Titration of starting amount down to 25 µg (N=3) still yields satisfactory number phosphopeptide IDs 70
3.4 Parameters selected for initial phosphopeptide enrichment optimization 70
3.5 Peptide IDs of 25 µg mouse brain samples for the selected parameters measured on Orbitrap Exploris 480, except for experiment number 2 and 4, that were measured on the timsTOF Pro2. Both instruments were operated in DDA mode. Phosphopeptides in blue and unmodified peptides in red. 74
3.6 Optimization of ERLIC phosphopeptide selectivity of 2000 µg mouse brain peptides by alternative counterion and convex gradient. Peptide counts per collected fractions for unmodified peptide (red), singly phosphorylated (blue), double phosphorylated (yellow), triply phosphorylated (green) and four phosphosites per peptide (purple). While the original method published by Alpert in 2008 shows a high overlap of unmodified and phosphorylated peptides in the middle frations, using a Mg^{2+} counter ion and a convex gradient show a very low number of unmodified peptides was present while still containing a high number of phosphorylated peptides. 78
3.7 Performance comparison of three *M.oryzae* biological replicates, each 1000 µg, enriched for phosphopeptides measured in DDA (Orbitrap Exploris 480) and DIA (timsTOF Pro 2) regarding A) peptide counts B) overlap of identified phosphopeptides C) overlap withing DDA and DIA replicates D) precursor quantity reproducibility and E) phosphosite identification confidence 80
3.8 Protein identifications and pearson correlation of their quantitative abundance of each 1000 µg *M.oryzae* tryptic digest, measured on Orbitrap Exploris 480 in DIA mode. Each color represents one sample type of the following: irreversibly adapted (red), loss-of-function (green), reversibly adapted (blue) and wild type (purple). Each bar represents one measured sample of the biological quadruplicates side by side, increasing time points from left to right (0 min - control, 10 min, 60 min, 4 h, 24 h) 85

3.9 Measures of data quality from the data set presented in figure 3.8. A) Overlap of the four biological replicates from WT 1440 min time point B) data completeness and C) variability of TIC and sum of protein intensities (=value) 87
3.10 Evaluation of different normalization and missing value imputation strategies for protein level of the dataset presented in figure 3.8. All observed protein quantitative values were subjected to the stated normalization strategy and the resulting measured and imputed value are shown as histogram of the log transformed quantitative values. 89
3.11 Comparison of statistical testing for the wild type proteome results presented in figure 3.8. For this comparison, all replicate samples from time point 24 h versus control time point were subjected to statistical testing. A) Student's t-test B) limma and C) the overlap of significantly changing instances from each test 91
3.12 Multivariate statistical analysis of the dataset presented in figure 3.8. A) Principal component analysis B) heatmap for sample clustering as quality control of all wild type proteome results combined 93
3.13 Volcano plots of changing protein levels for the dataset presented in figure 3.8 as fold changes and q-values for each time point compared to the control of *M.oryzae* wild type samples upon osmotic stress. 94
3.14 Gene ontology analysis of all signifiantly changing proteins of *M.oryzae* of the dataset presented in figure 3.8 upon osmotic stress during 24 h. A) in ClueGO and B) from STRING-DB 95
3.15 Diagnostics for clustering according to the k-means clustering algorithm applied on observed proteome quantities of *M.oryzae* of the dataset presented in figure 3.8 upon osmotic stress in the time course of 24 h. A) Silhouette plot B) Withinness plot C) PCA with color code of the two suggested groups in red and turquoise 97

3.16 Diagnostics for clustering according to the dbscan clustering algorithm applied on observed proteome quantities of *M.oryzae* of the dataset presented in figure 3.8 upon osmotic stress in the time course of 24 h. A) Distance plot B) PCA with color code of the two suggested groups in red and turquoise. Black dots represent outlines, that were excluded by the algorithm and not assigned to one of the groups 98

3.17 Diagnostics for clustering according to the hierarchical clustering algorithm applied on observed proteome quantities of *M.oryzae* of the dataset presented in figure 3.8 upon osmotic stress in the time course of 24 h. 99

3.18 Empirical clustering by k-means with k = 2 / 5 / 10 / 15 applied on observed proteome quantities of *M.oryzae* of the dataset presented in figure 3.8 upon osmotic stress in the time course of 24 h. The commonly used threshold for signifcantly changing instances log2(fold change) = 1 and -1 are marked with blue horizontal lines for better identifcation of signifcant response clusters. . . 100

3.19 Phosphopeptide identifications and Pearson correlation of their quantitative abundance of *M.oryzae* of the sample set presented in figure 3.8. The number of phosphopepties are shown in blue, not modified peptides in red. Each color code at the bottom represents one sample type of the following: irreversibly adapted (red), loss-of-function (green), reversibly adapted (blue) and wild type (purple). Each bar represents one measured sample of the biological quadruplicates side by side, increasing time points from left to right (0 min - control, 10 min, 60 min, 4 h, 24 h) 104

3.20 Measures of data quality of the sample set presented in figure 3.19. A) Overlap of the four biological replicates from WT 1440 min time point B) data completeness and C) variability of TIC 105

3.21 The sum of all phosphopeptide quantification values of the sample set presented in figure 3.19 is calculated for each phenotype dataset. Different data preprocessing strategies, such as VSN and the combination of kNN imputation and VSN are able to reduce the intra- and inter-dataset variability. On the left, the results for all observed values are included (All), no the right only peptide that were present in two out of four replicates (Filtered). 107

3.22 Histogram of phosphopeptide intensity values within the wild type samples only of the sample set presented in figure **??**. Evaluation of VSN, SVD, kNN and voom normalization and missing value imputation strategies do not significantly change the distribution of the intensity values compared to the raw intensities. 108
3.23 Comparison of statistical testing for the wild type phosphopeptide results 24h versus control time point of the sample set presented in figure 3.19. A) Student's t-test B) limma and C) the overlap of significantly changing instances from each test 110
3.24 Multivariate statistical analysis of the sample set presented in figure 3.19. A) Principal component analysis B) heatmap for sample clustering as quality control of all wild type proteome results combined 111
3.25 Volcano plots of changing phosphopeptide levels of *M.oryzae* during 24 h after osmotic stress of the sample set presented in figure 3.19 as obtained fold changes and q-values for each time point compared to the control 113
3.26 Temporal change of the dual phosphosites pY, pT and pY+pT of MoHog1 in four biological replicates upon salt stress of the sample set presented in figure 3.19 confirms the immediate response of the dual phosphorylation pY+pT. The pY only phosphosite also shows an upregulation, but with lower fold-change as the dual phosphosite pY+pT. The phosphosite pT does not show any immediate response, but decreases with low fold change at alter time points of more than 240 min. 114
3.27 ClueGO network for all significantly changing phosphopeptide instances in wild type *M.oryzae* upon salt stress of the sample set presented in figure 3.19. Predominant processes include cell cycle, protein phosphorylation and signal transduction. 115
3.28 STRING DB GO term enrichment of KEGG pathways for all significantly changing phosphopeptide instances in wild type *M.oryzae* upon salt stress of the sample set presented in figure 3.19. Among the significantly enriched terms, MAPK signaling pathway is present, also including the HOG pathway. This finding validates our findings in general as they are in line with previously published and confirmed data. 116

3.29 Temporal changes of phosphopeptide intensities of wild type *M.oryzae* of the sample set presented in figure 3.19 clustered by k-means clustering algorithm with arbitrary chosen k-cluster, displayed in k groups as spaghetti plot. On the y-axis the log2(fold change) is shown, while the x-axis shows each of the four time points as vertical lines. The commonly used threshold for significantly changing instances log2(fold change) = 1 and -1 are marked with blue horizontal lines for better identification of significant response clusters. Immediate responding phosphosites can be found in clusters 4 and 9, putatively regulating phosphosites in cluster 5 and phosphosites with functions in maintaining homeostatsis can maybe found in cluster 2 and 3. 118

3.30 Three biological replicates of control group and Ceritinib treated cultured HOS cells. Phosphopeptides enriched from 25 µg trypsin digested protein. Each sample measured in triplicates on timsTOF Pro 2 in DIA mode. The number of identified phosphopeptides in blue and non phosphorylated peptides in red show at least 4500 identified phosphopeptides with a enrichment efficiency of around 50 %. The pearson correlation of phosphopeptide abundance is shown right, with good intra-sample reproducibility and no obvious outliner sample. . 126

3.31 Analysis of phosphopeptides from Ceritinib treated versus control HOS cell culture samples of the sample set presented in figure 3.30: A) Multivariate analysis by principal components. Control samples and treated samples cluster together, separated over a combination of PC1 and PC2, which together account for 98 % of the observed variances. B) STRING protein interaction analysis of the top 10 loadings of PC1 and PC2 including first and second shell interactors shows not only separated proteins, but also proteins with known interaction. This indicates a functional relationship of those caused by the treatement. 128

3.32 A) ClueGO analysis of all significantly changing phosphoproteins after Ceritinib treatment in HOS cell culture compared to not treated control samples of the sample set presented in figure 3.30. B) Kinase substrate enrichment analysis 130

3.33 IMS enables separation on co-elution isobaric phosphopeptides of the sample set presented in figure 3.30 A) Number of identified co-eluting and ion mobilty separated isobaric phosphopeptide isomers with B) example of a chromatogram and corresponding ion mobilogram 134
3.34 Three biological replicates of control group and treated cultured murine Th17 cells. Phosphopeptides enriched from 25 µg trypsin digested protein. Each sample measured in triplicates on timsTOF SCP in DIA mode. The number of identified phosphopeptides in blue and non phosphorylated peptides in red show at least 6500 identified phosphopeptides with a enrichment efficiency of around 50 %. The pearson correlation of phosphopeptide abundance is shown right, with good intra-sample reproducibility, except for sample 3 of the control group. 137
3.35 A) Results from statistical testing (t-test) as volcano plot from treated versus control murine Th17 cells of the sample set presented in figure 3.34. B) Multivariate analysis by principal components. Control samples and treated samples cluster together, separated over a of PC1, which accounts for 89 % of the observed variances 138
3.36 ClueGO network from significantly enriched GO terms of phosphopeptides significantly changing in murine Th17 cells after treatment of the sample set presented in figure 3.34 A) from 25 µg. B) from 1000 µg. 139

6.1 Example chromatograms of 25 µg mouse brain phosphopeptides after enrichment A) without desalting B) after SePak tC18 desalting C) after Oasis HLB desalting 170
6.2 Gene ontology enrichment networks for each cluster of temporal proteome response of wild type *M.oryzae* during the time course of 24h after osmotic stress. 171
6.3 Gene ontology enrichment networks for each cluster of temporal phosphopeptide response (1/2) of wild type *M.oryzae* during the time course of 24h after osmotic stress. 172
6.4 Gene ontology enrichment networks for each cluster of temporal phosphopeptide response (2/2) of wild type *M.oryzae* during the time course of 24h after osmotic stress. 173

6.5 Volcano plots of temporal changes in proteome of the irreversibly adapted Hog1 deletion mutant during the time course of 24h after osmotic stress. . . . 174
6.6 Gene ontology enrichment networks for each cluster of temporal proteome response in the adapted Hog1 deletion mutant during the time course of 24h after osmotic stress. 175
6.7 Volcano plots of temporal changes in phosphopeptides of the irreversibly adapted Hog1 deletion mutant during the time course of 24h after osmotic stress. . . . 176
6.8 Gene ontology enrichment networks for each cluster of temporal phosphopeptide response in the adapted Hog1 deletion mutant (1/2) during the time course of 24h after osmotic stress. 177
6.9 Gene ontology enrichment networks for each cluster of temporal phosphopeptide response in the adapted Hog1 deletion mutant (2/2) during the time course of 24h after osmotic stress. 178
6.10 Volcano plots of temporal changes in proteome of the Hog1 deletion mutant (not adapted) during the time course of 24h after osmotic stress. 179
6.11 GO enrichment of proteome changes of the Hog1 deletion mutant (not adapted) during the time course of 24h after osmotic stress. 180
6.12 Volcano plots of temporal changes in phosphoprotein of the Hog1 deletion mutant (not adapted) during the time course of 24h after osmotic stress. . . . 181
6.13 GO enrichment of phosphoprotein changes of the Hog1 deletion mutant (not adapted) during the time course of 24h after osmotic stress. 181
6.14 Volcano plots of temporal changes in proteome of the Hog1 deletion mutant (reversibly adapted) during the time course of 24h after osmotic stress. 182
6.15 GO enrichment of proteome changes of the Hog1 deletion mutant (reversibly adapted) during the time course of 24h after osmotic stress. 183
6.16 Volcano plots of temporal changes in phosphoprotein of the Hog1 deletion mutant (reversibly adapted) during the time course of 24h after osmotic stress. 184
6.17 GO enrichment of phosphoprotein changes of the Hog1 deletion mutant (reversibly adapted) during the time course of 24h after osmotic stress. 185
6.18 Normalized histograms of co-eluting isobaric phosphopeptide isomers enriched from 24 µg human osteosarcoma cells with co-elution (red) and separated (blue) in ion mobility. 186

List of Tables

1.1 Measurement modes of quadrupole mass analyzers. Typically, three different modes can be used: SRM, Scan and RF. Within a quadrupole, only one of the modes can be present. Depending on the application, multiple quadrupoles have to be combined to make use of using multiple measurement modes sequentially. 23

1.2 Typical search parameters for database search always include a type of proteolysis (or no proteolysis, if no enzyme was used for digestion), an acceptable number of missed cleavage, biochemical modifications to the peptides (fixed or variable) as well as a source of protein sequences (typically a FASTA database) 32

1.3 A selection of quantification algorithms. Although this is not an exhaustive list, Top N, MaxLFQ and iBAQ are the most commonly used aglorithms. All algorithms have in common to take into account a sub-population of the identified peptides per protein. They differ mainly in the strategy how to select the peptides, based on their intensity (Top N), peptides changing in comparable manner (MaxLFQ) or normalization per protein length (iBAQ). Depending on the analytical need, the selection of the appropriate quantification strategy is key. 36

1.4 A selection of strategies for reducing batch effects and increasing robustness of statistical testing. Depending on the analytical aim, more than one strategy can be applied. Generally, normalization can be applied already during the acquisition or *in-silico* during analysis 38

1.5 A selection of strategies for reducing the number of missing values. The cause for missing values is key to choose the appripriate strategy. Two types of missing values can be differentiated, missing at random (MAR) and missing not at random (MNAR). Replacing MNAR values is generally more robust, as a rationale for their missingness is usually present. E.g. missing values due to low intensity. In this case a minimum value approach might aid to gain the correct conclusions from the dataset. On the other hand, replacing MAR values require more sophisticated approaches. More information have to be taken into account to increase the propability for correct replacement of the missing values. Often, phosphoproteomics experiments suffer from MAR values, as the sample preparation procedure is very sensitive towards slight changes in the experimental conditions. 39

2.1 Proteomics samples *M.oryzae* 1/2. Each sample type was measured in quadruplicates at 5 time points in DIA mode. 48

2.2 Proteomics samples *M.oryzae* 2/2. Each sample type was measured in quadruplicates at 5 time points in DIA mode. 48

2.3 Phosphoproteomics samples *M.oryzae* 1/2. Each sample type was measured in quadruplicates at 5 time points in DIA mode. 49

2.4 Phosphoproteomics samples *M.oryzae* 2/2. Each sample type was measured in quadruplicates at 5 time points in DIA mode. 49

2.5 Phosphoproteomics samples human osteosarcoma cells. Control and Ceritinib treated cells were measured in triplicates in DIA mode. 50

2.6 Phosphoproteomics samples murine Th17 cells. Control and treated cells were measured in triplicates in DIA mode. 50

2.7 Gradient conditions for a convex gradient in ERLIC in contrast to the linear gradient used in previous publications such as [126] 55

3.1 Identified optimal values for phosphopeptide enrichment using Zr^{4+}-IMAC magnetic beads. SP3, yielding comparably clean peptide samples, served as optimal digestion protocol before phosphopeptide enrichment. Moderate amount of additive and a lowered peptide to beads ratio did show the optimal balance for low amount enrichment. Surprisingly, the phosphopeptide loss during desalting outweighted the positive effect during chromatography, thus no deslating after enrichment is advised. 75

3.3 Summary of top 10 absolute PC1 loadings in HOS cells treated with ceritinib. Proteins involved in TNFa sysnthesis, RNA processing and EGF receptor trafficking anre included in the list, indicating changes in relevant cancer related processes. 127

6.1 Variable window sizes for DIA acquisition with Orbitrap Exploris 480. Applied for the DIA data acquisition of the *M.oryzae* osmostress resource 169

Acknowledgements

This work would not have been possible without the help of many people. I would like to take the chance here to express my gratitude:

To my supervising professor Dr. Stefan Tenzer, who believed in my skills right from the beginning and gave me the confidence to finish my work.

To my colleagues from AG Tenzer and the Institute of Immunology of the University Medical Center in Mainz, especially Ute Dister, Sabine Arndt, Matteo Lacki, Dana Hein, Malte Sielaff, Ruben Sporer, Christina Jung, Christian Leps, Anna Gabele, Assel Nurbekova, Marian Scherer, Lucas Kleinort, Claudia Darmstadt and David Gomez Zepeda. It has been a pleasure to work with all of you, I am thankful for all the nice discussions, not only the work and project related ones, the parties and the many µ of coffe we shared together!

To all colleagues from the ibwf, especially Dr. Stefan Jacob, Katharina Bersching and Sri Bühring. Without their enthusiasm, scientific input and practical help many party of this work would not have happened.

To all collaborators, I shared projects with. I was always happy to help and aid to gain insight in their data from proteomics perspective, it has always been fun and interesting.

A very special thank goes to my familiy for helping me in every thinkable way through good and bad times. Especially I would like to thank my wife Janina, who took care of our son (and everything else around our lives) while I was writing my thesis. That has been a very tough time and I acknowledge all the sacrifices she gave to make this PhD thesis happen. Thank you, I love you!

1 Introduction

1.1 Scope of this work

1.1.1 Aims

By definition, research aims to expand the previous knowledge. A common approach is the statement of a hypothesis, repeated experimental cycles and finally a conclusion, that validates or rejects the hypothesis. Another approach is the development of and elaboration on enabling techniques. The question here is, how can we make our techniques better? More precise? More informative? This thesis combined both approaches: Three different scientific problems with each their own hypothesis were solved by optimizing the enabling technique. This way, the added value of this research is double: Advancing enabling techniques and expanding knowledge in biological questions. In the following sections each scientific question is presented separately. But although the used techniques vary, the analytical aim and technology is common in all three problems: Proteome and phosphopeptide analysis using liquid chromatography coupled to mass spectrometry (LC-MS/MS). This work aims to find solutions for two prominent problems in phosphoproteomics: challenging samples such as tough cell wall structures of *Magnaporthe oryzae* and low sample amounts from very slowly growing cells *in-vitro* and from clinical samples as well as minute and precious sample amounts from animal experiments.

The problem of challenging sample types should be addressed by optimizing the lysis procedure and applying a data acquisition strategy that has not been widely applied yet for phosphopeptide analysis, *i.e.* data independent acquisition (DIA). In addition to that, novel data analysis strategies for the identification of relevant statistically significantly changing phosphopeptides should be characterized, *i.e.* linear models and missing value imputation strategies. This way, an optimized pipeline for challenging sample types should be developed.

Minute sample amounts have been addressed in the past and sample preparation techniques for phosphopeptide analysis have been developed to decrease the amount of required starting material to approximately 200 µg. Still, some sample types will require either a long time period to generate that much protein, such as human osteosarcoma cells (HOS), or will require pooling of biological replicates, such as sorted immune cells extracted from mice. This way, the biological information gets convoluted and more bi-

ological replicates are required. Therefore, a phosphopeptide enrichment method based on magnetic beads should be developed to decrease the required amount to 25 µg. In combination with DIA, the resulting biological outcomes and novel technical aspects that have recently been introduced, such as trapped ion mobility separation (TIMS), should be evaluated.

These challenges were addressed in three different biological projects and the main focus lays on the technical progress for the field, rather the biological outcome. Nevertheless, the results will be made available online and serve as potential resource for future in-depth bioinformatic analysis to gain even deeper insight into the underlying biological processes in the three presented projects.

1.1.2 Rapid evolutionary events in the model organism *Magnaporthe oryzae*

The first main scientific question of this research deals with a fundamental question in biology: How can organisms effectively adapt to new environmental conditions? In this case, the filamentous rice plant pathogen *Magnaporthe oryzae* serves as model organism to study a rapid adaptation phenomenon.

M. oryzae is one of the biological threads for rice crop production worldwide. When infected, the rice plants develop the rice blast disease which then accounts for typically 10 to 30 % of crop loss and regional epidemics can be devastating [1]. This makes *M.oryzae* a major risk for food supply and economy, as rice serves as main source of food for approximately half of the world population [2]. In order to prevent crop loss and maintain food supply, extensive scientific research to understand the plant-pathogen interaction as well as the biochemistry of *M.oryzae* had been necessary during the last decades. The high importance for nutrition and economy, an early availability and possibility of manipulation of the *M. oryzae* genome as well as its pathogenicity make it suitable and desirable as model organism for research purposes [3].

The fungus shows an interesting and mechanistically not yet understood adaptation behavior: When generating genetically modified phenotypes of *M.oryzae* that lack the function for osmostress regulation, these mutants will not grow on highly osmolarity media,

e.g. high concentrations of salt or sugar. After at least eight weeks of cultivating the loss-of-function (lof) phenotype on high osmolarity medium, it reproducibly regains the ability to respond appropriately to osmostress and begins to grow again. Instead of accumulating arabitol as intramolecular compensation of the osmotic gradient between medium and cytosole, the adapted lof phenotypes accumulate glycerol [4]. Thus, the mechanism of response must have changed in a rapid time frame (on an evolutionary scale), without mutations in the genome [4].

The high osmolarity glycerol (HOG) pathway is responsible for the regulation and response to osmotic stress. This pathway has been extensively studied in yeast and many insights seem to be transferable to *M.oryzae* and other fungi [5]. The molecular mechanism of action involves environmental sensing by a hybrid histidine sensor kinase (HIK), which is autophosphorylated at a conserved histidine residue upon normal environmental conditions. The phosphoryl group is transferred to an aspartic acid residue within the same protein and then subsequently transported via a his-phosphorylated phosphotransfer protein to a asp-phosphorylated response regulator protein [6]. Osmostress sensing initiates the dephosphorylation of a conserved histidine amino acid residue of the HIK, which in turn activates a MAPK cascade. In consequence, the mitogen activated protein (MAP) Kinase Hog1 is phosphorylated and translocates into the nucleus, where it acts as transcriptional response regulator [7]. Interestingly, the cited study [4] also shows, that no matter which protein of the signaling cascade is defective, all mutants are able to regain osmoregulation ability.

The mechanism and cause for the ability to regain functions is yet unknown. The current hypothesis is, that signaling pathways are rewired so that other HIKs or other sensor proteins can take over the role of stress sensors while redirecting the response to the accumulation of a different intracellular solute. In contrast to the very well studied model organism *Saccharomyces cerevisiae*, where only Sln1 as HIK is known, the genome of *M.oryzae* contains 10 putative HIK coding sequences of which many maybe take over the role of MoSln1 as osmostress sensor [7]. The main questions addressed this study were:

1. What is the proteomic response in wild type and adapted mutant?
2. Which pathways are involved in each genotype and phenotype?

1.1.3 Phosphoproteomic profiling of human osteosarcoma cells

Osteosarcoma is the most frequently occurring form of malignant sarcoma in children and young adults between 5 and 20 years and is associated with very poor long term prognosis of 20% survivors [8]. The first line treatment is a combination of chemotherapy and surgery, but the recurrence intervals are comparably short and often metastasis in other soft tissue, in most cases the lung, become dominant [9]. Although the molecular mechanisms leading to aberrant cell proliferation and enhanced cell motility have been addressed by genetics and transcriptomics, options for potential therapeutic targets are limited [10]. Most commonly, the tumor suppressor gene TP53 shows genetic rearrangements leading to inactivation of p53 [11]. Furthermore, fusion genes of LRP1-SNRNP25 and KCNMB4-CCND3 were found to promote osteosarcoma cell motility [12]. Recent studies show the involvement of intracellular anaplastic lymphoma kinase (ALK) and the insuline-like growth factor 1 receptor (IGF1R) in sarcoma, regulating cellular growth, proliferation, and survival [13, 14]. Inhibition of these proteins have been proposed as promising pharmaceutical targets in cancer therapy [15, 16, 17]. This opens the possibility for therapeutic agents such as Ceritinib and Dasartinib. Ceritinib was originally FDA approved in 2014 for the treatment of non-small cell lung cancer (NSCLC) and was proposed to be an ALK inhibitor [18]. Nevertheless, it has been shown that Ceritinib also displays inhibitory effects on IGF1R [19]. *In-vitro* and *in-vivo* clinical observations suggest that inhibitory effect of the monotherapy with Ceritinib leads to a bypass Src activity counteracting the IGF1R inhibition effect [20]. As higher Src activity is associated with cancer progression [21], a combination treatment with Src inhibitory agents, such as Dasartinib, have proven very effective in treating *in-vitro* cell culture of primary tumor cells and experimental therapy for very few patients [22].

One of the patients was a 16 years old girl, that already received all conventional therapy options and agreed to participate in this experimental study where she was treated with the combination of Ceritinib and Dasartinib. Her disease progression was closely monitored and biopsy samples were taken from the lung metastasis before treatment and after treatment, to monitor intra-tissue drug concentration. On this occasion, primary tumor cells of the biopsy samples were taken into cell culture to grow enough cells to perform phosphoproteomic analysis for the elucidation of the current (aberrant) cellular

signaling status. At this point of time, roughly 1000 µg protein material was required to perform a successful phosphoproteomics experiment using TiO_2 spin tips. As the primary lung metastasis cells were growing at a very slow rate, it took several weeks to meet the necessary amount of protein.

Therefore, a method was *urgently* needed for robust and comprehensive phosphoproteomic analysis, that requires much less amount of input material, in order to overcome issues with unnecessary delay times for clinical samples in the future. Nevertheless, phosphoproteomics in general and especially with low amount of starting material is still very challenging [23]. Recent publications showed promising results in terms of downscaling of phosphoproteomics experiments [24, 25]. In the cited studies, the authors have already achieved the identification of 3000 to 4000 phosphopeptides from as little as 25 µg protein material, which is 40-fold less than required by most common phosphoproteomics workflows [26]. In addition to that, recently developed Zirconium based immobilized metal ion affinity chromatography (Zr-IMAC) magnetic beads promise to increase the quality of the analysis while being excellently scalable at the same time [27]. This served as starting point for the development of a downscaled phosphoproteomics workflow in our laboratory. For this, a low cell number of cell culture samples of a commercially available human osteosarcoma (HOS) cell line were treated with Ceritinib and 25 µg of the resulting protein amount after cell lysis was used for phosphopeptide enrichment using a newly developed Zr-IMAC method. This way we could benchmark the new phosphopeptide enrichment workflow with the literature values and validate the biological findings, as we expected to identify the effects of ALK / IGF1R inhibition on PTM level.

In addition to the validation of pinpointing the expected biological processes despite massive downscaling, we demonstrate with this dataset a novel technical approach for the elucidation of coeluting and isobaric phosphoisomer pairs. Roughly 20 % of all detected phosphopeptides share the same amino acid sequence with one or more phosphopeptides, that differs in the identified position of the phosphosites. By nature, they have the same molecular weight and thus display the same m/z values in mass spectrometry, they are isobaric. Nearly half of those isobaric positional isomers can not be resolved by chromatography, they elute at the very same time from the analytical column. Thus, their m/z signal intensities are convoluted and also fragments supporting both positional iso-

forms are present in the resulting MS2 spectra. In consequence, the site localization confidence is decreased and a proper quantification and identification is impossible. Additional measures have to be utilized to properly resolve coleluting isomer peptides and in the past, ion mobility spectrometry (IMS) had been demonstrated as possible solution for this separation problem [28]. The recently introduced Bruker timsTOF Pro 2 offers an integrated IMS solution for molecule separation based on their ion mobility after elution from the analytical column and ionization, before analysis in the mass spectrometer. Thus, this instrument has the potential to increase the identification and site localization confidence and make the separate quantification of both ionspecies even possible. Here, we examine the dataset for such cases and evaluate the outcome in comparison to traditional LC-MS/MS without the possibility for ion mobility separation, i.e. the Thermo Fisher Orbitrap Exploris 480.

In summary, the adressed questions in this study were:

1. Can we achieve a competitive identification rate compared to the recently published numbers?
2. Can we validate the expected biological responses of the results?
3. Is it possible to resolve coeluting and isobaric phosphopeptide isomers?
4. Does the confidence in identification and site localization increase due to the use of IMS?

1.1.4 Isolated Th17 cells from mice

For the successful defense of invading pathogens and disease prevention in any living organism, a complex interplay of specialized functions is required. In the human body, a multitude of different specialized cells take over those functions and serve together as the immune system. One part of the immune system is already present at birth, the innate immune system. The whole innate immune system serves as first line defense against pathogens and is unspecific. In order to increase specificity, a second system exists that develops during lifetime and adapts towards more specific targets - the adaptive immune system.

The adative immune system orchestrates a plethora of specialized cells, such as T cells, B cells and Antigen presenting cells. Usually, an invading pathogen is processed by antigen presenting cells, where its proteins undergo proteolysis into peptides (*i.e.* antigens) that are presented at the surface of these cells. T cells are able to recognize the antigen as non-endogenous and start secreting signaling molecules (*e.g.* cytokines such as interleukins and interferons) to attract and/or differentiate additional T cells, in order to start the pathogen defense. Therefore, the naïve T cells differentiate into more polarized T cell subtypes, such as T helper cells (T_{H1}, T_{H2}, T_{H17} and many more), cytotoxic T cells (T_C) or regulatory T cells (T_{reg}). T cells can have pro-inflammatory effects, which are necessary during infection to neutralize the pathogen, or anti-inflammatory effects, to regulate the inflammatory state. The mechanisms to regulate pro- and anti-inflammatory processes are well balanced and highly complex. A perturbation of this sensitive system has devastating consequences.

Roughly up to 8 % of the world population suffers from some kind of such perturbations of the immune system [29]. *Id est* that the acquired immune system recognizes endogenous cells, tissues or molecules as hostile pathogens and consequently maintain an inflammatory state at the affected areas. In severe cases, multiple organs can be affected simultaneously and are impaired in their function, which can be life threatening and hospitalization of the patients is often required. Anti-inflammatory or immunosuppressive medication is needed to control the patients immune system and prevent auto-immune reactions. In consequence, typically mild infections can be a major threat to the patients health and triggers the need for more specialized immunoregulatory therapy to maintain the patients resistance towards pathogens while minimizing auto-inflammatory reactions.

In the presented study, a Casein Kinase II (CKII) inhibitory agent is tested on isolated naïve T cells from mice, that are treated with differentiating agents. The hypothesis here is, that the inhibition of CKII leads to differentiation to a T cell with anti-inflammatory phenotype or regulatory T cell. To study the consequences and draw conclusions about the phenotype of the resulting T cell, phosphoproteomic analysis is required. Typically 1000 µg of peptides are necessary for a successful phosphopeptide enrichment using TiO_2 spin tips. To obtain sufficient amounts of protein, differentiated T cells from multiple mice have to be pooled, which causes an increased demand of animal resources and information

convolution of the mouse specific phenotype. Recently developed enrichment strategies allow the use of up to 40-fold less peptide consumption and still obtain a high number of phosphopeptides. Nevertheless, the more peptide is available for the enrichment, the more phosphopeptides will be identified. Most importantly, the biological conclusions from the reduced number of phosphopeptides should be similar. For the validation of this approach, a comparison of both phosphopeptide enrichment methods has been done and the following questions have been addressed:

1. Can we achieve a competitive identification number
   comparing 25 µg and 1000 µg starting material?
2. Is the nature of the identified phosphopeptide dependent on the enrichment type?
3. Is the biological conclusion the same for both enrichment types?

1.2 Techniques in bottom-up proteome analysis

1.2.1 Introduction to the analysis of the proteome by LC-MS/MS

Parts of this chapter have been published in [26]

For a comprehensive understanding of complex biological processes, it is necessary to link information of multiple levels, such as transcriptome, proteome, metabolome and Post-Translations-Modifications (PTMs) [30]. Especially the analysis of protein phosphorylation is key to understand cellular signaling [31]. Nowadays, LC-MS/MS approaches allow the identification and quantification of thousands of peptides in a single analysis, but until a decade ago, the science of proteomics was very tedious and rather insensitive: For protein identification it was necessary to prepare two-dimensional gels (2D-GE) that were difficult to handle, time consuming, low resolution and difficult to reproduce. Identification and quantification of single proteins could be done using more or less specific antibodies or other techniques that came either with safety issues or error prone and thus far from robust [32]. The formerly relatively expensive mass-spectrometric (MS)-based identification suffered from high instrument cycle times and was only operable by highly specialized staff. Nevertheless, during the last years the instrument prices went down and the operability was simplified, making MS-based proteomics the method of choice for global, comprehensive protein analysis. Current generation instruments reproducibly quantify thousands of proteins with high sensitivity, throughput and robustness, rendering them superior to classical 2D gel approaches in most aspects [33].

The protein analysis by MS became more and more popular, but the basic principle of the strategy remained the same: Proteins undergo digestion by proteases like trypsin and Lys-C yielding smaller peptides. Following separation by reversed-phase liquid chromatography (RP-LC) and electrospray ionization (ESI), the peptide mixtures are analyzed online by the mass spectrometer. In nanoscale ultra-high-performance liquid chromatography (nanoUPLC) systems, peptides are separated by gradients (typically between 30 and 180 min length) and elute over a short time of 5–30 s into the MS. In data-dependent acquisition (DDA), the mass over charge ratios (m/z) and intensities of the eluting peptides are measured first. Subsequently, the most intense signals are selected for fragmenta-

tion in the collision cell. Bioinformatic tools allow to reconstruct the underlying peptide sequences from the obtained intact and fragment m/z values. Finally, proteins can be identified by mapping the peptide sequence to entries from protein databases. In label-free quantification, intensity information or spectral counts of the peptides are used for peptide and protein quantification [34, 35].

With transcriptional and translational changes of protein expression many functions of living organisms can be steered, but this process is time and resource intensive. Another level of information adds up by the use of PTMs, which enables rapid and resource saving regulation of cellular processes. While the generation of a new protein would take minutes to hours to react to a certain stimulus, a phosphorylation event can be rapidly catalyzed by specialized proteins (*i.e.* kinases) and is reversible by other enzymes (*i.e.* phosphatases). Structural changes of the substrate can lead to activation, deactivation or aggregation and stabilization of proteins, that serve the cells as response [36]. The phosphorylation event is an esterification with phosphorous acid or phosphate (typically in form of adenosine-triphosphate) with a hydroxy-, amide-group or even thiol-groups of the amino acids serine, threonine and tyrosine (S/T/Y) or arginine, lysine, aspartic acid, glutamic acid and cysteine (R/K/D/E/C). Phosphorylation on serine, threonine and tyrosine are the most abundant and studied phosphorylated sites in eukaryotes while phosphorylated histidines have traditionally rather been recognized in prokaryotes and plants, but recent research has proven that phosphorylated histidines are equally common in eukaryotes [37, 38].

The analysis of phosphorylated peptides is essential for the elucidation of signal transduction pathways, but remains challenging. Due to their low abundance in the peptidome, phosphopeptides will be difficult to detect in the presence of the signals that derive from non-phosphorylated species. Additionally, during electrospray ionization, phosphopeptides ionize less efficiently, which causes low efficiency in simultaneous identification of non- and phosphorylated peptides. Last but not least, the neutral loss of the phospho-group during peptide fragmentation process makes the correct identification of the phosphorylation site difficult [39, 23].

Thus, for the comprehensive analysis of the phosphoproteome, additional sample preparation steps for phosphopeptide enrichment are necessary. These include immunoprecipi-

tation (IP), metal oxide/immobilized metal ion affinity chromatography (MOAC/IMAC), and fractionation strategies such as high-pH reversed-phase chromatography (high pH RP), strong cation exchange (SCX), or electrostatic repulsion hydrophilic interaction chromatography (ERLIC) [40, 41].

Other methods than LC-MS/MS for the elucidation of cellular signaling through protein phosphorylation are available, such as western blot or kinase arrays. Those techniques are biased and require previous knowledge of the underlying mechanisms. Which antibody and which phosphosite have to be addressed in the western blot? How quantitative is the staining? Such typical questions can be solved using an unbiased and discovery like approach: bottom-up phosphoproteomics by LC-MS/MS. With this analytical strategy, not only very specific and targeted questions can be addressed, but also previously unknown phosphosites might be discovered, that might have been missed with western blot analysis or other analytical strategies.

The general approach in bottom-up proteomics by LC-MS/MS requires the proteolytic digest of the proteome into smaller peptides, where their mass and the mass of the amino acid fragments can be accurately measured by high accuracy and high resolution mass spectrometers. Bioinformatic search engines analyze the resulting spectra, compare to known databases and allow the identification and quantification of the proteins in shotgun style [42]. Although recent initiatives promote the analysis of intact proteins (top-down strategy), the resulting spectra are unequally complex and are - up to this date - extraordinary challenging to interprete [43, 44]. Especially when the research aims to identify specific sites and forms of post-translational modifications, such as protein phosphorylation, bottom-up peptide analysis serves as more applicable approach. However, the analysis of phosphorylated peptides remains challenging, as the number of phosphorylated peptides as subpopulation within a digested whole proteome samples is small compared to the number of unphosphorylated peptides. In addition to that, their physicochemical properties prevent efficient ionization and fragmentation in the LC-MS/MS analysis [45].

Furthermore, phosphopeptide analysis was traditionally very sample consuming, but recent developments allow the analysis with much less peptide material. This opens the door to multiple sample preparation strategies, such as filter aided sample preparation (FASP) or single pot solid phase sample preparation (SP3) that require much less peptide

material and yield peptides with high purity. In addition to that, alternative enrichment strategies - each with their own benefits - can be applied depending on the analytical need. Serine and Threonine phosphopeptide analysis for example can be done using TiO_2 spin tips, while the analysis of Tyrosine phosphosites requires Immunoprecipitation (IP). Next, the phosphopeptide separation parameters and mass spectrometry data acquisition strategy heavily influence the result, but are also demanding by chromatography and computational means.

In summary, a plethora of methods, parameters and challenges have to be taken into account for the successful analysis of phosphorylated peptides, which will be introduced in the following chapters.

1.2.2 Cell lysis and protein digest

For proteome analysis by LC-MS/MS, the proteins have to be isolated from their biological surrounding, being tissue, a cell or any other (bio)fluid / solid. Typically, aqueous buffers in combination with chemical and/or physical treatment are used to facilitate the cell / tissue lysis and protein release. In order to enable an efficient protein digest, all disulfide bonds of the cysteine residues within/between the proteins have to be reduced and re-stabilization by oxidation has to be prevented by chemical modification. Subsequently, the denatured and modified proteins are subjected to proteolytic digest using a protease.

A widely applied approach is the addition of aqueous solutions of chaotropic agents to denature the proteins and keep them in solution. Urea and Thiourea are commonly used as chaotropic agents, where Urea is known to denature the proteins preferrably by intercalating into the hydrophobic parts of the proteins, thus interrupt tertiary and secondary structures [46]. Although heat is also known to promote protein denaturation, high temperature is unfavourable in lysis conditions involving the chaotropic agent Urea as it will lead to partial carbamylation of the proteins on their Lysine (K) and Arginine (R) residues. Ultimately, this prevents the proteolysis with the protease trypsin, that relies on accesible and unmodified K and R residues for an efficient cleavage. If elevated temperature is desired or not avoidable during sample preparation, a possible solution to this issue is the use of guanidine hydrochloride as chaotropic agent, which comes with the downside of less effective protein denaturation and solubilization [47, 48]. In general, less stable protein/peptide modifications, such as phosphorylation on histidine and arginine residues, will be lost during treatment with elevated temperature due to the higher energy intake and the temperature dependency of chemical reactions according to the Van't Hoff equation [49].

The second important chemical lysis strategy includes the use of detergents like sodium dodecylsulfate (SDS), sodium deoxycholate (NaDOC) or the zwitterionic CHAPS (a taurin derivate) are used solely or in combination with chaotropic agents to increase the denaturation of proteins and promote their solubilization. Detergents typically consist of two distinct structural parts, that vary by a certain degree in their hydrophobicity. The hydrophobic part interrupts intra-protein hydrophobic interactions, usually in the inner core of the protein in case of cytosolic/secreted proteins or the transmembrane domains of

membrane bound proteins. Thus, the proteins are unfolded and the hydrophobic sections are covered by the hydrophobic part of the detergent. The hydrophilic part of the detergent then facilitates the solubilization in aqueous lysis buffers [50]. In addition to that, detergents are able to penetrate and solubilize the lipid layer of the cell wall and thus undermining the membrane integrity. This aids the thorough lysis of the cell, increasing the protein release from the cells and increases the solubilization of membrane bound proteins [51]. When combining the detergent based lysis buffer with reducing agents like dithiothreitol (DTT) and heating up to 95 °C the lysis efficiency can be significantly increased as demonstrated in later chapters. Although the solubilization efficiency of boiling SDS in combination with DTT serve as excellent denaturating and solubilizing conditions, the detergent is incompatible with the downstream analysis. When SDS is not removed from the peptides before LC-MS/MS analysis, SDS acts as ion-pairing reagent [52]. This alters the retention mechanism of the stationary phase, converting the chromatographic mode to a cation exchange condition. Consequently, the analytes can not be sufficiently separated for the subsequent MS analysis. This effect is irreversible, due to the strong affinity of the hydrophobic part of the detergent to the stationary phase and a new analytical column is required for furher analysis, which is costly and inefficient. Furthermore, detergents often cause ion suppression that reduces the ionization efficiency of the analytes and thus decreases the sensitivity to a great extent[53]. Contamination of the analytes with detergents are critical and have to be avoided, so additional sample preparation steps have to be implemented to separate the detergent from the analytes.

Suitable strategies to separate the detergent and the analyte are a) precipitation of the proteins, b) precipitation of the detergent and c) solid-phase extraction (SPE) in combination with molecular weight cut-off (MWCO) filters. The precipitation of the proteins is one of the oldest and simplest and thus most commonly used techniques for protein clean up. As many (cytosolic) proteins display a minute solubility in organic solvents, such is added to the lysis buffer until the maximum solubility of the proteins is exceeded. The proteins thus precipitate while the detergent remains in solution, which can easily be removed. Commonly used are cold acetonitrile (ACN) with over night incubation in the freezer or methanol/chloroform ($MeOH/CHCl_3$) mixtures for efficient precipitation. If NaDOC is used as detergent, it can be precipitated by lowering the pH. Last but not least, either commercially available SPE products like S-trap or alternative digestion

strategies like filter-aided sample preparation can be used for detergent separation prior to LC-MS/MS analysis [54].

Especially for challenging sample types, the combination of chemical treatment with physical treatment has proven very effective. Depending on the analytical aim, different types of forces can be applied for the disruption of cellular components, thus releasing more proteins and/or increasing the robustness of the workflow. For instance, the French pressure cell can be used for the disruption of biological membranes from suspension cells, that are more or less homogeneous. In order to homogenize tissues and disrupt the cells in one step, bead beating devices served best. An alternative strategy is the use of ultrasound, introducing shearing forces through pressure waves and strong focused forces via cavitation [55]. Especially ultrasound has proven beneficial for phosphopeptide analysis, because typically interfering substances in the enrichment step such as chromatin and DNA is sheared during this process and thus less likely to interfere with the phosphopeptide enrichment and analysis. The downside of using physical treatment as part of the sample preparation process is the possibility of heat dissipation into the lysis buffer. If heating of the lysis buffer can not be prevented, *e.g.* with active cooling, aforementioned issues and solutions have to be considered for lysis buffer design.

Following the successful release and denaturation of the proteins, their status has to be stabilized for the proteolytic digest. The first step is the reduction of existing intra- and intermolecular disulfide bonds between cysteine (C) residues of the proteins. By cleaving the disulfide bonds, larger areas become accessible for the protease and increase the digestion efficiency and protein coverage. Commonly used agents for the reduction are Mercaptoethanol and DTT, which typically are incubated with the proteins of interest for a period of 30 to 60 min under elevated temperature like 30 °C to 60 °C. More recent approaches use Tris(2-carboxyethyl)phosphine (TCEP) that offers a faster reaction at room temperature conditions [56]. As the reduced cysteins tend to re-oxidize to disulfide bonds, those have to be chemically modified to prevent this reaction. Typically, an alkylating agent like iodoacetamide (IAA) or chloracetamide (CAA) are used to introduce carbamidomethylation to the sulfur group of the cysteines as fixed artificial modification to the protein / peptide [56].

The proteolytic digest is performed by a protease, in most cases trypsin. Trypsin has the

advantage to cleave the amino acid sequences after each appearing lysine and arginine, unless it is followed by a proline (which sterically hinders the cleavage [57]). This creates peptides of managable length for the MS analysis and additionally ensures that at least one basic amino acid residue (*i.e.* K and R) is present in the resulting peptide, which increases the probability for efficient ionization during electrospray-ionization before MS analysis significantly. Typically, the digest is very efficient and does not yield many cases of missed cleavages. An increased number of missed cleavage serves an indicator for inefficient sample preparation beforehand (denaturation or reduction/alkylation), as those possible cleavage sites are believed to be less accessible for cleavage. It has been shown, that the presence of organic solvents increase the efficiency of the proteolytic digest [58] and meanwhile specialized products are commercially available that minimize the autolysis products [59] and thus creating less peptide artifacts with non-sample origin in the LC-MS/MS analysis.

1.2.3 Phosphopeptide enrichment

When analysing the resulting peptides from the proteolytic digest directly using LC-MS/MS, very few phoshopeptides can be found. The reason for this finding is mainly the low ionization efficiency of phosphorylated peptides in general, especially in presence of non-phosphorylated peptide species [23]. Due to the acidic nature of the additional phosphate group, the likelihood for attracting and stably harbouring additional positive charges in positive ionization mode is decreased compared to unmodified peptides. A very simple solution would be to switch the polarity to negative mode, but in general the ionization efficiency is even less and has proven to be not useful for phosphopeptide analysis [60]. In addition to that, the relative abundance of phosphorylated peptide species is significantly lower than unmodified peptides. Thus, their low signal intensity and also lower number in relation to signals from unmodified species make the effective selection for MS/MS identification in the mass spectrometer very challenging. Therefore, a different solution is commonly applied: the separation (*i.e.* chromatography) or enrichment of phosphopeptides from unmodified peptides before the analysis with LC-MS/MS. The physicochemical properties that can be used with current techniques for this chromatographic procedure are either the acidic nature / negative charge of the phosphorylated amino acid or the 3-dimensional structure.

It has been shown that phosphorylated peptides selectively bind to metal ions by establishing a bidentate bridging between two oxides of the phosphate group and the metal ion [61]. When utilizing this affinity mechanism, the pH of the chromatographic condition should be adjusted so that the phosphate group harbours a negative charge. The isoelectric point (pI) of the phosphopeptides is dependent on the neighbouring amino acid composition and is usually around 2.5 to 3.5, so an optimal pH value for an efficient chromatography should be buffered one pH unit higher than the phosphopeptide pI values to preserve the deprotonated state. The negative charge and polarity of the phosphate group shows naturally a high affinity to positively charged or uncharged polar ion species and molecules. The functional material can be bound to chromatographic resin filled in HPLC columns, Spin-tips or simply centrifugable slurry or magnetic beads. Either way, phosphorylated peptides can bind to the functional material while the unmodified species can be washed away (flow through). As functional material, metal ions such as Fe^{3+},

Ti^{4+} or Zr^{4+} are immobilized by chelation within a nitrilotriacetic acid (NTA) matrix (immobilized metal ion affinity chromatography - IMAC) or metal oxides such as TiO_2 or ZrO_2 as chromatograpic resin (metal oxide affinity chromatography - MOAC) are used for the affinity chromatography of phosphopeptides [62]. All of the proposed materials show high affinity towards phosphopeptides in general, but the differ in their exact specificity. Therefore, a sequential combination of different enrichment techniques, *e.g.* first enrichment using TiO_2 MOAC followed by a second enrichment using the flow-through with Zr^{4+} IMAC, have proven to yield a more comprehensive picture of the phosphoproteome [63, 64].

Naturally, not only phosphorylated peptides show affinity towards this functional material, but any other molecule harboring certain groups that introduce acidity. This includes peptides containing acidic amino acid residues such as aspartic or glutamic acid, but also DNA strands that are build around a heavily phosphorylated backbone. Two strategies are applied to prevent unspecific binding towards the metal oxides. First, competitive agents like small organic acids (citric, lactic, oxalic or dihydroxybenzoic acid and others) will reduce the binding of acidic peptides while leaving the binding of phosphopeptides unaffected. The reason for this is a minor difference in the mode of binding towards the metal ion between organic acids and phosphogroups. While organic acids tend to form bidentate chelates, the phosphogroup forms bidentate bridges [65]. Also, the 3-dimensional structure of the complexes are slightly different (*e.g.* because of the different angles of the binding dentates towards the metal) and organic acids mimic the structure for acidic amino acid residues in greater conformance. Second, to aviod the coverage of binding sites with contaminants such as DNA, they have to be removed before the phosphopeptide enrichment. Usually, digesting enzymes such as DNAse I or nuclease A (commercialized as benzonase) can be incubated with the sample before tryptic digest of the sample, but in this case the used nuclease concentration has to be monitored and optimized, as this will create artifact peptides with non sample origin in the analysis.

The aforementioned methods are well established, widely used and comparably easy to implement while requiring a high, but reasonable amount of protein as starting material. On the other hand, this analysis will lead to the identification of mostly serine, followed by threonine and very few tyrosine phosphosites, due to the naturally occurring

abundances and to some extent also the enrichment bias of the selected method [66]. In many research areas, such as cancer signaling, tyrosine kinases play a crucial role and the information about the affected phosphosites will not be reflected by MOAC or IMAC approaches. A solution to this problem is the use of phosphotyrosine targeting antibody based immunoprecipitation (IP) enrichment. In this case, the relatively large and feature rich 3-dimensional structure of the phosphorylated phenolic side chain serves as suitable target for specific antibodies [67]. In this approach, the pY-harbouring phosphopeptides are precipitated and eluted separately while the flow through can be used for sequential enrichment with MOAC or IMAC. Typically, for such IP pull down experiments as much as 10 mg of peptides are required to obtain a resonable number of tyrosine-phosphosites. Often this is a major limitation for clinical studies, where *e.g.* needle biopsies deliver only few milligrams of samples, of which most has to undergo traditional differential diagnostic procedures. An other considerable downside of currently widespread MOAC/IMAC approaches is the incompatibility to preserve fragile phosphorylation related PTMs such as phosphorylations on histidines, arginines and lysines, aspartic and glutamic acid and even cystein. Recent research could achieve a proteolytic digest and phosphopeptide enrichment with very mild conditions, that preserve fragile phosphosites, demonstrated on phospho-histidines of *Escherichia coli* [68]. Specifically in *M.oryzae*, a two component system harboring phospho-histidines is known to regulate the functions studied in one of the featured research questions, which can not be addresses with the used analytical methods.

For the successful and comprehensive elucidation of the phosphoproteome, many factors have to be considered. Similar to the proteolytic digest, one single method is not sufficient and has to be adapted to specifically answer the scientific question of interest. Ultimately, the method of choice is also mainly driven by the availability of protein and consequently peptide amount. Therefore, downscaling of the required material while maintaining information depth is one of the major challenges in the field of phosphoproteomics in the near future, especially in respect to single cell and spatial proteomics. Consequently, this issue was also addressed in this featured research and validated with a biological question, where we were able to obtain the same biological deduction while reducing the required starting material 40-fold.

1.2.4 Peptide analysis by liquid chromatography and mass spectrometry

Proteolytic digests of whole proteome samples are highly complex and have to be simplified before analysis which can be achieved *e.g.* by electrophoresis or liquid chromatography. Dduring the laste decades, peptide separation by ultra high performance liquid chromatography (UPLC) became increasingly popular due to increased robustness and throughput. Furthermore, downscaling of the analytical conditions have proven to increase sensitivity, thus currently microliter- and even nanoliter per minute flowrate instruments (µUPLC and nUPLC) are commonly used for peptide separation in proteomics. The principle of separation of based on the hydrophobicity of the analytes. In nUPLC systems, the separation takes place in a column, that is filled with functional material. In most cases, silicagel particels of around 1.8 µm in size, that are modified with C18 alkyl chains on the surface, are used as functional material. The peptides from the sample are loaded onto the column using aqueous solvents. In consequence, the partly hydrophobic peptides tend to interact with the C18 side chains and are rather retained on the column, then eluted together with the aqueous solvent. When gradually increasing the amount of organic solvent, the peptides elute in dependency of their hydrophobicity. Controlled by the steepness of the gradient (*i.e.* change in % organic solvent per time), the number of coeulting peptides that leave the separating column is reduced to a manageable number [69].

As the total number and diversity of the analytes is quite high, a chromatographic separation before mass analysis is always a trade off between generic and easy applicability, speed and accuracy. But especially when analyzing sub-populations of peptides, such as phosphopeptides, the chromatographic procedure has to be thoroughly evaluated. The additional phosphogroup within the analytes introduces an additional hydrophilic component, thus the retention of the phosphopeptides is reduced compared the unmodified peptides. On the other hand, the complexity of the sample is reduced and thus the gradient steepness and time might be adjusted for optimal analyte distribution over the gradient time.

After successful separation, the analysis of the peptides is performed by mass spectrometry. The development of the soft ionization technique electrospray ionization (ESI) and

increased operability / accessibility of mass spectrometers made it possible to identify the mass-to-charge ratios (m/z) of intact proteins and peptides as well as their amino acid fragments in a large scale [70]. However, the technical principles of analyte separation by m/z and ion detection are diverse and each technique serves certain applications. In most proteomics applications, three types of mass separators and analyzators are used: quadrupole, time-of-flight and orbitraps. The separation and measurement principles are often integrated into one hybrid instrument such as triple-quadrupole or quadrupole and TOF and many more combinations to meet complex analytical requirements in proteomics as summarised in figure 1.1.

Quadrupole instruments are typically cost effective, highly sensitive and robust but on the other hand offer a limited m/z separation power. They consist of at least four metal rods, that create en electrical field by superposition of a constant direct voltage (DC) and a alternating current voltage (AC). When positively or negatively charged ions enter the quadrupole from one side, the design of the prevalent electrical field causes the ions to follow certain paths that guides only ions of a certain m/z ratio stably to the other side of the quadrupole. The flight path is dependent on the applied ratio between DC and AC, thus quadrupoles act as m/z filters, which can easily be calibrated. In order to detect the number of ions that successfuly pass the quadrupole a detector has to be placed at the outlet of the quadrupole, usually an electron multiplier tube [71]. The filtering ability of quadrupole allows three operation modes, as described in table 1.1.

A commonly used sequential combination of three quadrupoles (figure 1.1 D) is often used in targeted proteomics, where the analytes are already known and characterized and specially targeted within all present peptides. For this, the first quadrupole is used in SRM mode for a previously selected peptide m/z of interest. Consequently, only that specific peptide and other analytes with the same m/z ratio can pass through the quadrupol and serve as precursor for the following fragmentation. The second quadrupole serves as collision cell, it is filled with gas molecules and set to RF mode. While passing through the second quadrupole, the precursor collides with gas molecules. This collision induced fragmentation (CID) generates amino acid oligomer fragments of various length. The third quadrupole can be operated in Scan mode, to obtain information about all generated fragments (to confirm identity by sequencing the amino acid fragments) or in SRM mode,

for the accurate, robust and most sensitive quantification by a selected fragment only [72].

The disadvantage of QqQ instruments is a low m/z resolution compared to TOF and Orbitrap instuments. It is calculated by dividing the observed m/z value by the full width of the peak at half maximum (FWHM) [73]. Typically, quadrupoles can reach a resolution ranging from 500 to 5000, which is not enough to calculate the theoretical sum formulas or match peptide masses from databases with adequate probability. Therefore, the main use of these instruments is the robust and sensitive quantification of known and characterized substances, *e.g.* the monitoring of known biomarker peptides/proteins of patient serum samples in the clinic.

For the unbiased discovery of proteome wide changes, a considerably fast and accurate measurement of peptide m/z values with higher resolution is required. Time-of-flight mass spectrometers fit for this purpose, as their measurement principle provides typically a resolution in the range of 10 000 up to 50 000 [74], in some instruments also up to 300 000. The obtained m/z values for peptide precursors are usually within the range of ±15 ppm, which adds the required confidence to the correct database search of the theoretical peptide mass compared to the measured m/z values. Furthermore, TOF instruments typically are able to measure mass spectra very fast with a frequency of up to 120 Hz. The measurement principle is quite simple: charged ions are accelerated into a drift tube and the time the ions need to reach the end of the tube is measured. As the kinetic energy

Mode	**Description**
SRM/MRM	Single/Multiple reaction monitoring. A single or multiple fixed AC/DC (*i.e.* m/z value) are set up and solely measured.
Scan	The continuous flow will pass through the quadrupole while the AC/DC ratios (calibrated to known m/z values) are ramped. Thus, unknown m/z can be identified.
RF	Radiofrequency only mode. The quadrupole is set in a way, that all ions will be transmitted, independent on their m/z ratio.

Table 1.1: Measurement modes of quadrupole mass analyzers. Typically, three different modes can be used: SRM, Scan and RF. Within a quadrupole, only one of the modes can be present. Depending on the application, multiple quadrupoles have to be combined to make use of using multiple measurement modes sequentially.

is the same for each charge state, the flight time is solely dependent on the analyte mass. As depicted in figure 1.1 B, some instrument types include additionally reflectrons into the flight path to increase the length and thus the resolution. In general, the resolution is constant over the whole m/z range in contrast to quadrupole and orbitrap instruments and there is no theoretical upper limit for the measurement of m/z simultaneously. Having the flight path length as main driver for the resolution and m/z, it is crucial that it remains constant after calibration. Even small temperature changes or movements of the instrument make a new calibration necessary. A possible strategy to overcome such robustness issues is the use of an internal calibrant, that is spiked in during the data acquisition. Data processing software can then use the known m/z information of the calibrant to correct all acquired mass spectra if needed. This strategy requires additional measurement time, that is therefore not available for the analytes. Other instruments make use of temperature control and heavy insulation, but still calibration is frequently required. Often, quadrupol separators are installed before TOFs (figure 1.1 F). This allows the high resolution precursor m/z measurement while the quadrupol is in RF mode, but also allow the preselection using SIM mode with subsequent fragementation to obtain high resolution m/z values of the resulting fragments.

In summary, time-of-flight instruments serve as excellent trade off between low cycle time (high scan frequency), sensitivity and also flexibility for combination with other separation techniques such as ion mobility. On the other hand, they suffer from susceptibility to environmental changes and thus require frequent calibration. In addition to that, the space requirements are comparably large, as the flight tube requires large physical space.

The third commonly used type of mass analyzer is the Orbitrap exclusively distributed by Thermo Fisher Scientific. The Orbitrap consists of an inner core and two shells, that are isolated against each other (figure 1.1 C). Between core and shell an electrical field is applied, that causes the injected ions to circulate around the core while oscillating along the core axis at the same time. The oscillation frequency is dependent on the m/z ratio of the analyte and causes a measureable potential difference between the isolated outer shells. As all frequencies are superimposed in the transient diagram, fourier-transformation has to be applied to obtain the raw frequencies, that can be calibrated with known m/z ions [75]. It allows the acquisition of high resolution mass spectra with adjustable resolution,

that is dependent on the m/z value. The higher the measured m/z value, the worse the resolution. Resolutions up to 480 000 are possible, but are unpracticable in proteomics, as a higher resolution requires substantially higher measurement time. Still, typically applied resolutions range from 15 000 to 60 000 while operating at a slightly lower m/z scan frequency than TOF instruments. In general, Orbitraps are always coupled to ion routing systems, traps and quadrupoles (figure 1.2). After ionization, the ions are guided trough a separating or RF mode quadrupole, after which the charge counter leads a predefined amount of charges into a trapping chamber. For the measurement of precursor m/z, the ions are routed to the Orbitrap device and measured. For the measurement of peptide fragments, the peptide m/z of interest is already separated in the quadrupole, guided through the charge counter into the trap, where the fragmentation takes place. In this case, higher energy collision induced dissociation (HCD) is applied to generate the peptide fragments [76]. Afer that, all ions are measured in high resolution in the orbitrap.

An Orbitrap mass spectrometer offers a very high resolution while being robust to environmental changes. In addition, they cover a broad dynamic range [77] and are able to resolve reporter ions for labelled peptide quantification, such as tandem mass tags (TMT, branded by Thermo Fisher). On the other hand, the cycle time per each scan is increased and the instruments are solely available from one manufacturer.

1.2.5 Ion mobility spectrometry

Due to the fast m/z scan time of TOF instruments, additional separation techniques that require separation times that are between liquid chromatography (several seconds for one peptide) and mass spectrometry (1/120 s for one spectrum) can be interposed. Typically, ion mobility separation (IMS, figure 1.3) fits that spot. IMS enables the measurement of an additional feature for each peptide (retention time, ion mobility and m/z values) within milliseconds, which increases the confidence in identification and opens the door for substances that coeulte from the column (equal retention time) and are isobaric (equal m/z) and would thus not be identified separately without the additional IMS feature. An important example are phosphopeptide isomers, that differ by the position of the phosphate group within the amino acid sequence. Several principles of ion mobility have been used together with mass spectrometry in the past years [80]. The recently developed

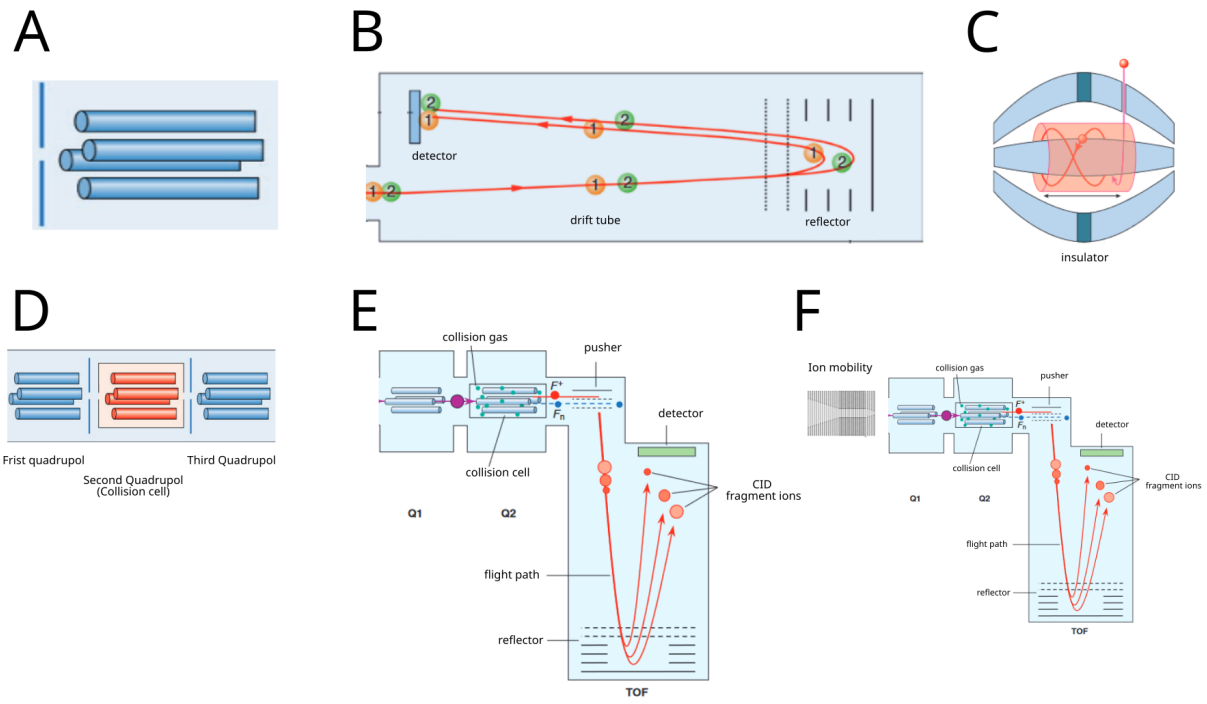

Figure 1.1: Overview of different mass spectrometer types. A) Single quadrupol MS. B) Time-of-flight MS. C) Orbitrap MS. D) Triple quadrupol MS. E) Quadrupol coupled to time-of-flight MS (qTOF). F) Tandem ion mobility separation coupled to qTOF. Images adapted from [78, 79]

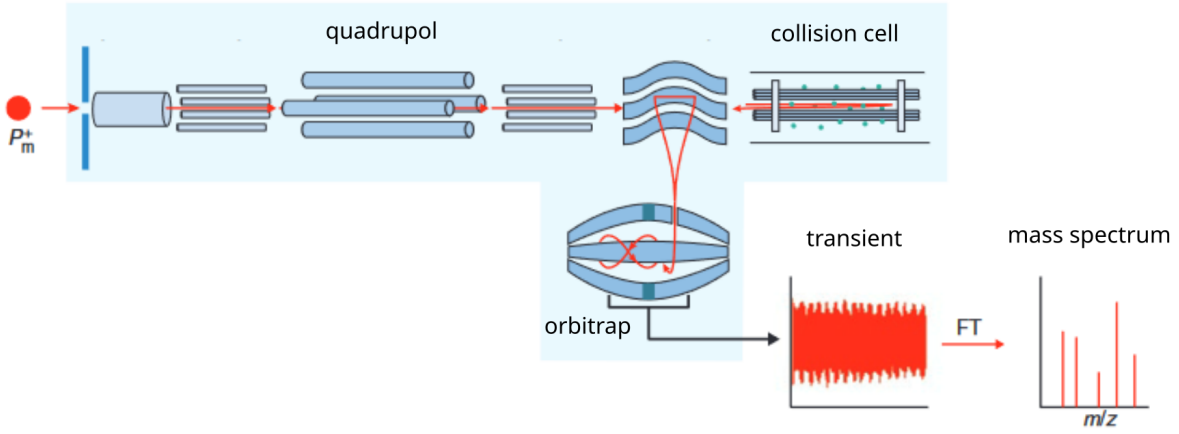

Figure 1.2: Operating principle of the Thermo Fisher Exploris 480 exemplary for Orbitrap Mass Analyzers as shown in [78]. The ionized molecules can be separated by the integrated quadrupol first, before fragmentation. After back-transfer of the ions from the collision cell, the ions enter an electric field inside the orbitrap, that causes a circular and oscillating movement of the ions along and around the inner core rod. By measuring the frequency of the voltage differences between two isolated outer shells caused by the oscillation of the ions, the m/z ratio can be determined after Fourier-Transformation and calibration.

trapped ion mobility spectrometry coupled to a TOF instrument (timsTOF from Bruker Daltonics) enables the trapping of ions in the IMS device before measurement, which increases sensitivity. The principle is simple: charged ions are pushed by a constant gas stream into the IMS tube, while a counter directional electric field with increasing field strength is applied along the IMS path. Thus, ions are trapped at the position where the counter directed electric force equals the pushing force caused by the constant gas flow. The greater the collisional cross section (CCS) of the molecule is, the greater is the pushing force and the greater the counter directed electric field has to be. In plain words, the bigger the molecule, the further it can move along the IMS tube. By lowering the potential from the proximal end of the IMS tube sequentially, the ions are released into the qTOF instument for measurement. This feature has proven to increase the sensitivity and number of identifiable analytes in proteomics samples, but has also proven benefits in lipidomics or metabolomics [81].

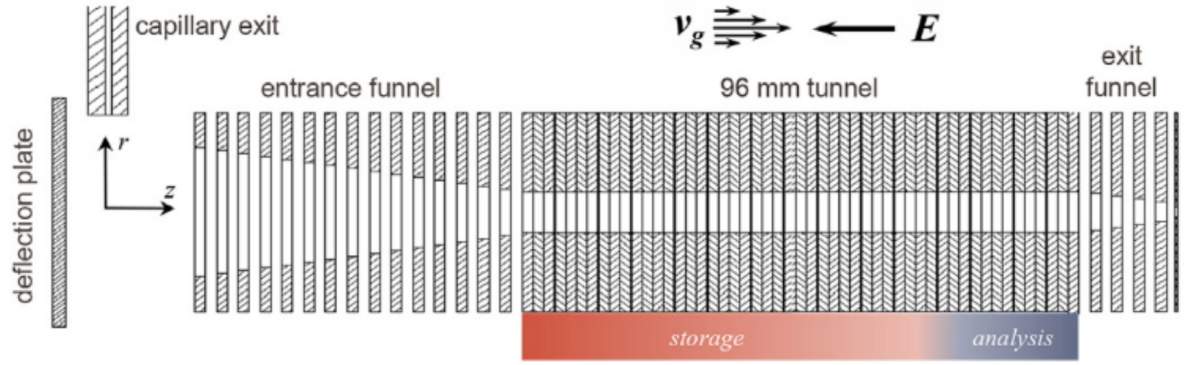

Figure 1.3: Scheme of a trapped ion mobility spectrometer (tims) adapted from [82]. Ionized molecules enter from left and are pushed by gas flow into the tims device. There, a counter directed electric field traps ions at the position where the counter directed force by the electric field equals the forward directed force on the ions caused by the pushing gas flow. The molecule geometry in the gas phase determines the area that is affected by the gas flow. Thus, larger molecule geometry, *i.e.* cross collisional section (CCS), causes greater forces acting on the ions, the ion mobility is increased. In consequence, a gradient electric field along the tims device is applied to trap ions at different positions along the tube, separated by their ion mobility.

1.2.6 Visualization of mass spectrometry data

The visualization principles for LC-MS/MS raw data is similar for all MS types. The most basic type of chromatogram is the total ion count (TIC). For this visualization, the sum of each signal in every acquired MS scan (typically MS1 level only) is calculated and this

value is plotted against the time point it was acquired (*i.e.* retention time) as depicted in figure 1.4 A. This way, a general overview is given and can be compared in each sample. Also, it quickly answers technical questions, such as ESI spray stability and ionization efficiency as shown in figure 1.4 B. Often, MS1 spectra are dominated by one very high intensity precursor, the so-called base peak. When extracting not the sum of all intensities, but solely from the base peak, we can obtain chromatographic information on precursor information, as can be seen in figure 1.5. In more detail, a base peak chromatogram (BPC) shape that follows nearly Gaussian distribution is desirable, but can also show tailing or fronting. This often indicates a void volume within the system that causes this peak broadening or fouling of the analytical column. In consequence, all connections have to the re-fitted and column aging and performance has to be monitored closely. A third very important visualization principle is the extraction of signal intensities for distinct masses only throughout every MS1 spectrum. This principle is called extracted ion chromatogram (XIC) and helps to identify single precursors, such as known analytes and contaminants with known m/z ratios from the raw file directly.

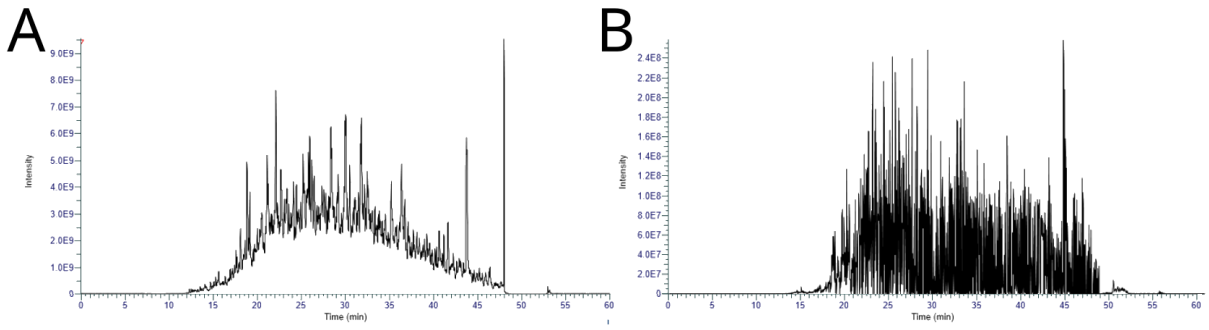

Figure 1.4: Examples of a typical proteomics peptide chromatography run with A) TIC of technically successful raw data B) spray instability during data acquisition

1.2.7 Data dependent versus data independent acquisition mode

Typically, the full m/z spectrum for any time point during chromatography is measured, harboring the information about the ion intensities and m/z ratios for all coeluting peptides at this specific retention time, which is called MS1 scan. But for the successful identification of the eluting peptides, the precursor m/z is not sufficient. Rather, amino acid composition by analysis of the contained peptide fragments by processing software or manually gives satisfactory confidence and proof for the correct identification. This can be achieved using the CID or HCD capabilities in modern mass spectrometers, the peptides

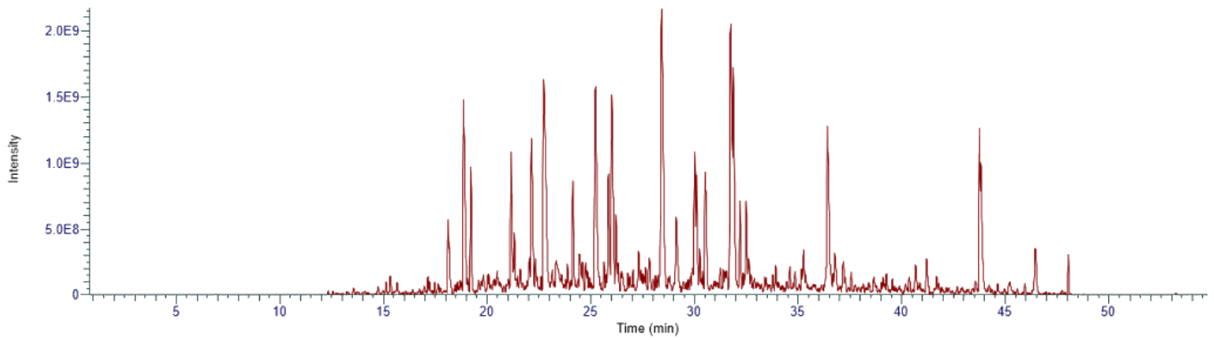

Figure 1.5: Example for a base peak chromatogram of a typical proteomics peptide chromatography. The most intense signal in each MS1 is displayed, in proteomics experiments these are typically peptides.

are fragmented and yield fragments of various length as described in figure 1.6 A, which is called MS2 scan. This is also the rationale behind the abbreviation LC-MS/MS, as two different kinds of mass spectra are collected after LC separation. For the acquisition of fragment spectra, currently two strategies are employed: data dependent acquisition (DDA) or data independent acquisition (DIA) as described in figure 1.6.

In data dependent acquisition, the top N most intensive m/z ions are identified from the MS1 scan (precursor spectrum, in proteomics typically precursors are peptides) by the operating software of the mass spectrometer and sequentially selected with very narrow window (*e.g.* ± 0.5 Thompson) by the quadrupole for fragmentation, so their MS2 spectra (fragment spectra) can be collected. The selected number N of most intense ions is typically between 10 to 25 and can be chosen depending on the instrument speed and on the analytical need. When short LC gradients and highly complex MS1 spectra are present, a high N is needed for deep peptide coverage. On the other hand, a high N costs measurement time and MS1 quantification accuracy. In general, the DDA strategy decides depending on the MS1 information, which precursors are selected for fragmentation. It provides clean and high quality spectra, that can also be used for denovo sequencing. In addition to that, the data processing is not computationally intensive and implements easy and straight forward algorithms that are accessible to a broad community.

In contrast to that, in data independent acquisition strategy, no preselection of the precursors is performed. The fragmentation is independent of any MS1 information present. Instead of choosing a very narrow window for selecting the precursors for fragmentation,

a wider window of precursor m/z are allowed to pass through the quadrupole. This way, multiple precursor cofragment and create chimeric MS2 spectra, where the assignment of precursor and their corresponding fragments is not easily possible. More complex bioinformatic algorithms have to be applied to elucidate the amino acid evidence for each precursor [83]. This also includes the use of spectral libraries, that contain a set of pre-aquired high quality spectra of already assgined peptide annotations. Spectral libraries are either labor intensive or computationally intensive to create. Recent developments in the proteomics community show improvements in algorithms and software to actually process DIA generated rawdata in a comprehensive and user friendly way [84] as well as the accessibility to high performing computer systems have paved the way for increasing use of DIA. The major advantage of DIA is a robust and accurate quantification as well as the decrease of missing values, due to the fact that no selection of precursors is performed and instead also borderline signal intensities are fragmented and have the chance to be identified and quantified.

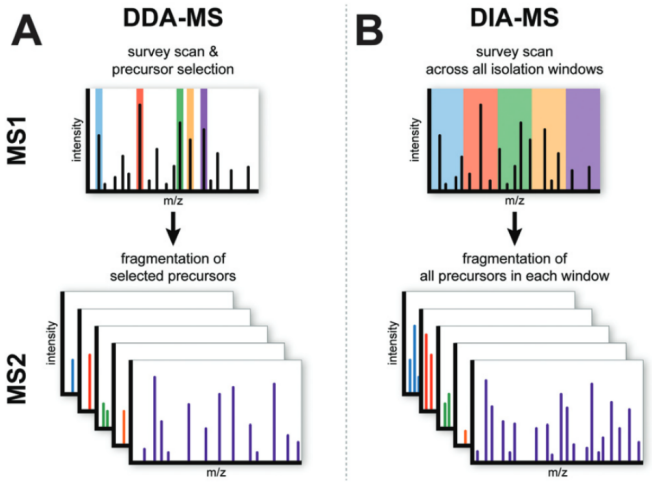

Figure 1.6: Principles of data acquisition. A) In data dependent acquisition (DDA) mode, single precursors are selected and fragmented for identification. This mode yields comparably clean spectra, but with limited capacity. B) In data independent acquisition (DIA) mode, predefined m/z ranges are fragmented simultaneously. Thus, the resulting MS spectra are more complex compared to DDA, but all peptides within this m/z range can possibly be detected. Figure adapted from [83]

1.3 Peptide identification, quantification and bioinformatics

Ultimately, proteomics research aims to gain unbiased insight into biological systems. The path towards this insight includes the correct identification of measured peptides, combining this information on proteome level or PTM level and quantify the results. This way, the protein and PTM quantities can be used to identify relevant changes by using simple statistical methods including t-test, linear models and multivariate statistics such as principal component analysis (PCA), Uniform Manifold Approximation and Projection for Dimension Reduction (UMAP), t-Distributed Stochastic Neighbor Embedding (t-SNE) and other unsupervised machine learning algorithms like hierarchical clustering, clustering by k-nearest neighbors, dbscan and random forest. Databases with previously collected knowledge are then necessary to conclude what those changes and relationships actually mean and to combine the findings with the observed biology. A plethora of methods and tools are available, but most often share the same principles.

The following sections want to introduce the used tools and the principles behind them as well as promising advancements and limitations.

1.3.1 Processing of DDA data

With a protein database at hand, e.g. from Uniprot [85], the protein sequences can be digested *in-silico* according to the sample preparation protocol, the user defines search parameters e.g. as noted in table 1.2. The exact settings are variable and depending on the experimental designs, the question at hand and also the used software, but in general these are the most important and usually the minimum required input from the user. The *in-silico* digest provides a list of possible, expectable peptide masses that are consequently compared to the measured intact peptide masses (i.e. precursor mass before fragmentation). Usually, multiple candidates will fit to the measured peptide mass which makes additional features for the correct identification necessary. Fragmentation of the intact peptides delivers possible combinations of amino acids from the peptide sequence. The DDA strategy ideally provides clean MS2 spectra deriving from only one selected precursor after fragmentation. The mass differences between adjacent fragment

m/z values can only meet the mass of one naturally occurring amino acid or a multitude and combinations of those. The fragmentation patterns are often not complete, fragments are missing or types of ions are favored in certain fragmentation techniques. Therefore, scoring approaches are necessary to evaluate the probability of correct identification.

Proprietary as well as open source software is commonly used in the proteomics community. Each software regardless the respective publication status has its own strengths, although they serve the same purpose. The most commonly used open source software is MaxQuant developed by the Cox group [86], which accepts most mass spectrometer raw file formats with and without ion mobility information available. Especially the label free quantification (LFQ) strategy MaxLFQ [87] is known to be very strict on one hand, but also very accurate on the other [88] and is also used in other software for quantification. Another commonly used open source search engine is MSFragger by the Nesvizhskii group [89]. It is known to provide search results very fast and is compatible with many other

Parameter	**Typical value**
Digestion enzyme	Trypsin
Allowed missed cleavages	1
Fixed modifications	Carbamidomethylation (*i.e.* alkylation on cystein)
Variable modifications	Oxidation on methionines N-terminal acetylation Phosphorylation on S/T/Y
Max. no. of variable modifications	4
Database	FASTA file *e.g.* Homo sapiens, downloaded on 24.04.2022 from Uniprot reference proteomes (UP000005640) containing 20 577 protein sequences and including 172 most common contaminants

Table 1.2: Typical search parameters for database search always include a type of proteolysis (or no proteolysis, if no enzyme was used for digestion), an acceptable number of missed cleavage, biochemical modifications to the peptides (fixed or variable) as well as a source of protein sequences (typically a FASTA database)

proteomics tools that can be incorporated in the FragPipe software suite. In addition to that, by using FragPipe for the library generation for DIA experiments it has been shown to boost the performance of DIA identifications significantly [90]. A very popular proprietary software for the comprehensive analysis of DDA data is PEAKS by Bioinformatic Solutions Inc. [91]. PEAKS is able to perform and provide DeNovo sequencing information together with the database search which aids confident identification and to answer discovery related questions such as novel identifications of mutation or isoform discovery. PEAKS has also been shown to perform well with unspecific searches such as required by immunopeptidomics [92].

1.3.2 Processing of DIA data

In contrast to the straight forward identification procedure using DDA data, DIA requires sophisticated algorithms to perform the identification. Two different strategies are applied for the processing of DIA data: Spectrum centric and peptide centric approach [93].

The spectrum centric approach is similar to the analysis of DDA data. During the analysis, peptides typically elute over a time frame of several seconds and follow ideally a gaussian curve. Consequently, also their fragment ions follow this pattern. When comparing the overlay of fragment spectra with the precursor elution profiles, pseudo-spectra can be extracted that look similar to DDA like data [94]. This way, the resulting pseudo-MS2 spectra can be searched either via database search or can be compared to previously generated (by DDA measurements) spectral libraries. In contrast to this approach, the peptide centric approach assumes all theoretical peptides of a digest can be found in the raw data [95]. Therefore, peptide centric algorithms search for evidence (*e.g.* retention time, ion mobility, fragments present ...) of the peptides of interest and reports scoring values.

Peptide centric approaches work best using a spectral library. The generation of such a spectral library requires the acquisition of high quality MS2 spectra for every possible peptide for the particular species, cell line or project. Peptide spectra that are not present in the library, can not be identified in the samples of interest. Therefore, much effort is required to provide a broad coverage of the proteome. Usually, a pool of all samples is fractionated by strong-cation exchange (SCX) or high-pH reversed phase fractionation

(hpH-RP) [83], and the fractions are measured in DDA mode to obtain sufficient quality. By distributing the number of peptides through the fractions while measuring each fraction in DDA with high peptide capacity (Top N) and long gradients, the total number of identifiable peptides is increased. The downside is the time intensive measurement and data processing, that has to be done before the first sample of interest can be measured. But once this library is generated, it can be used for many consecutive analysis, as long as the sample type remains the same. Therefore this approach is a valuable alternative when large scale studies are of interest. To reduce the required wet lab work and measurement time, effort has been made to predict spectra *in-silico* from FASTA files [96]. This way, also smaller projects can profit from the benefits of DIA, without the need for intensive preliminary work. It has been shown, that although wet-lab generated spectral libraries still achieve the highest numbers, *in-silico* predicted spectral libraries have increased in performance over the past years. Very recent research has proven that the neuronal-network assisted library-free analysis of DIA data can even exceed the identification performance of wet-lab libraries [84]. Still, the generation of predicted spectral libraries requires exceptional computational performance and thus very long analysis runtimes. Especially, when adding PTM level information to the library entries, the number of possible precursors to predict grows exponentially. The generation of a predicted library including up to four phosphosites requires roughly one week on a computer with 128x cores of 2.9 GHz and 256 GB of RAM with DDR4 technology, but once it is generated it can be used for many follow-up studies.

The complexity of the chimeric spectra is already very high when analyzing unmodified peptides which is even increased in the presence of modified peptides. Especially phosphoproteomics DIA experiments create supercomplex MS2 spectra that were unsearchable few years ago without wet-lab spectral library, as the required computational resources for the prediction were not affordable. But the generation of a spectral library requires high amounts of sample material, that often is not available in phosphoproteomics experiments. Very recent software improvements and the availability of affordable computational resources made it possible to investigate the broader use of DIA for phosphoproteomics. Few publications are published up to date addressing the issues and give guidance with DIA in phosphoproteomics [97, 98], so this featured work is one of the few DIA phosphoproteomics datasets available that shed light on remaining questions with this.

In general, the processing of DIA data requires either a laborious library generation or sophisticated computing resources. In addition to that, software requirements for the actual identification and visualization of the results are not as trivial as for DDA data. In exchange for that, the number of missing values is reduced and the robustness of the dataset is increased, which is especially relevant for large scale and long term studies. Common software for DIA data processing include the open-source Skyline and DIA-NN [84, 99]. A widely applied proprietary software is Spectronaut from Biognosis.

1.3.3 Quantification strategies

For identification and quantification, usually a feature is constructed that includes all possible information about the precursor such as retention time, m/z, isotope pattern, ion mobility, peptide sequence and signal intensity. First, the identification of features can be transferred between runs (*i.e.* match between runs - MBR), where the feature is identified confidently but no MS2 spectrum proves the identification. This way, the number of missing values is reduced, but no direct prove for the peptide can be provided in some samples. Next, all features of each sample can be used to normalize their intensity values, if required. The assumption behind this is, that no major changes happen to the majorities of peptides. One strategy for this uses the sum of signal intensities (total ion count - TIC) for a given MS1 spectrum (*i.e.* retention time) for normalization. Others implement internal standard peptides, that assign a fixed normalization factor for the whole sample. Sophisticated normalization using variance stability normalization or cyclic loess [100] can also be applied after data processing, before statistical analysis. In the last step, peptides are selected for calculation of protein abundance in each sample. Each software has own featured quantification algorithms as described in table 1.3. Other quantification strategies involve the chemical or metabolic isotopic labeling of peptides and combination of samples (multiplexing) before analysis, but this strategy is not applied here and can be reviewed in [101]. Advantage of such labeled quantification is a reduced number of missing values. On the other hand, the reagents are cost intensive and multiplexing is limited to a maximum of 18 samples [102]. Furthermore, very high resolution in low mass regions (typically the mass range for the reporter ions of the labels) is required where TOF instruments perform less effective than Orbitrap instruments, thus the efficient use of instrument time is reduced as Orbitrap instruments typically require longer measurement

times. Therefore, a label-free quantification strategy was applied to the featured sample sets.

1.3.4 Preprocessing, missing values and statistical testing

In proteomics, often perturbation experiments are conducted to elucidate the proteomic and PTM response of a biological system to a stimulus. In the most simple case, a control group of samples is compared to a sample group including perturbations *e.g.* drug treatment or loss-of-function, each with adequate number of replicates. For proteomics experiments, often three replicates are sufficient to gain statistical confidence, as such workflows have proven very robust especially using the DIA strategy. The analysis of PTMs such as phosphorylation requires the data anlysis on peptide level, thus a higher variability in the raw data values is expected. In addition to that, especially phosphopeptide enrichment introduces an additional source for variability, that has to be accounted for. In these cases at least four, preferably five biological replicates are required to achieve sufficient confidence.

Algorithm	**Description**
Top N	Takes the average or sum of the N top intensity peptides as protein quantity. The advantage is an easy implementation but the approach can not cover broad dynamic ranges, low intensity signals are often neglected. Therefore, some peptides might provide intensities outside the linear range of the instrument.
MaxLFQ	Compares the fold-changes of each peptide individually per sample and selects those peptides, that follow similar patterns. [87]
iBAQ	Intensity based absolute quantification. The sum of all observed peptides is normalized to the number of theoretically identifiable peptides, which approximates the absolute protein abundance. [103]

Table 1.3: A selection of quantification algorithms. Although this is not an exhaustive list, Top N, MaxLFQ and iBAQ are the most commonly used aglorithms. All algorithms have in common to take into account a sub-population of the identified peptides per protein. They differ mainly in the strategy how to select the peptides, based on their intensity (Top N), peptides changing in comparable manner (MaxLFQ) or normalization per protein length (iBAQ). Depending on the analytical need, the selection of the appropriate quantification strategy is key.

Many statistical methods are similar in -omics sciences and can be transferred from broadly applied and well known technologies, such as transcriptomics and genomics. Before identification of the statistically significantly changing events, data preprocessing is sometimes necessary to reduce batch effects or compensate for fluctuating values throughout time (*e.g.* temperature changes caused by day and night might influence the flight path of TOF instruments and thus influence the measured values, clogging or fouling of the analytical chromatography column or the ESI emitter and many more). An overview of possible measures to account for systematic but compensateable changes can be found in table 1.4. A comprehensive review for protein level data can be found in [100].

Another critical and controversial discussed strategy is the compensation for missing values. The question is in this case: Is data completeness required for the successful analysis of the dataset? The answer to this is not trivial and strongly dependent on analytical aim of the study. While statistical testing such as t-test or linear models are to some extent robust to missing values, advanced multivariate statistical analysis such as unsupervised clustering, PCA *et al.* require data completeness. In these cases, either very stringent filtering and/or the artificial calculation of replacement values (imputation) has to be done before further analysis.

Whereas in other -omics technologies missing value imputation plays a minor role, it becomes relevant in proteomics. Even when applying the DIA strategy, missing values per sample of up to 10 % are common. In most cases, proteins with a high number of missing values across the dataset typically are either very small and thus yield less detectable peptides or are expressed in comparably low abundance and thus yield signal intensities below the lower limit of quantification/detection (LLOQ/LLOD) which is strongly dependent on the dynamic range of the sample, the data set can be described as left-censored. Both scenarios create missing not at random values (MNAR) and for their missing value imputation multiple robust methods are available, as summarized in table 1.5.

A typical phosphoproteomics dataset measured in DDA comprises about 75 % missing values. The reason for the comparably high number of values is on one side the fact that missing value information is counted on peptide level, instead of protein level. On protein level, missing peptides can be compensated, as multiple peptides can be considered for protein identification and quantification whereas on peptide level no other source of infor-

Strategy	Type	Description
Pre-measurements	Acquisition	The sample small quantity is analyzed in advance to the main measurements, the resulting total ion counts (TIC) are compared normalized and the normalization factors are used to adjust the injection volume for the main measurements, so an equal amount of analytes in injected in each run.
Randomization	Acquisition	Randomized injection of samles distributes possible batch effects across the whole dataset, robustness is increased.
Internal standards	Acquisition	Addition of internal standards *e.g.* in the final resuspension solution to compensate for changes in ESI efficiency and mass spectrometer differences.
Sample pool QC	Acquisition	Pool small quantity of each sample and inject every N sample aqcuisition. By monitoring the intensities of the pool peptides over the whole dataset, a normalization factor can be estimated if necessary.
Variance stability (VSN)	*In-silico*	Transformation of the intensity values to minimize differences in variance. [104]
Local regression (loess)	*In-silico*	Linear regression for distinct intensity regions, as it is assumed that the bias to be corrected is different for each local intensity range. [105]

Table 1.4: A selection of strategies for reducing batch effects and increasing robustness of statistical testing. Depending on the analytical aim, more than one strategy can be applied. Generally, normalization can be applied already during the acquisition or *in-silico* during analysis

mation is available than the peptide itself, this source of missingness can be regarded as MNAR. On the other hand, the phosphopeptide enrichment sample preparation step not only introduces an additional source of variability, but also a random selection of peptides to be enriched. The nature of the enriched phosphopeptides is also strongly dependent on the present contamination such as salts or detergents [23]. Therefore, missing values in phosphopeptides analysis is regarded as missing at random (MAR) and the imputation

is not trivial. A selection of methods for imputation of MAR values is described in 1.5.

Strategy	Type	Description
Filtering	all	Remove observations with missing values from the data matrix.
Minimum value	MNAR	The minimum value of the dataset is used to replace missing values.
Stochastic minimal value	MNAR	A Gaussian distribution based on a given minimal value as average is created, random values from this distributions are used to replace the missing values.
k-nearest neighbors based (kNN)	MAR	Observations with missing values in sample A, that have at least a predefined number of samples B (C, D, ...) with measured values, can be estimated by identifying k similarly (nearest) behaving entities with existing values in A as well as B (C, D, ...). Based on the existing values, the missing value is replaced. [106]
Singular value decomposition based (SVD)	MAR	In the first step, all missing values are substituted by the row mean followed by SVD analysis, which similarly to principal component analysis, creates eigenvectors to describe the dataset. Missing values are then imputed by calculation from the eigenvectors. Repeated until the change in the matrix falls below a threshold of 0.01. [106]

Table 1.5: A selection of strategies for reducing the number of missing values. The cause for missing values is key to choose the appripriate strategy. Two types of missing values can be differentiated, missing at random (MAR) and missing not at random (MNAR). Replacing MNAR values is generally more robust, as a rationale for their missingness is usually present. E.g. missing values due to low intensity. In this case a minimum value approach might aid to gain the correct conclusions from the dataset. On the other hand, replacing MAR values require more sophisticated approaches. More information have to be taken into account to increase the propability for correct replacement of the missing values. Often, phosphoproteomics experiments suffer from MAR values, as the sample preparation procedure is very sensitive towards slight changes in the experimental conditions.

For the statistical testing of significantly changing observations, missing value imputation is facultatory, as long as a sufficient number of measured values is present. Two major approaches are dominating the statistical tests in -omics sciences: t-test and linear models.

In general, the statistical tests provide a p-value for each peptide/protein by calculating a test-statistic from parameters of the data sets. Depending on the desired stringency, observations with a p-value below 0.05 can be considered as statistically significant, as commonly accepted in many scientific disciplines [107].

A very easy and straight forward method to identify if peptides or proteins are different in two conditions is the well known Student's t-test. By comparing average and standard deviation of two independent data sets, a probability is calculated whether both data sets belong to the same distribution (*i.e.* no effect of the condition) or belong to two different distributions (*i.e.* the condition has an effect on the peptide/protein). Prerequisite is that the measured values of both conditions follow the Gaussian distribution, which has to be verified before. Furthermore, the classical t-test has been proven to work best when the variances are equal in both conditions, whereas alternatives such as Welch test recently were discussed to have general superior power compared to t-test and should be used instead [108]. In case of t-test, the required parameters for the test statistics to calculate the p-value is the average value in each group, the average of all observations, the number of observations and the standard deviation [109]. The hypothesis to be tested is usually two-sided, *i.e.* 'the average of both data sets is equal' is the null hypothesis and that 'they are not equal' is the alternative hypothesis, as it is not known *a priori* which peptide or protein is up- or downregulated. A special case to be considered is sample pairs *e.g.* when the same individual sample is used as control (before treatment) and as condition (after treatment). In this case, a paired t-test can be performed which increases the statistical power of the resulting p-value by de-noising the datasets [110]. The statistical power of the t-test is tied to the number of missing values. In theory, if enough values are present over all data sets, and equal variance is assumed, this is sufficient to perform the test. Practically, it is not always obvious that equal variances are present. In consequence, at least three values in each condition are mandatory.

In contrast to the straight forward t-test or Welch test, more sophisticated statistical methods have been evaluated in -omics sciences. Decades ago, linear models have been identified as superior method in transcriptomics and algorithms have been implemented in R packages that are available from Bioconductor [111]. Here, linear regression is used to build a linear model between the data sets. The statistical test is performed with the

hypothesis that the slope, which is the coefficient of the linear model, is not zero and requires the slope and the residual error calculated from the regression as input values. One advantage of linear models over classical t-test is the robustness of the resulting p-value towards missing values. This approach is able to identify potential changes with more accuracy, with decreased sensitivity towards a high number of missing values. Well renown R packages include linnomr, edgeR, DSeq2 and limma, while all of them are widely applied in transcriptomics already, only the latter became increasingly popular in proteomics [112].

Disregarding the missing value problem, a general strategy about the acceptance criteria for significantly changing peptide/proteins has to be evaluated for the scientific question at hand. Usually, the ratio of average value or median values for each sample group is calculated (fold change - FC) and only 2-fold and 0.5-fold changes are accepted, *i.e.* a factor of two. This is an arbitrary value that was historically estimated to reflect the maximum variability introduced by sample preparation and measurement. But with increasingly robust methods, this dogma begins to change and with consequent validation of the analytical method also other FC thresholds can be accepted. In addition to that, a minimum number of measured values (or maximum number of missing values) can be evaluated, which is well feasible with proteomics data, but becomes problematic for phosphopeptide data, as the number of missing values is usually much higher. For proteomics datasets, usually at least 60 % of the measurements of sample condition (or even of one single sample in case of technical replicates) should be present, whereas in phosphopeptide datasets 50 % across all measured data serves as acceptable trade off between data completeness and statistical analysis.

Another important consideration is the number of false positive identifications. When applying a significance threshold of p-value < 0.05, it is known that by chance 5 % of all null hypothesis rejections is wrong, *i.e.* a false positive. When testing thousands of proteins, the number of false positives is significant, but strategies to adjust p-values for adjusting the p-value are available [113]. The aim of all strategies is the elevation of all p-values to a certain degree, so the number of significant instances and subsequently also the number of false positives is reduced. A widely applied strategy for p-value adjustment is the Benjamini-Hochberg method, also known as FDR correction. For this, all p-values

are ranked and multiplied by the ratio between the total number of tests and the p-value rank [114]. An adjusted p-value calculated this way is abbreviated as q-value in the following.

1.3.5 Data visualization and result analysis

After performing the statistical test, several data visualization tools help to get an overview of the conducted experiment, serve as quality control for the applied statistical strategy and allow biological conclusions. Basic tools for typical perturbation experiments include correlation and volcano plots, data reduction strategies such as principal component analysis (PCA), clustering as well as protein-protein interaction networks (PPI) and gene ontology enrichment (GO) or gene set enrichment analysis (GSEA).

Apart from the number of identified peptide and proteins, their correlation in abundance or rank is a most basic visualization for a quick evaluation of the data quality. Furthermore, it is possible to identify outliner samples or unravel unexpected patterns. For this, either the Pearson (correlation of absolute values) or Spearman (correlation of ranks) correlation between all sample pairs is calculated and visualized in a grid-like manner where the color corresponds to the respective value. Commonly accepted is the Pearson correlation, as it proves more robust for small absolute quantitative differences between the protein abundances. As visible in figure 1.7 A, a good correlation within one sample condition is desired and ideally, this differs from the other conditions. In general, a strategy has to be identified how to deal with missing values. Either, only complete observations can be chosen for the correlation, which reduced the informative value of the visualization, or missing values are imputed or replaced by 0's. In any case, this serves as one diagnostic tool for missing value imputation assessment.

For the overview of the statistical testing, a volcano plot is widely applied. A volcano plot as shown in figure 1.7 B shows the log2 of the fold-change on the abszissa and the negative decadic logarithm of the adjusted p-value *e.g.* q-value on the ordinate. Ideally a shape from an erupting volcano is generated, thus the naming. The plot provides information about the general distribution of the data, how much peptide/proteins are changing in total? Is there a bias towards down or upregulated or is the distribution skewed, which hints towards a normalization problem? In a more advanced step, the datapoints can be

colored or labeled according to previous or generated knowledge to further understand the data.

Proteomics datasets consist of typically thousands of descripting observations covering a large dynamic range. Consequently, observations with high quantitative values usually impact the follow-up analysis more compared to low quantity values. To reduce this bias, the data is often log transformed (typically log10 or log2) and normalized (*i.e.* centered). In addition to that, a row-wise normalization can be applied, such as the z-score, where the absolute intensity value is normalized by the standard deviation and the mean of the dataset. Thus, *e.g.* z-score allows a better visualization on heatmaps, as the differences are more visible and independent of the absolute value. The following techniques are usually used with transformed and normalized values, but in general, the raw data should also be evaluated.

Multivariate statistics data reduction techniques aid the identification of different and equal behaving data sets and classification to certain groups based on all descriptors. A commonly used approach is the principal component analysis (PCA), where the descriptors are condensed to few principal components, each with certain loading of the descriptors, that can be used for data visualization as shown in figure 1.7 C [115]. This way, the principal components aim to depict the variance in the dataset and usually a very high percentage (more than 90 %) of explained variance is desired using just two components. Prerequisite for the application of PCA and other multivariate data reduction techniques is a complete dataset. Missing values are not tolerated and have to be imputed in any case. Thus, when a large number of missing values is present, the descriptive power of the PCA is limited. Nevertheless, a successful separation of the sample conditions using principal components is a promising indication of strong evidence for meaningful proteome changes, whereas failing to achieve a separation does not allow to draw any conclusions about the data quality. More recent approaches include Uniform Manifold Approximation and Projection for Dimension Reduction (UMAP), t-Distributed Stochastic Neighbor Embedding (t-SNE), that are already widely used in other -omics disciplines and gained more attention in proteomics lately. The aim of such techniques is the same, but the reduction principle varies and can be reviewed in [116, 117].

An additional approach to identify similarities between samples and proteins/peptides is

given by unsupervised clustering algorithms, such as k-means, dbscan and hierarchical clustering [118]. Figure 1.7 D shows an exemplary heatmap with clustered rows. A good indication for a successful experiment is the clustering of the distinct perturbation groups, but similarly to PCA results, failing to achieve a reasonable clustering does not allow conclusions about the quality of the dataset. The clustering of peptides and proteins often aids the identification of functional groups, especially in time series experiments. The optimal clustering strategy has to be tested and evaluated, especially because data completeness is a prerequisite for most clustering algorithms.

The functional analysis of significantly differentially abundant peptides and proteins requires previous knowledge and a hypothesis with reduced number of involved proteins. In some cases, discovery like experiments do not have such knowledge at hand, thus techniques are required to extract previously unknown knowledge from the dataset. Typically, protein-protein networks already give the first insight into the underlying biology. A commonly used database for this analysis is STRING-DB [119]. There, a list of potentially interesting proteins (in most cases the significantly changing proteins) can be searched for annotated experimental or predicted evidence of interaction with each other or with other proteins. This way, a network of interaction is created, that potentially hints towards heavily involved proteins in the dataset, as shown in figure 1.7 E. An other important and widely used strategy for the biological interpretation of the dataset is the gene ontology (GO) enrichment analysis. For most proteins, annotations of information is available in databases such as their biological function, the cellular component, molecular function or involved pathways. When comparing the number of identified GO terms associated with the proteins that are significantly changing to the number that we would expect based on the whole proteome, enrichment of certain ontologies can be identified. To validate the findings, statistical testing ist performed and adjusted p-values are reported. STRING-DB offers GO enrichment functions with limited features, alternatives such as PANTHER [120] or the Cytoscape Plugin ClueGO [121] offer flexible and feature rich analysis options. Especially ClueGO offers the possibility to extract information from different database sources such as Reactome, WikiPathways and many more simultaneously and cluster the resulting terms by their functional meaning. The resulting GO cluster are combined into a network with overlapping genes / functions as shown in figure 1.7 F. Often already at first sight of the GO terms biological insights can be concluded and the corresponding

proteins can be extracted and validated in the processed data.

PTMs add extra level of information, that can be analyzed in specialized tools to gain insight about the identified modification. In case of phosphorylation, databases with previous knowledge aid the interpretation. Phosphositeplus [122] serves as curated database with detailed information about discovered phosphosites for human, mouse and rat (to some extend also cow, rabbit, chicken, hamster *et al.*). Identified phosphosites can be found in the database and downstream and upstream effects can be investigated along with the respective publication. Nevertheless, this is tedious work to do for all identified phosphosites and requires previous knowledge about the underlying biology by the analyst. A more unbiased and discovery approach is kinase and substrate enrichment analysis (KSEA). Similar to gene ontology enrichment analysis, databases are inquired to gain information about how many phosphosites would be expected and compared to the dataset at hand. Especially in the case of tyrosine kinases, often immediate downstream phosphosites can not be identified in the dataset. But an enrichment of further downside serine or threonine phosphosites can still indicate hyperactivity of the respective upstream or downstream kinase. A very simple and powerful KSEA tool is KSEA App [123], that can be accessed online and programmatic in R, which investigates the phosphositeplus database as well as NetworKIN, a database containing also predicted kinase-substrate relationships. A more advanced alternative is the integrative inferred kinase activity score (INKA score) [124], which takes more information into account such as information about kinase activating phosphorylation loops, downstream and upstream evidence and calculates a score, which indicates the kinase activity. This tool is especially suitable for clinical phosphoproteomics study, as it provides a scoring based on one sample only, whereas KSEA App requires calculated fold changes. The disadvantage of INKA score is the limited input formats, as it only accepts MaxQuant output as input. In addition to that, we found that it is not robust towards small changes from newer MaxQuant versions and does not provide proper error feedback. Additionally, the integration of proteomic change (as response to phosphorylation governed stimulus) and the phosphorylation status in the cell is key to understand causal relationships. A recently developed tool, CausalPath [125], allows to interrogate the dataset about causal and also conflicting relationships that are explained with the presented dataset. Unfortunately, so far this tool is only available for human -omics datasets. CausalPath accepts tailored input from various sources, takes the exact

phosphosite/protein relationship into account and the resulting causal pathways can be easily visualized and customized in Cytoscape. This makes it the ideal tool for discovery phosphoproteomics analysis from human proteome data.

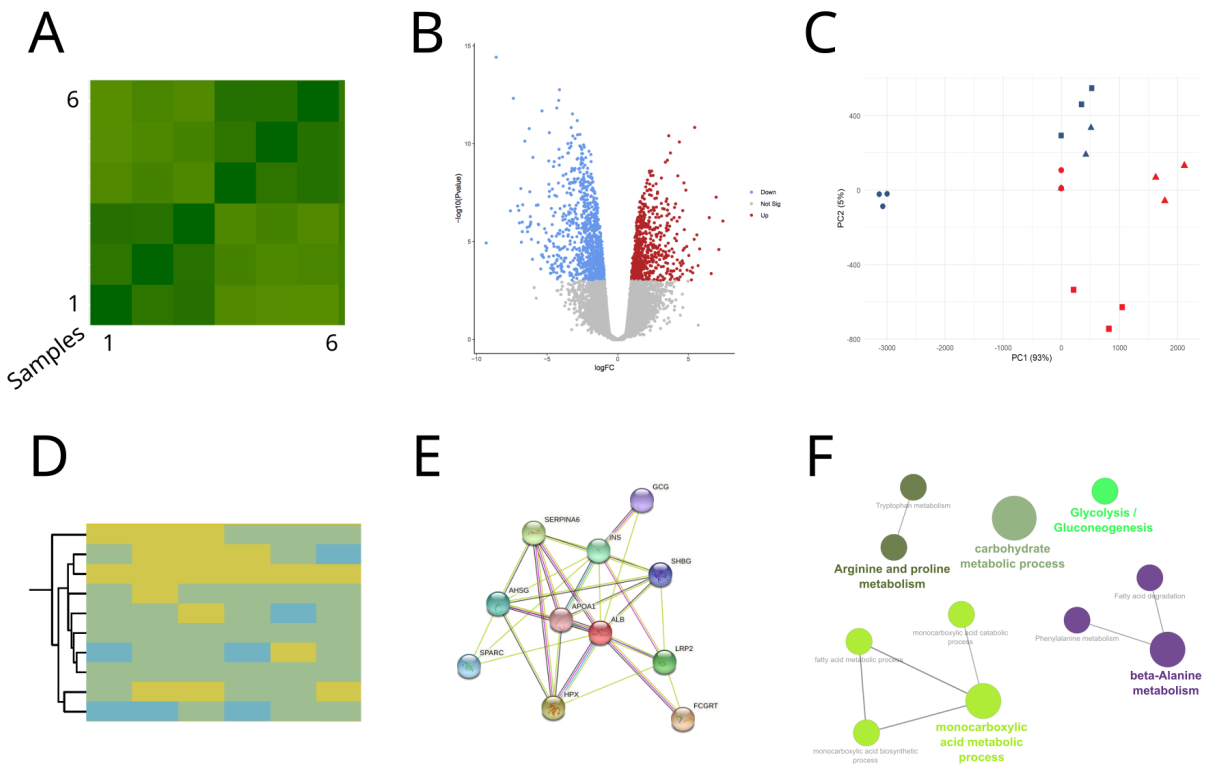

Figure 1.7: Overview of possible data visualization strategies A) correlation plot B) volcano plot C) principal component analysis D) clustered heatmap E) Protein-protein interaction network F) Gene ontology clustering by Cytoscape/ClueGO

2 Materials and methods

2.1 Sample overview

Sample #	Phenotype	Condition (KCl stress)	Bio Replicate	Type	LC-MS/MS	Inj.vol [µL]	Instrument	rawfile LC-MS Replicate 1
2022-003-01 TM	WT	control	1	Proteome	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	0.5	Orbitrap Exploris 480	V220319_04
2022-003-02 TM	WT	control	2	Proteome	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	0.5	Orbitrap Exploris 480	V220319_05
2022-003-03 TM	WT	control	3	Proteome	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	0.5	Orbitrap Exploris 480	V220319_06
2022-003-04 TM	WT	control	4	Proteome	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	0.5	Orbitrap Exploris 480	V220319_07
2022-003-05 TM	WT	10 min	1	Proteome	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	0.5	Orbitrap Exploris 480	V220319_08
2022-003-06 TM	WT	10 min	2	Proteome	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	0.5	Orbitrap Exploris 480	V220319_09
2022-003-07 TM	WT	10 min	3	Proteome	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	0.5	Orbitrap Exploris 480	V220319_10
2022-003-08 TM	WT	10 min	4	Proteome	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	0.5	Orbitrap Exploris 480	V220319_11
2022-003-09 TM	WT	60 min	1	Proteome	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	0.5	Orbitrap Exploris 480	V220319_12
2022-003-10 TM	WT	60 min	2	Proteome	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	0.5	Orbitrap Exploris 480	V220319_13
2022-003-11 TM	WT	60 min	3	Proteome	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	0.5	Orbitrap Exploris 480	V220319_14
2022-003-12 TM	WT	60 min	4	Proteome	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	0.5	Orbitrap Exploris 480	V220319_15
2022-003-13 TM	WT	240 min	1	Proteome	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	0.5	Orbitrap Exploris 480	V220319_16
2022-003-14 TM	WT	240 min	2	Proteome	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	0.5	Orbitrap Exploris 480	V220319_17
2022-003-15 TM	WT	240 min	3	Proteome	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	0.5	Orbitrap Exploris 480	V220319_18
2022-003-16 TM	WT	240 min	4	Proteome	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	0.5	Orbitrap Exploris 480	V220319_19
2022-003-17 TM	WT	1440 min	1	Proteome	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	0.5	Orbitrap Exploris 480	V220319_20
2022-003-18 TM	WT	1440 min	2	Proteome	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	0.5	Orbitrap Exploris 480	V220319_21
2022-003-19 TM	WT	1440 min	3	Proteome	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	0.5	Orbitrap Exploris 480	V220319_22
2022-003-20 TM	WT	1440 min	4	Proteome	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	0.5	Orbitrap Exploris 480	V220319_23
2022-003-21 TM	ΔHOG1	control	1	Proteome	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	0.5	Orbitrap Exploris 480	V220322_07
2022-003-22 TM	ΔHOG1	control	2	Proteome	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	0.5	Orbitrap Exploris 480	V220322_08
2022-003-23 TM	ΔHOG1	control	3	Proteome	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	0.5	Orbitrap Exploris 480	V220322_09
2022-003-24 TM	ΔHOG1	control	4	Proteome	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	0.5	Orbitrap Exploris 480	V220322_10
2022-003-25 TM	ΔHOG1	10 min	1	Proteome	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	0.5	Orbitrap Exploris 480	V220322_11
2022-003-26 TM	ΔHOG1	10 min	2	Proteome	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	0.5	Orbitrap Exploris 480	V220322_12
2022-003-27 TM	ΔHOG1	10 min	3	Proteome	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	0.5	Orbitrap Exploris 480	V220322_13
2022-003-28 TM	ΔHOG1	10 min	4	Proteome	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	0.5	Orbitrap Exploris 480	V220322_14
2022-003-29 TM	ΔHOG1	60 min	1	Proteome	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	0.5	Orbitrap Exploris 480	V220322_15
2022-003-30 TM	ΔHOG1	60 min	2	Proteome	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	0.5	Orbitrap Exploris 480	V220322_16
2022-003-31 TM	ΔHOG1	60 min	3	Proteome	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	0.5	Orbitrap Exploris 480	V220322_17
2022-003-32 TM	ΔHOG1	60 min	4	Proteome	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	0.5	Orbitrap Exploris 480	V220322_18
2022-003-33 TM	ΔHOG1	240 min	1	Proteome	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	0.5	Orbitrap Exploris 480	V220322_19
2022-003-34 TM	ΔHOG1	240 min	2	Proteome	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	0.5	Orbitrap Exploris 480	V220322_20
2022-003-35 TM	ΔHOG1	240 min	3	Proteome	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	0.5	Orbitrap Exploris 480	V220322_21
2022-003-36 TM	ΔHOG1	240 min	4	Proteome	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	0.5	Orbitrap Exploris 480	V220322_22
2022-003-37 TM	ΔHOG1	1440 min	1	Proteome	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	0.5	Orbitrap Exploris 480	V220322_23
2022-003-38 TM	ΔHOG1	1440 min	2	Proteome	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	0.5	Orbitrap Exploris 480	V220322_24
2022-003-39 TM	ΔHOG1	1440 min	3	Proteome	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	0.5	Orbitrap Exploris 480	V220322_25
2022-003-40 TM	ΔHOG1	1440 min	4	Proteome	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	0.5	Orbitrap Exploris 480	V220322_26

Table 2.1: Proteomics samples *M.oryzae* 1/2. Each sample type was measured in quadruplicates at 5 time points in DIA mode.

Sample #	Phenotype	Condition (KCl stress)	Bio Replicate	Type	LC-MS/MS	Inj.vol [µL]	Instrument	rawfile LC-MS Replicate 1
2022-003-41 TM	Reversible Adapted	control	1	Proteome	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	0.5	Orbitrap Exploris 480	V220322_30
2022-003-42 TM	Reversible Adapted	control	2	Proteome	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	0.5	Orbitrap Exploris 480	V220322_31
2022-003-43 TM	Reversible Adapted	control	3	Proteome	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	0.5	Orbitrap Exploris 480	V220322_32
2022-003-44 TM	Reversible Adapted	control	4	Proteome	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	0.5	Orbitrap Exploris 480	V220322_33
2022-003-45 TM	Reversible Adapted	10 min	1	Proteome	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	0.5	Orbitrap Exploris 480	V220322_34
2022-003-46 TM	Reversible Adapted	10 min	2	Proteome	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	0.5	Orbitrap Exploris 480	V220322_35
2022-003-47 TM	Reversible Adapted	10 min	3	Proteome	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	0.5	Orbitrap Exploris 480	V220322_36
2022-003-48 TM	Reversible Adapted	10 min	4	Proteome	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	0.5	Orbitrap Exploris 480	V220322_37
2022-003-49 TM	Reversible Adapted	60 min	1	Proteome	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	0.5	Orbitrap Exploris 480	V220322_38
2022-003-50 TM	Reversible Adapted	60 min	2	Proteome	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	0.5	Orbitrap Exploris 480	V220322_39
2022-003-51 TM	Reversible Adapted	60 min	3	Proteome	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	0.5	Orbitrap Exploris 480	V220322_40
2022-003-52 TM	Reversible Adapted	60 min	4	Proteome	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	0.5	Orbitrap Exploris 480	V220322_41
2022-003-53 TM	Reversible Adapted	240 min	1	Proteome	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	0.5	Orbitrap Exploris 480	V220322_42
2022-003-54 TM	Reversible Adapted	240 min	2	Proteome	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	0.5	Orbitrap Exploris 480	V220322_43
2022-003-55 TM	Reversible Adapted	240 min	3	Proteome	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	0.5	Orbitrap Exploris 480	V220322_44
2022-003-56 TM	Reversible Adapted	240 min	4	Proteome	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	0.5	Orbitrap Exploris 480	V220322_45
2022-003-57 TM	Reversible Adapted	1440 min	1	Proteome	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	0.5	Orbitrap Exploris 480	V220322_46
2022-003-58 TM	Reversible Adapted	1440 min	2	Proteome	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	0.5	Orbitrap Exploris 480	V220322_47
2022-003-59 TM	Reversible Adapted	1440 min	3	Proteome	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	0.5	Orbitrap Exploris 480	V220322_48
2022-003-60 TM	Reversible Adapted	1440 min	4	Proteome	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	0.5	Orbitrap Exploris 480	V220322_49
2022-003-61 TM	Stable adapted	control	1	Proteome	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	0.5	Orbitrap Exploris 480	V220322_53
2022-003-62 TM	Stable adapted	control	2	Proteome	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	0.5	Orbitrap Exploris 480	V220322_54
2022-003-63 TM	Stable adapted	control	3	Proteome	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	0.5	Orbitrap Exploris 480	V220322_55
2022-003-64 TM	Stable adapted	control	4	Proteome	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	0.5	Orbitrap Exploris 480	V220322_56
2022-003-65 TM	Stable adapted	10 min	1	Proteome	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	0.5	Orbitrap Exploris 480	V220322_57
2022-003-66 TM	Stable adapted	10 min	2	Proteome	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	0.5	Orbitrap Exploris 480	V220322_58
2022-003-67 TM	Stable adapted	10 min	3	Proteome	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	0.5	Orbitrap Exploris 480	V220322_59
2022-003-68 TM	Stable adapted	10 min	4	Proteome	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	0.5	Orbitrap Exploris 480	V220322_60
2022-003-69 TM	Stable adapted	60 min	1	Proteome	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	0.5	Orbitrap Exploris 480	V220322_61
2022-003-70 TM	Stable adapted	60 min	2	Proteome	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	0.5	Orbitrap Exploris 480	V220322_62
2022-003-71 TM	Stable adapted	60 min	3	Proteome	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	0.5	Orbitrap Exploris 480	V220322_63
2022-003-72 TM	Stable adapted	60 min	4	Proteome	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	0.5	Orbitrap Exploris 480	V220322_64
2022-003-73 TM	Stable adapted	240 min	1	Proteome	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	0.5	Orbitrap Exploris 480	V220322_65
2022-003-74 TM	Stable adapted	240 min	2	Proteome	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	0.5	Orbitrap Exploris 480	V220322_66
2022-003-75 TM	Stable adapted	240 min	3	Proteome	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	0.5	Orbitrap Exploris 480	V220322_67
2022-003-76 TM	Stable adapted	240 min	4	Proteome	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	0.5	Orbitrap Exploris 480	V220322_68
2022-003-77 TM	Stable adapted	1440 min	1	Proteome	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	0.5	Orbitrap Exploris 480	V220322_69
2022-003-78 TM	Stable adapted	1440 min	2	Proteome	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	0.5	Orbitrap Exploris 480	V220322_70
2022-003-79 TM	Stable adapted	1440 min	3	Proteome	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	0.5	Orbitrap Exploris 480	V220322_71
2022-003-80 TM	Stable adapted	1440 min	4	Proteome	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	0.5	Orbitrap Exploris 480	V220322_72

Table 2.2: Proteomics samples *M.oryzae* 2/2. Each sample type was measured in quadruplicates at 5 time points in DIA mode.

Sample #	Phenotype	Condition (KCl stress)	Bio Replicate	Type	LC-MS/MS	Inj.vol [µL]	Instrument	rawfile LC-MS Replicate 1
2022-003-81 TM	WT	control	1	Phospho	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	3	Orbitrap Exploris 480	V220325_08
2022-003-82 TM	WT	control	2	Phospho	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	3	Orbitrap Exploris 480	V220325_09
2022-003-83 TM	WT	control	3	Phospho	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	3	Orbitrap Exploris 480	V220325_10
2022-003-84 TM	WT	control	4	Phospho	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	3	Orbitrap Exploris 480	V220325_11
2022-003-85 TM	WT	10 min	1	Phospho	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	3	Orbitrap Exploris 480	V220325_12
2022-003-86 TM	WT	10 min	2	Phospho	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	3	Orbitrap Exploris 480	V220325_13
2022-003-87 TM	WT	10 min	3	Phospho	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	3	Orbitrap Exploris 480	V220325_14
2022-003-88 TM	WT	10 min	4	Phospho	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	3	Orbitrap Exploris 480	V220325_15
2022-003-89 TM	WT	60 min	1	Phospho	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	3	Orbitrap Exploris 480	V220325_16
2022-003-90 TM	WT	60 min	2	Phospho	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	3	Orbitrap Exploris 480	V220325_17
2022-003-91 TM	WT	60 min	3	Phospho	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	3	Orbitrap Exploris 480	V220325_18
2022-003-92 TM	WT	60 min	4	Phospho	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	3	Orbitrap Exploris 480	V220325_19
2022-003-93 TM	WT	240 min	1	Phospho	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	3	Orbitrap Exploris 480	V220325_20
2022-003-94 TM	WT	240 min	2	Phospho	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	3	Orbitrap Exploris 480	V220325_21
2022-003-95 TM	WT	240 min	3	Phospho	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	3	Orbitrap Exploris 480	V220325_22
2022-003-96 TM	WT	240 min	4	Phospho	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	3	Orbitrap Exploris 480	V220325_23
2022-003-97 TM	WT	1440 min	1	Phospho	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	3	Orbitrap Exploris 480	V220325_24
2022-003-98 TM	WT	1440 min	2	Phospho	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	3	Orbitrap Exploris 480	V220325_25
2022-003-99 TM	WT	1440 min	3	Phospho	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	3	Orbitrap Exploris 480	V220325_26
2022-003-100 TM	WT	1440 min	4	Phospho	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	3	Orbitrap Exploris 480	V220325_27
2022-003-101 TM	ΔHOG1	control	1	Phospho	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	3	Orbitrap Exploris 480	V220325_31
2022-003-102 TM	ΔHOG1	control	2	Phospho	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	3	Orbitrap Exploris 480	V220325_32
2022-003-103 TM	ΔHOG1	control	3	Phospho	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	3	Orbitrap Exploris 480	V220325_33
2022-003-104 TM	ΔHOG1	control	4	Phospho	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	3	Orbitrap Exploris 480	V220325_34
2022-003-105 TM	ΔHOG1	10 min	1	Phospho	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	3	Orbitrap Exploris 480	V220325_35
2022-003-106 TM	ΔHOG1	10 min	2	Phospho	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	3	Orbitrap Exploris 480	V220325_36
2022-003-107 TM	ΔHOG1	10 min	3	Phospho	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	3	Orbitrap Exploris 480	V220325_37
2022-003-108 TM	ΔHOG1	10 min	4	Phospho	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	3	Orbitrap Exploris 480	V220325_38
2022-003-109 TM	ΔHOG1	60 min	1	Phospho	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	3	Orbitrap Exploris 480	V220325_39
2022-003-110 TM	ΔHOG1	60 min	2	Phospho	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	3	Orbitrap Exploris 480	V220325_40
2022-003-111 TM	ΔHOG1	60 min	3	Phospho	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	3	Orbitrap Exploris 480	V220325_41
2022-003-112 TM	ΔHOG1	60 min	4	Phospho	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	3	Orbitrap Exploris 480	V220325_42
2022-003-113 TM	ΔHOG1	240 min	1	Phospho	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	3	Orbitrap Exploris 480	V220325_43
2022-003-114 TM	ΔHOG1	240 min	2	Phospho	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	3	Orbitrap Exploris 480	V220325_44
2022-003-115 TM	ΔHOG1	240 min	3	Phospho	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	3	Orbitrap Exploris 480	V220325_45
2022-003-116 TM	ΔHOG1	240 min	4	Phospho	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	3	Orbitrap Exploris 480	V220325_46
2022-003-117 TM	ΔHOG1	1440 min	1	Phospho	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	3	Orbitrap Exploris 480	V220325_47
2022-003-118 TM	ΔHOG1	1440 min	2	Phospho	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	3	Orbitrap Exploris 480	V220325_48
2022-003-119 TM	ΔHOG1	1440 min	3	Phospho	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	3	Orbitrap Exploris 480	V220325_49
2022-003-120 TM	ΔHOG1	1440 min	4	Phospho	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	3	Orbitrap Exploris 480	V220325_50

Table 2.3: Phosphoproteomics samples *M.oryzae* 1/2. Each sample type was measured in quadruplicates at 5 time points in DIA mode.

Sample #	Phenotype	Condition (KCl stress)	Bio Replicate	Type	LC-MS/MS	Inj.vol [µL]	Instrument	rawfile LC-MS Replicate 1
2022-003-121 TM	Reversible Adapted	control	1	Phospho	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	3	Orbitrap Exploris 480	V220325_54
2022-003-122 TM	Reversible Adapted	control	2	Phospho	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	3	Orbitrap Exploris 480	V220325_55
2022-003-123 TM	Reversible Adapted	control	3	Phospho	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	3	Orbitrap Exploris 480	V220325_56
2022-003-124 TM	Reversible Adapted	control	4	Phospho	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	3	Orbitrap Exploris 480	V220325_57
2022-003-125 TM	Reversible Adapted	10 min	1	Phospho	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	3	Orbitrap Exploris 480	V220325_58
2022-003-126 TM	Reversible Adapted	10 min	2	Phospho	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	3	Orbitrap Exploris 480	V220325_59
2022-003-127 TM	Reversible Adapted	10 min	3	Phospho	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	3	Orbitrap Exploris 480	V220325_60
2022-003-128 TM	Reversible Adapted	10 min	4	Phospho	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	3	Orbitrap Exploris 480	V220325_61
2022-003-129 TM	Reversible Adapted	60 min	1	Phospho	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	3	Orbitrap Exploris 480	V220325_62
2022-003-130 TM	Reversible Adapted	60 min	2	Phospho	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	3	Orbitrap Exploris 480	V220325_63
2022-003-131 TM	Reversible Adapted	60 min	3	Phospho	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	3	Orbitrap Exploris 480	V220325_64
2022-003-132 TM	Reversible Adapted	60 min	4	Phospho	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	3	Orbitrap Exploris 480	V220325_65
2022-003-133 TM	Reversible Adapted	240 min	1	Phospho	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	3	Orbitrap Exploris 480	V220325_66
2022-003-134 TM	Reversible Adapted	240 min	2	Phospho	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	3	Orbitrap Exploris 480	V220325_67
2022-003-135 TM	Reversible Adapted	240 min	3	Phospho	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	3	Orbitrap Exploris 480	V220325_68
2022-003-136 TM	Reversible Adapted	240 min	4	Phospho	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	3	Orbitrap Exploris 480	V220325_69
2022-003-137 TM	Reversible Adapted	1440 min	1	Phospho	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	3	Orbitrap Exploris 480	V220325_70
2022-003-138 TM	Reversible Adapted	1440 min	2	Phospho	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	3	Orbitrap Exploris 480	V220325_71
2022-003-139 TM	Reversible Adapted	1440 min	3	Phospho	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	3	Orbitrap Exploris 480	V220325_72
2022-003-140 TM	Reversible Adapted	1440 min	4	Phospho	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	3	Orbitrap Exploris 480	V220325_73
2022-003-141 TM	Stable adapted	control	1	Phospho	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	3	Orbitrap Exploris 480	V220325_77
2022-003-142 TM	Stable adapted	control	2	Phospho	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	3	Orbitrap Exploris 480	V220325_78
2022-003-143 TM	Stable adapted	control	3	Phospho	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	3	Orbitrap Exploris 480	V220325_79
2022-003-144 TM	Stable adapted	control	4	Phospho	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	3	Orbitrap Exploris 480	V220325_80
2022-003-145 TM	Stable adapted	10 min	1	Phospho	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	3	Orbitrap Exploris 480	V220325_81
2022-003-146 TM	Stable adapted	10 min	2	Phospho	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	3	Orbitrap Exploris 480	V220325_82
2022-003-147 TM	Stable adapted	10 min	3	Phospho	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	3	Orbitrap Exploris 480	V220325_83
2022-003-148 TM	Stable adapted	10 min	4	Phospho	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	3	Orbitrap Exploris 480	V220325_84
2022-003-149 TM	Stable adapted	60 min	1	Phospho	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	3	Orbitrap Exploris 480	V220325_85
2022-003-150 TM	Stable adapted	60 min	2	Phospho	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	3	Orbitrap Exploris 480	V220325_86
2022-003-151 TM	Stable adapted	60 min	3	Phospho	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	3	Orbitrap Exploris 480	V220325_87
2022-003-152 TM	Stable adapted	60 min	4	Phospho	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	3	Orbitrap Exploris 480	V220325_88
2022-003-153 TM	Stable adapted	240 min	1	Phospho	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	3	Orbitrap Exploris 480	V220325_89
2022-003-154 TM	Stable adapted	240 min	2	Phospho	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	3	Orbitrap Exploris 480	V220325_90
2022-003-155 TM	Stable adapted	240 min	3	Phospho	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	3	Orbitrap Exploris 480	V220325_91
2022-003-156 TM	Stable adapted	240 min	4	Phospho	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	3	Orbitrap Exploris 480	V220325_92
2022-003-157 TM	Stable adapted	1440 min	1	Phospho	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	3	Orbitrap Exploris 480	V220325_93
2022-003-158 TM	Stable adapted	1440 min	2	Phospho	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	3	Orbitrap Exploris 480	V220325_94
2022-003-159 TM	Stable adapted	1440 min	3	Phospho	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	3	Orbitrap Exploris 480	V220325_95
2022-003-160 TM	Stable adapted	1440 min	4	Phospho	20210604_DIA_21_window_345_1250_30K_60min_35B_trap_linear_3_55C	3	Orbitrap Exploris 480	V220325_96

Table 2.4: Phosphoproteomics samples *M.oryzae* 2/2. Each sample type was measured in quadruplicates at 5 time points in DIA mode.

Sample #	Condition	Bio Replicate	Type	LC-MS/MS	Inj.vol [µL]	Instrument	MS mode	rawfile
2021-237-01	control	1	Phospho	DI_20210617_linearstep_47mingdtBruker_IonOptics_37B_50C_Inj600Bar.m?HyStar_LC	2	timsTOF Pro 2	DIA	F211112_027
				20210608_proteomics_oC_6.2.4_timsTOF-SCP_lowSampleAmount.m?OtofImpacTEMControl				F211112_034
								F211112_041
2021-237-02	control	2	Phospho		2	timsTOF Pro 2	DIA	F211112_028
								F211112_035
								F211112_042
2021-237-03	control	3	Phospho		2	timsTOF Pro 2	DIA	F211112_029
								F211112_036
								F211112_043
2021-237-04	Ceritinib treated	1	Phospho		2	timsTOF Pro 2	DIA	F211112_030
								F211112_037
								F211112_044
2021-237-05	Ceritinib treated	2	Phospho		2	timsTOF Pro 2	DIA	F211112_031
								F211112_038
								F211112_045
2021-237-06	Ceritinib treated	3	Phospho		2	timsTOF Pro 2	DIA	F211112_032
								F211112_039
								F211112_046

Table 2.5: Phosphoproteomics samples human osteosarcoma cells. Control and Ceritinib treated cells were measured in triplicates in DIA mode.

Sample #	Condition	Bio Replicate	Type	LC-MS/MS	Inj.vol [µL]	Instrument	MS mode	rawfile
2021-236-01	control	1	Phospho	DI_20210617_linearstep_47mingdtBruker_IonOptics_37B_50C_Inj600Bar.m?HyStar_LC	2	timsTOF Pro 2	DIA	F211112_005
				20210608_proteomics_oC_6.2.4_timsTOF-SCP_lowSampleAmount.m?OtofImpacTEMControl				F211112_012
								F211112_019
2021-236-02	control	2	Phospho		2	timsTOF Pro 2	DIA	F211112_006
								F211112_013
								F211112_020
2021-236-03	control	3	Phospho		2	timsTOF Pro 2	DIA	F211112_007
								F211112_014
								F211112_021
2021-236-04	treated	1	Phospho		2	timsTOF Pro 2	DIA	F211112_008
								F211112_015
								F211112_022
2021-236-05	treated	2	Phospho		2	timsTOF Pro 2	DIA	F211112_009
								F211112_016
								F211112_023
2021-236-06	treated	3	Phospho		2	timsTOF Pro 2	DIA	F211112_010
								F211112_017
								F211112_024

Table 2.6: Phosphoproteomics samples murine Th17 cells. Control and treated cells were measured in triplicates in DIA mode.

2.2 Sample preparation

2.2.1 Cell lysis and protein digest

If not stated otherwise, all reagents were used in LC-MS/MS grade from common vendors, such as Carl Roth, Merck *et cetera*.

The sample preparation for all *Magnaporthe oryzae* samples has been performed as described in [26]. A sample aliquot of 200 mg lyophilized and grinded mycelium was suspended in 2500 µL boiling lysis buffer (5 % SDS / 5 mM DTT / 100 mM Tris–HCl adjusted to pH 8.5) , incubation at 90°C for 30 min (5 min heating / 5 min cooling at room temperature, vortexed samples in between). 160 µL alkylation reagent (200 mM iodoacetamide) and incubated with the sample for 30 min in the dark at room temperature. Afterwards, the samples were treatment with ultrasound (30 s on / 30s off cycles, high power mode) for 15 min in the Bioruptor (Diagenode, Belgium). After centrifugation of the samples, the supernatant was transferred to a 50 mL Eppendorf Protein LoBind conical tube where proteins were precipitated by addition of 10 mL methanol and 2.5 mL chloroform. The samples were vortexted, 7.5 mL of water was added, vortexed and the samples were centifuged at 4°C at 4600 rpm for one minute. The upper phase was discarded and 10 mL methanol was added. After vortexing and centrifugation with same conditions as before, the remaining liquid above the formed precipitate was discarded. The precipitated proteins were resuspended in 2500 µL urea buffer (7M urea / 2M thiourea / 100 mM ammonium bicarbonate) and an 10 µL aliquot was taken for protein quantification by Pierce 660 nm assay (Thermo Fisher, USA) according to the manufacturers instructions. The samples were diluted 1:4.44 with buffer (50 mM ammonium bicarbonate), remaning DNA/RNA was removed by addition 500 units benzonase and tryptic digest was performed over night at room temperature by addition TPCK-treated trypsin with an 1:25 ratio of trypsin to protein, according to the previously determined protein concentration. The next day, the digest was acidified by addition of trifluoroacetic acid (TFA) to a final concentration of 0.5% TFA.

For the following desalting procedure, always 2/3 of the cartridge volume available (4 mL) were applied. Waters Sep-Pak tC18 500 mg sorbent/6 mL volume were used. First, the cartridges were flushed with methanol, conditioned with 50 % ACN / 0.1% TFA and

equilibrated with 0.1% TFA in water. The samples were loaded as slowly as possible. After binding of the samples, the cartridges were washed five times with 0.1% TFA in water. The samples were eluted by addition of 4 mL 50 % ACN / 0.1% TFA. Taking into account the previously determinde protein amount in the samples, aliquots of the eluate were taken a) 20 µg for proteome analysis and b) 1000 µg for phosphopeptide enrichment. Both aliquots were lyophilized. The remaining aliquot has been stored at $-80°C$ for later usage as back-up.

The whole proteome analysis aliquot of 20 µg was resuspended in 40 µL of 0.1 % FA and the phosphopeptide aliquot was resuspended in 20 µL of 0.1 % FA for LC-MS analysis.

Human osteosarcoma cells and murine Th17 cells (both of various and unknown cell count) were lyzed in 50 µL boiling 1% SDS and the samples were treated with ultrasound with the same conditions as *M. oryzae* samples. Reduction and acetylation of free cysteines was performed by addition of DTT (final concentration of DTT in the sample: 10 mM) and incubation at $45°C$ for 30 min, followed by addition of IAA (final concentration of IAA in the sample: 40 mM) , DNA/RNA was digested by addition of 600 Units benzonase. After protein quantification by Pierce 660 nm Assay according to manufacturers instructions, an aliquot of 30 µg was subjected to an on-bead precipitation based tryptic digest (single pot solid phase sample preparation - SP3). For this, sample was added to 250 µg of 1:1 mixture magnetic beads with functional surface (Sera-Mag carboxylate-modified magnetic particles, hydrophobic and hydrophilic obtained from GE Healthcare, USA). Proteins were precipitated on beads by addition of acetonitrile (ACN) to achieve a final concentration of 80 % ACN. After 20 min incubation at room temperature while shaking at 800 rpm, the supernatant was removed and the precipitated proteins were washed two times with 200 µL of 80 % ACN and one time with pure ACN. The proteins were digested over night at $32°C$ by addition and incubation with 20 µL 50 mM ammonium bicarbonate buffer containing 1 µg trypsin. After digest, the supernatant was collected in a new tube and trifluoroacetic acid (TFA) was added to a final concentration of 1 % before desalting using SepPak tC18 µElution plate (Waters, USA). After desalting, an aliquot of 5 µg for later whole proteome analysis and 25 µg for later phosphopeptide enrichment were separated and lyophilized.

2.2.2 Phosphopeptide enrichment

Phosphopeptide enrichment for *Magnaporthe oryzae* samples was performed using 1000 µg peptide and commercially available spin-tips Titansphere Phos-TiO with 1 mg sorbent bed (GL Sciences, Japan) following the manufacturers instructions. Clean-up of the enriched phosphopeptides was performed using Pierce Graphite Spin Tips (Thermo Fisher Scientific, Waltham, MA USA) following the manufacturers instructions.

Phosphopeptide enrichment for Human osteosarcoma cells and murine Th17 cells was performed using Zr-IMAC high performance beads (MagReSyn, SA). As described in the following Results and Discussion, method parameters were optimized for low amount phosphopeptide enrichment, resulting in the following procedure.

25 µg lyophilized peptides were resuspended in 100 µL *Loading Buffer* (80 % ACN, 5 % TFA, 0.5 M glycolic acid) in a 2 mL Eppendorf tube. 62.5 µg Zr-IMAC beads (*i.e.* ratio 1:2.5 peptide:beads) were prepared by binding the beads on a magnet rack, discarding the supernatant. The beads were then washed with 100 µL *Loading Buffer*, collection on magnet, discarding the supernatant. Resuspended peptides were added to the washed beads and incubated at 40°*C* for 30 min while shaking at 800 rpm. After incubation and bead capture on magnets, the supernatant can be collected for proteome measurement. The bound phosphopeptides on the magnetic beads are further washed with 50 µL *Wash Buffer 1* (80 % ACN, 1 % TFA) and *Wash Buffer 2* (10 % ACN, 0.2 % TFA) at 40°*C* for 10 min each, shaking at 800 rpm and combining the supernatant with previous supernatant for proteome analysis. The bound phosphopeptides are eluted by incubation with 25 µL *Elution Buffer* (1 % NH_4OH) at room temperature for 15 min, which is repeated to obtain a final elution volume of 50 µL. The high pH of the eluted phosphopeptides are immediately neutralized by adding the elution directly into previously prepared 15 µL of 10 % formic acid (FA). After lyophilization, the phosphopeptides are resuspended in 25 µL of 0.1 % FA for subsequent LC-MS/MS analysis.

2.2.3 ERLIC chromatography parameters

The used chromatographic parameters were first published in [126] and modified as indicated in the following. In this example, 2 mg of lyophilized mousebrain tryptic peptides were prepared following the same protocol as for the *Magnaporthe oryzae* samples described earlier. The peptides were resuspended in 1 mL of Eluent A (20 mM $Na-MePO_4$, pH 2.0, 70 % ACN). 0.45 µm syringe filters were conditioned by flushing with at least 1 mL of Eluent A. The sample was completely filtered through the preconditioned syringe filters and an appropriate volume of Eluent A was applied after the sample to reach 1 mL final syringe eluate volume. The filtered sample was applied with a sample loop to an Äkta pure 20 HPLC system (GE Healthcare, Chicago, USA). The peptides were separated at 16 °C and fractions of 1 mL were collected over a gradient of 30 min starting with 100 % Eluent A to 100 % Eluent B (200 mM triethylammonium phosphate, pH 2.0, 60 % ACN), followed by 15 min of equilibration with Eluent A, that was not part of the collected fractions. The collected fractions were lyophilized and each desalted using Waters Sep-Pak tC18 500 mg sorbent cartridges with the procedurce described earlier. The desalted peptides were lyophilized again and resuspended in 20 µL of 0.1 % FA for LC-MS analysis.

The first variation of the chromatographic conditions was the use of magnesium hydroxide solution instead of the usually used sodium hydroxide solutions for the pH adjustment of each Eluent, offering a different counter-cation for the chromatography mode. For the second variation, in addition to the changes from the first variation, was the use of a convex gradient. The gradient was designed as shown in 2.7.

Time [min]	% Eluent A
0	100
5	80
10	60
15	40
20	30
25	20
30	10
35	0
35	0
36	100
50	100

Table 2.7: Gradient conditions for a convex gradient in ERLIC in contrast to the linear gradient used in previous publications such as [126]

2.3 Peptide identification

2.3.1 LC-MS/MS of *M.oryzae* as resource for osmostress signaling research

0.5 µL of the reconstituted peptides for whole proteome analysis or 2 µL of the reconstituted phosphopeptides were separated on an Ultimate 3000 nanoUPLC (Thermo Scientific, Waltham, USA) with 300 nL/min by a reversed phase C18 column (HSS-T3 C18 1.8 µm, 75 µm × 250 mm, Waters Corporation) at 55°C using a 45 min linear gradient from 95 % Eluent A (0.1 % TFA, 3 % DMSO in water) to 35 % Eluent B (0.1 % TFA, 3 % DMSO in ACN) followed by ionization in positive mode using a Nanospray Flex electrospray ionization source (Thermo Scientific). Mass-to-charge analysis of the eluting peptides was performed using an Orbitrap Exploris 480 (Thermo Scientific) in data independent acquisition (DIA) mode. MS1 scans were acquired with a resolution of 120 000 @ 200 m/z in for a range of 345 - 1250 m/z. RF lens was set to 40 % and AGC target to 300 % (*i.e.* corresponding to 3x 10^6 charges). DIA MS2 scans were acquired with a resolution of 30 000 @ 200 m/z with a variable window scheme as shown in supplementary table 6.1. The normalized collision energy was set to 27 %, RF lens to 40 % and AGC target to 1000 % (*i.e.* corresponding to 10x 10^6 charges).

2.3.2 LC-MS/MS of *M.oryzae* in DDA for comparison to DIA

2 µL of the reconstituted phosphopeptides were separated on an Ultimate 3000 nanoUPLC (Thermo Scientific) with 300 nL/min by a reversed phase C18 column (HSS-T3 C18 1.8 µm, 75 µm × 250 mm, Waters Corporation) at 55°C using a 45 min linear gradient from 95 % Eluent A (0.1 % TFA / 3 % DMSO / Water) to 35 % Eluent B (0.1 % TFA / 3 % DMSO / ACN) followed by ionization using a Nanospray Flex electrospray ionization source (Thermo Scientific). All samples were measured in triplicates. Mass-to-charge analysis of the eluting peptides was performed using an Orbitrap Exploris 480 (Thermo Scientific) in data dependent acquisition (DDA) mode. Full scan MS1 spectra were collected over a range of 350 - 1600 m/z with a mass resolution of 60 000 @ 200 m/z using an automatic gain control (AGC) target of 100 %, maximum injection time was set to "Auto" and RF lens to 40 %. Within a fixed cycle time of 1.5 s the most intense peaks (number of peaks selected is automatically determined by the instrument) above the signal threshold of 2 x10^4 , harboring a charge of 2 - 6, were selected within an isolation window of 1.4 Da as precursors for fragmentation using higher energy collisional dissociation (HCD) with normalized collision energy of 30. The resulting fragment ion m/z ratios were measured as MS2 spectra over a automatically selected m/z range with a mass resolution of 15 000 @ 200 m/z, AGC target was set to "Standard" and maximum injection time to "Auto".

2.3.3 LC-MS/MS of *M.oryzae* in DIA for comparison to DDA

3 µL of the reconstituted phosphopeptides were separated on a nanoElute LC system (Bruker Corporation, USA) at 400 nL/min using a reversed phase C18 column (Aurora UHPLC emitter column, 25 cm x 75 µm 1.6 µm, IonOpticks) which was heated to 50°C. Peptides were loaded onto the column in direct injection mode at 600 bar. Mobile phase A was 0.1 % FA (v/v) in water and mobile phase B 0.1 % FA (v/v) in ACN. Peptides were separated running a linear gradient from 2 % to 37 % mobile phase B over 39 min. Eluting peptides were analyzed in positive mode ESI-MS using parallel accumulation serial fragmentation (PASEF) enhanced data-independent acquisition mode (DIA) on a timsTOF Pro 2 mass spectrometer (Bruker Corporation). The dual tims was operated at a fixed duty cycle close to 100 % using equal accumulation and ramp times of 100 ms each spanning a mobility range from $1/K_0 = 0.6$ Vs cm^{-2} to 1.6 Vs cm^{-2}. We defined 36×

25 Th isolation windows from m/z 300 to 1165 resulting in fifteen diaPASEF scans per acquisition cycle. The collision energy was ramped linearly as a function of the mobility from 59 eV at $1/K_0$ = 1.3 Vs cm^{-2} to 20 eV at $1/K_0$ = 0.85 Vs cm^{-2}

2.3.4 LC-MS/MS of HOS / Th17

3 µL of the reconstituted phosphopeptides were separated on a nanoElute LC system (Bruker Corporation, USA) at 400 nL/min using a reversed phase C18 column (Aurora UHPLC emitter column, 25 cm x 75 µm 1.6 µm, IonOpticks) which was heated to 50°C. Peptides were loaded onto the column in direct injection mode at 600 bar. Mobile phase A was 0.1 % FA (v/v) in water and mobile phase B 0.1 % FA (v/v) in ACN. Peptides were separated running a linear gradient from 2 % to 37 % mobile phase B over 39 min. Eluting peptides were analyzed in positive mode ESI-MS using parallel accumulation serial fragmentation (PASEF) enhanced data-independent acquisition mode (DIA) on a timsTOF SCP mass spectrometer (Bruker Corporation). The dual tims was operated at a fixed duty cycle close to 100 % using equal accumulation and ramp times of 166 ms each spanning a mobility range from $1/K_0$ = 0.7 Vs cm^{-2} to 1.3 Vs cm^{-2}. We defined 29× 25 Th isolation windows from m/z 280 to 990 resulting in ten diaPASEF scans per acquisition cycle. The collision energy was ramped linearly as a function of the mobility from 59 eV at $1/K_0$ = 1.6 Vs cm^{-2} to 20 eV at $1/K_0$ = 0.6 Vs cm^{-2}.

2.3.5 Data processing parameters

Peptides measured in DIA mode were identified and label-free quantification (LFQ) of proteins was performed using DIA-NN (v1.8).

Full proteome samples from *M.oryzae* were processed using library free mode with standard parameters, except for tryptic cleavage sites considering no cleavage before proline. The FASTA protein database contained 12 790 protein entries of the *M.oryzae* reference proteome and 172 common contaminant proteins and was obtained on 01^{st} September 2021 from Uniprot.

For phosphopeptide analysis of either species, *M.oryzae*, mouse and human, a phosphopeptide spectral library was predicted *in-silico* using the built-in library free prediction algorithm provided by DIA-NN. For *M.oryzae*, the aforementioned FASTA database was

used as basis, for mouse a FASTA database was downloaded from uniprot.org on 09^{th} November 2021 containing 17 082 reviewed proteins to which 172 common contaminant proteins were added. The human FASTA database was downloaded on 03^{rd} March 2021 and included 20 365 reviewed protein entries to which the 172 common contaminant proteins were added. The spectra libraries were predicted with the precursor charge range set between 1 - 4 and the range for fragment ions and precursor mass to charge ratio was limited to 250 - 1250 m/z. The peptide length was set to 7 - 30. Tryptic cleavage considering no cleavage after the lysine or arginine is followed by proline, maximum one missed cleavage was allowed. N-terminal methionine excision was enabled and cysteine carbamidomethylation was set as fixed modification. The maximum number of variable modifications was set to 3, allowing only UniMod:21 modifications, *i.e.* mass delta of 79.9663 corresponding to phosphorylation at serine, threonine and tyrosine. The generated spectral libraries were used for follow up identification and quantification in DIA-NN using the standard settings.

The DDA rawfiles were processed by PEAKS X Pro (BSI, Canada) using the FASTA file described above, precursor tolerance and fragment ion tolerance were set to 15 ppm and 0.03 Da respectively, two missed cleavages were allowed, carbamidomethylation at cysteins was set as fixed modification while oxidation on methionine and phosphorylation on serine, threonine and tyrosine were set as variable modifications with a maximum of 4 variable modifications per peptide

2.3.6 Availability of raw files and R code

All raw files, DIA-NN settings and output files as well as R codes have been uploaded via JPOST [127] to be retrievable at proteomeXchange [128].

Data for the *M.oryzae* osmostress resource are available via the identifier PXD034481. The dataset can be accessed via

https://repository.jpostdb.org/preview/179221543262a4a1ccb2393

using the access key 3542.

Data for the DDA/DIA comparison are available via the identifier PXD038605 The dataset can be accessed via

https://repository.jpostdb.org/preview/13118931046394b6ae69c06

using the access key 9526.

3 Results and discussion

3.1 Improved sample preparation and measurement

3.1.1 Optimized cell lysis

In large scale experiments such as required for the adaption investigation in *M.oryzae*, reproducibility is key to ensure the correct identification over a large sample batch and over time inter-experiment, but also intra-experiment caused by time intensive data acquisition. Mainly, the robustness if governed by the variability introduced during sample handling, where working with small quantities or large volumes is undesirable due to unspecific binding to plasticware *et al.* Therefore, an efficient lysis and protein stabilization/solubilization strategy for *M.oryzae* is of utmost importance. In contrast to tissues or mammalian cell culture samples, where cells are fragile and easily lysable by chemical and gentle physical treatment (if at all), fungi and plants often need harsh lysis conditions to access the proteins from the cells, due to their tough cell wall structre [129]. Three parameters were investigated experimentally: 1) The protein yield after lysis, assessed by relating the measured protein amount, after lysis and clean up, to the crude sample weight. 2) The lysis volume and relation to the crude weight and 3) Inhibition of unspecific proteolysis.

The assessment of protein yield in dependence of the chemical (chaotropic and denaturating agents) and physical (sonication, bead beating and heat) treatment is summarized in figure 3.1. A) shows the protein yield from mouse brain tissue as control example. The mouse brains were homogenized under liquid nitrogen before weight into tubes for lysis. The combination of Urea/Thiourea containing lysis buffer with ultrasound treatment in the Diagenode Bioruptor yields 5.5 % median protein relative to the crude weight, that increases in presence of SDS as detergent to a median of 8.5 %, although due to the need to remove the detergent precipitation by $CHCl_3$/MeOH is required. When changing from Urea/Thiourea to DTT as chemical treatment, the yield is around 12 %. In contrast to that, a simple Urea/Thiourea treatment yields only a median of 0.5 % protein yield in *M.oryzae* as shown in figure 3.1 B . Bead beating as physical treatment yields a slightly higher protein yield of 0.75 %, but this treatment is known to introduce undesirable heat to the sample, which can cause carbamidomethylation. Although bead beating systems are available with cooling option, the marginal increase in yield is in no relation to the potential increase in chemical variability that is further propagated in

downstream phosphopeptide enrichment. Nevertheless, as cell wall disruption is believed to be a major issue preventing higher protein yield, bead beating as most harsh physical treatment yielding more protein underlines this hypothesis. The use of an alternative chaotropic agent, guanidine hydrochloride (GuHCl), did yield a even less protein lysis efficiency with a yield of around 0.25 % with apparently higher reproducibility. The lower denaturation strength is in line with previously published results [130], and the use of higher lysis temperature of 95 °C could not compensate for this effect. By introducing SDS to the Urea/Thiourea buffer, no increase could be observed, which is also true for the use of NaDoc alone in combination with heat. Interestingly, either the combination of Urea/Thiourea and CHAPS as detergent as well as the combination of DTT, SDS and heating lead to a significant increase in yield to a median of around 1 %. The hypothesis is, that Urea/Thiourea and SDS act in similar way as solubilizing and denaturating agents and thus do not show synergetic effects. In contrast to SDS, CHAPS is known as non-denaturating has a substantially different three-dimensional structure and being zwitterionic also physicochemical character [130]. Thus, the affected solubilized area while using CHAPS is different, while at the same time having a stiff structure which makes the kinetics of lysis and solubilzation putatively slower compared to the flexible dodecyl-chain of SDS. In addition to that, heat can not be applied due to the side reactions of Urea/Thiourea and the use of Urea/Thiourea with CHAPS seems to provide higher CVs, although this finding might be intrigued by the higher number of experiments (N=12) compared to DTT/SDS (N=6). DTT and SDS consequently show a similar disruptive denturating strength. Presumably the kinetics of both, the reduction of disulfide bonds by DTT and the solubilization and denaturation by SDS, are more similar than the combination of Urea/Thiourea and SDS and thus yield higher efficient lysis, especially as high temperature is possible with this lysis buffer type.

It is known, that fungi often display a high unspecific proteolytic activity [131]. As the downstream bioinformatic analysis requires peptides to be from tryptic digest origin (N-terminal R and K) for systematic *in-silico* digest or spectra prediction with successful and efficient database search, a unspecific proteolytic activity before tryptic digest will decrease reproducibility and identification efficiency dramatically. Consequently, inhibition of such activity is required. While in western blotting the chemical inhibition by commercially available reagent cocktails is commonly applied, this is contraproductive in

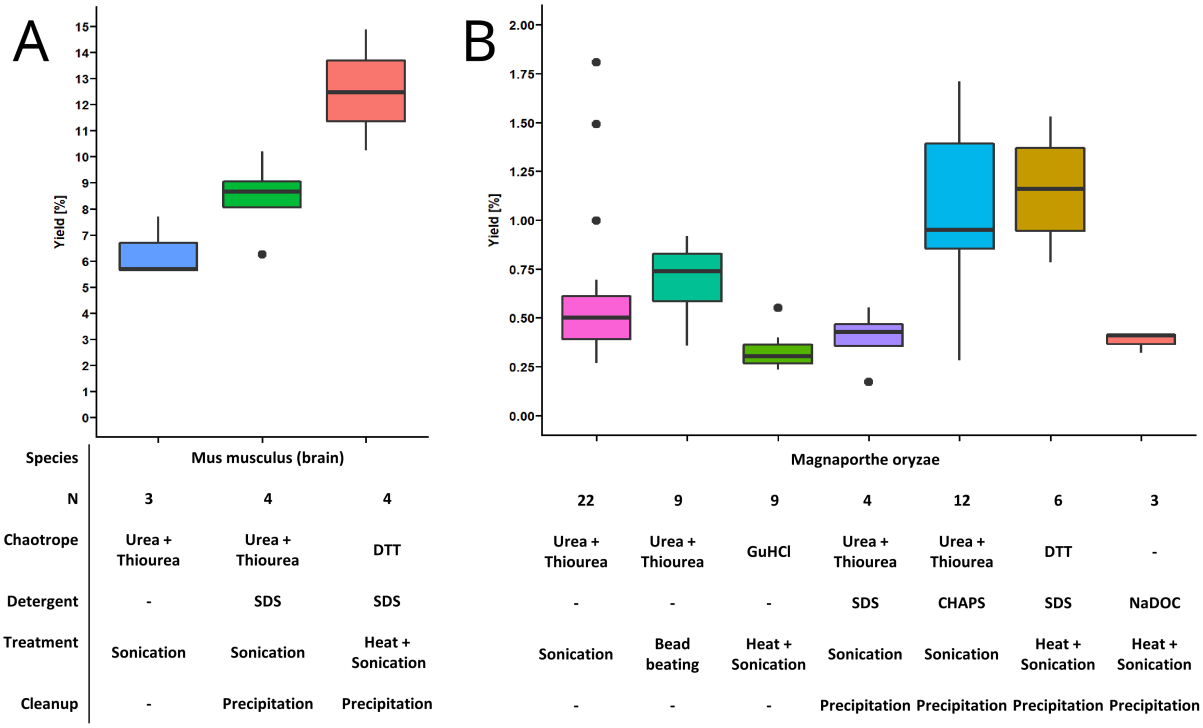

Figure 3.1: Overview of different lysis buffer strategies for A) mouse brain as control and B) *Magnaporthe oryzae*. The yield was calculated by measurement of protein content in the lysate by Pierce 660nm Assay compared to the crude sample weight that was used for lysis. While the least effecitve lysis strategy for the mouse brain samples still yields around 6 % protein amount of the used crude sample, the most effective lysis strategy for the fungal samples never exceeds 2 % protein amount of the sample weight.

proteomics research, as tryptic digest still has to be possible. For this analytical problem, specially developed cocktails (*e.g.* without EDTA) can be used, but do not provide full protection against unspecific proteolysis. On the other hand, heat as a simple denaturating technique can serve as inhibitor of proteolytic enzymes right from the beginning of the sample preparation. A measure of proteolytic activity is the unspecific search of peptides from DDA experiments. By calculating the ratio of tryptic peptides to the total number of identified peptides (including those from unspecific cleavage) an estimation of pre-tryptic unspecific proteolytic acivity can be made. From figure 3.2 A is obvious, that mouse brain tissue samples that are not expected to display high proteolytic activity do accordingly show a similar ratio of tryptic peptides when either using inhibitory cocktail or heat. Thus, in this case the treatment inhibits this activity ito a similar degree. In case of *S.bayanus*, a yeast strain that is available in out laboratory, shows a better inhibitory effect of heat compared the inhibitor cocktail. The reason for this can be explained by the partial activity of metal-proteases that usually are inhibited by the contained EDTA,

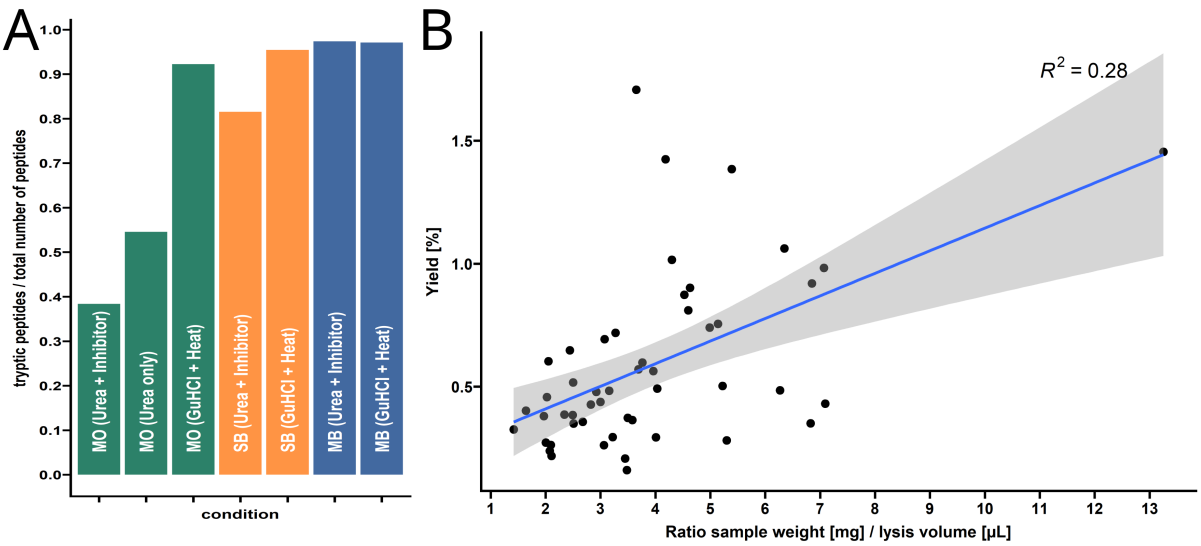

Figure 3.2: A) Effects of heat treatment to unspecific protease activity (MO: *Magnaporthe oryzae* in green / SB: *Saccharomyces bayanus* in orange / MB: Mouse brain from *Mus musculus* in blue). Fungal and murine samples were lysed according to the named treatment, tryptically digested and peptides mesaured by LC-MS. The mass spectra were serched with unspecific cleavage allowed. A high number of unspecifically cleaved peptides indicates a high protease activity before sample preparation. Based on the observed results, heat treatment serves as effective measure to inhibit unspecific protease activity B) The effect of the ratio between sample weight and lysis volume on protein yield of *Magnaporthe oryzae*. Samples were weight with very variable amounts, due to the inhomogeneous nature of the grinded mycellium. A very weak correlation could be observed, with no clear conclusion possible. An arbitraty ratio of sample weight to lysis buffer volume of 1:4 has been used for further experiments.

which can not be used in classical bottom-up proteomics. Interestingly, the use of inhibitor cocktails seems to have an even deceasing effect for the efficiency of the tryptic digest in *M.oryzae*, as the ratio of tryptic peptides is even lower compared to a lysis buffer without inhibitor cocktail. This effect can be explained by the lower concentration of trypsin that is used during the digest compared to mouse brain. Presumably, the inhibitory effect is also governed by a inhibiton kinetic on trypsin, that is not metal related. Therefore, a much higher relative activity in the mouse brain samples lead to an efficient digest before it is inhibited while the kinetics of the tryptic digest are not fast enough to compensate the inhibitory effect of the cocktail in the *M.oryzae* samples. Strikingly, by using heat we could achieve a similar efficiency ratio of tryptic peptides compared to the other sample types. Consequently, a lysis buffer that is compatible with heat is most desirable and with the previous results the combination of heated DTT and SDS provide a satisfactory solution to this problem, although the required protein precipitation as clean up intro-

duces a new source of variability. In respect of the used lysis volume, we obeserved a very weak correlation of the ratio between sample weight and lysis volume versus the measured protein yield as shown in figure 3.2 B. Thus, no clear conclusion could be drawn and a ratio of sample weight to lysis buffer of at least 1:4 has been used.

Protease and phosphatase inhibitors are widely applied in proteomics to preserve the status quo during sample preparation. Not only we demonstrate in our dataset a negative effect of protease inhibitiors on the proteolytic digest, but both phosphatase and protease inhibitors have been shown to lead to inefficient phosphopeptide enrichment or total depletion of phosphopeptides [62]. The reason for this is yet unknown and elucidation of mechanisms will require extensive work due to the vast number of inhibitory reagents. Furthermore, the sole use of phosphatase inhibitors without applying kinase inhibitors at the same time might lead to false positive hits, but kinase inhibitors are less common to be applied during cell lysis. Thus, enzyme denaturation by using heat is an excellent solution to circumvent proteolysis and enrichment issues right from the beginning.

In conclusion, cell wall disruption does not seem to be a problem with *M.oryzae*, rather protein solubilization (or protein content in crude weight) and protease activity, where a combination of DTT / SDS and heat treatment provide a satisfactory solution. Nevertheless, ultrasound is a valuable element during sample preparation as it potentially disrupts chromatin proteins which are critical contaminants in phosphopeptide enrichment. Here, optimization potential is given with newer instruments that work in high-throughput 96-well format. The issue hereby is the minimal sample volumes of less than 200 µL. Due to the low protein concentrations at hand, additional preparation steps such as buffer reduction by MWCO filters are required, that potentially loose a low mass subproteome and might be time intensive due to the complex matrix, if it is possible at all. Furthermore, due to the large lysis volume and low protein concentrations, unspecific binding to plasticware might become an issue.

3.1.2 Advancements in phosphopeptide enrichment methods

Two major challenges of phosphopeptide enrichment are a) the large amount of starting material required and b) low reproducibility and bias towards phosphopeptide subpopulations introduced by the enrichment step. Therefore, alternative strategies for a) downscal-

ing the phosphopeptide enrichment by using magnetic beads have been evaluated and b) a feasibility study for an alternative chromatography mechanism for potentially replacing the reversed phase chromatography was conducted.

For the successful downscaling of the phosphopeptide enrichment the most essential parameter is the ratio of functional groups per peptides. When the number of potential binding sites exceeds the number of phosphopeptides present in the sample, the remaining sites are consequently occupied by unspecifically binding peptides and thus, the enrichment efficiency and subsequently ionization efficiency decreases. On the other hand, less binding sites than phosphopeptides will lead to incomplete and unreproducible enrichment. Although strategies have been developed to adjust the amount of TiO_2 material onto tips (*i.e.* STAGE tips [132]), the manual assembly of such tips suffers from high variability and low shelf-life, as TiO_2 is hygroscopic. Recently, magnetic beads became popular as they offer a better and more reproducible scalability, fast and easy handling also in high throughput format (which is not easily feasible with tips) and provide better intermediary and inter-laboratory reproducibility due to their prolonged shelf-life [24].

Figure 3.3 A shows an overview of tested bead types with the respective peptide loading amount compared to the common TiO_2 spin tip performance. All samples were brain tissue from mouse, after digest and enrichment resuspended in 20 µL, 2 µL injected for 120 min LC runtime (90 min gradient) measured on the Orbitrap Exploris 480 in DDA mode. The TiO_2 and Ti^{4+}-IMAC beads show similar or moderately better performance in number of identified phosphopeptides compared to the TiO_2 spin tips, while reducing the peptide amount to 500 µg. In the case of TiO_2 beads, the handling of the beads stock solution was impaired due to beads aggregation and clogging. Thus, also one replicate did not show any peptides identified, which can be explained by inhomogeneous bead distribution during sample preparation. Interestingly, the enrichment efficiency is worse in bead type enrichment compared to tip type enrichment. A possible reason or this might be differences in the particulate structure and non-functional surfaces on both materials. While the spin tips are filled with pure TiO_2 particles of inhomogeneous and uncontrolled particle sizes, magnetic beads bind the functional groups on their surface (either covalently for MOAC or as complex for IMAC). Therefore, matrix structures or incomplete loading onto the beads surface might enable interactions of the sample matrix (salts, detergents,

lipids, buffer *et cetera*) with the surface and/or the phosphopeptides.

Even when decreasing the starting amount to 250 µg, a competitive performance of Ti^{4+}IMAC beads compared to TiO_2 spin tips has been observed, although with consequently lower enrichment efficiency. Interestingly, Zr^{4+}-IMAC beads provide superior performance compared to Ti^{4+}-IMAC beads using equal starting amount and superior performance compared to the TiO_2 spin tips while using just $^1/_4$ of peptide amount required. In addition to that, an increased enrichment efficiency has been observed for Zr^{4+}-IMAC beads. The reason for this observation is yet unclear: Is the superior identification performance due to reduced ionization suppression or is the total number of enriched peptides increased? The first would mean that Ti^{4+}-IMAC beads might provide a similar enrichment performance, that is suppressed by a greater extend of unspecific co-enrichment, while the latter would indicate a true performance difference, possibly caused by stronger affinity or altered retention mechanism. A possible way to address this question would be the enrichment of an artificial peptidome that consists of a sufficient number of known amounts of peptides, that have to be enriched without possible contaminants and with certain spike-in levels of contaminants. Supposedly high impact contaminants are detergents that are not completely cleaned up from the sample lysis such as SDS and CHAPS that have the potential to decrease the surface load of functional material on the beads or shield the surface of the beads towards the sample matrix. Second, salts with higher oxidative state such as magnesia might compete with the chelated metal ions, which in turn also would decrease the functional load on the magnetic beads. This way, the source of such difference can be evaluated, but has not been done yet due to time and resource constraints.

A possible reason for the difference in performance is an increased stability of the Zr^{4+} ion and bead matrix complex. It has been shown, that the cation-dipole interaction of complexes is a variant of the ion-ion attraction force which can be described with the Coulomb equation [133]. In both cases, the strength of the force is antiproportional to the square of the distance between ion/dipol and the other ion. As Zr^{4+} has a greater atom size compared to Ti^{4+}, the three-dimensional space within the binding site of the bead matrix is occupied to a greater extend and thus the distance between the Zr^{4+}-Ion and the chelating sites must be decreased. In consequence, the attraction force is increased and Zr^{4+}-Ions form a more stable complex. This way, the functional surface

remains unaffected with higher contaminant load of either source (detergent or salts). To our knowledge, this hypothesis has not been verified yet experimentally. Furthermore, it is known that the strength of a covalent (and non-covalent) bond is dependent on the overlap and the resulting hybridization in covalent binding [134]. Due to its atom size, the Zr^{4+}-Ion offers larger and more flexible orbitals for overlap with either the immobilization matrix or the phosphopeptide. In consequence the chances fot overlap and the overlap size is increased with Zr^{4+}-Ions in comparison to Ti^{4+}-Ions. These hypotheses are in line with the observation, that Zr^{4+}-IMAC beads offer a higher shelf-life compared to Ti^{4+}-IMAC. Thus, in all optimization experiments, Zr^{4+}-IMAC were used unless stated otherwise.

Benchmarking the downscaling of peptide amount was performed using the Ti^{4+}-IMAC beads in triplicates with equal conditions as the previous enrichment of 250 µg and is shown in figure 3.3 B. Interestingly, down to 100 µg the decrease in phosphopeptide IDs is still not significant, whereas a major decrease of phosphopeptides can be observed below 100 µg. But even with as low as 25 µg of starting peptide amount a reasonable number of phosphopeptides could be measured. Therefore, 25 µg were set as target amount for the downscaling optimization. Figure 3.4 shows an overview of the seven parameters that were addressed in the optimization. The resulting number of identified phosphopeptides, enrichment efficiency and reproducibility were selected as performance indicators of each optimization step.

The sample sets were analyzed on different LC-MS/MS, based on the availability to ensure short turn-over time. Thus, the absolute numbers of peptides can not be compared directly, but rather within the experiments only. Nevertheless, the bioinformatic processing and downstream analysis of the raw files was performed in FragPipe and R, equal for all samples. In general, except for the first experiment, the identified numbers of peptides are consistently low and suffer from high variance. This effect correlates with the amount of previously measured plasma samples. Contaminating agents from the plasma sample preparation or carry-over of peptides from previous runs might alter the retention behavior of phosphopeptides by shielding the chromatographic surface for the anyway weak interaction with phosphopeptides. In addition to that, carry-over unmodified peptides will reduce the ionization efficiency of phosphopeptides. Furthermore, such peptides can also enhance the solubility of trace metals in the chromatographic system

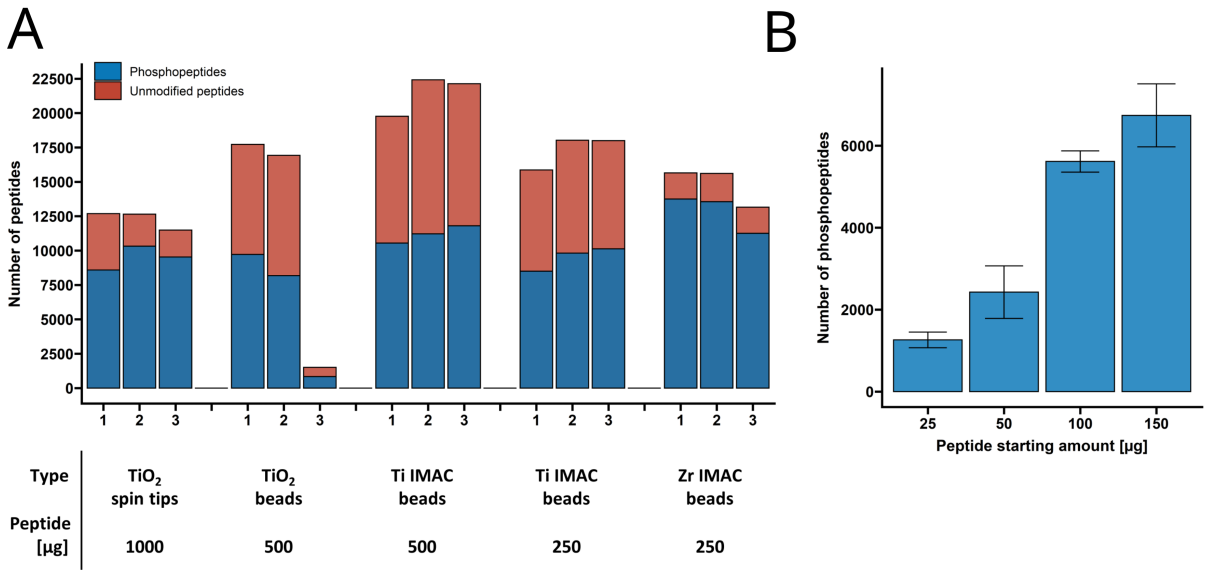

Figure 3.3: Overview of phosphopeptide enrichment of indicated amount of mouse brain peptides performed with magnetic beads measured on an Orbitrap Exploris 480 in DDA mode. A) Performance of different functional material and starting amount compared to TiO_2. B) Titration of starting amount down to 25 µg (N=3) still yields satisfactory number phosphopeptide IDs

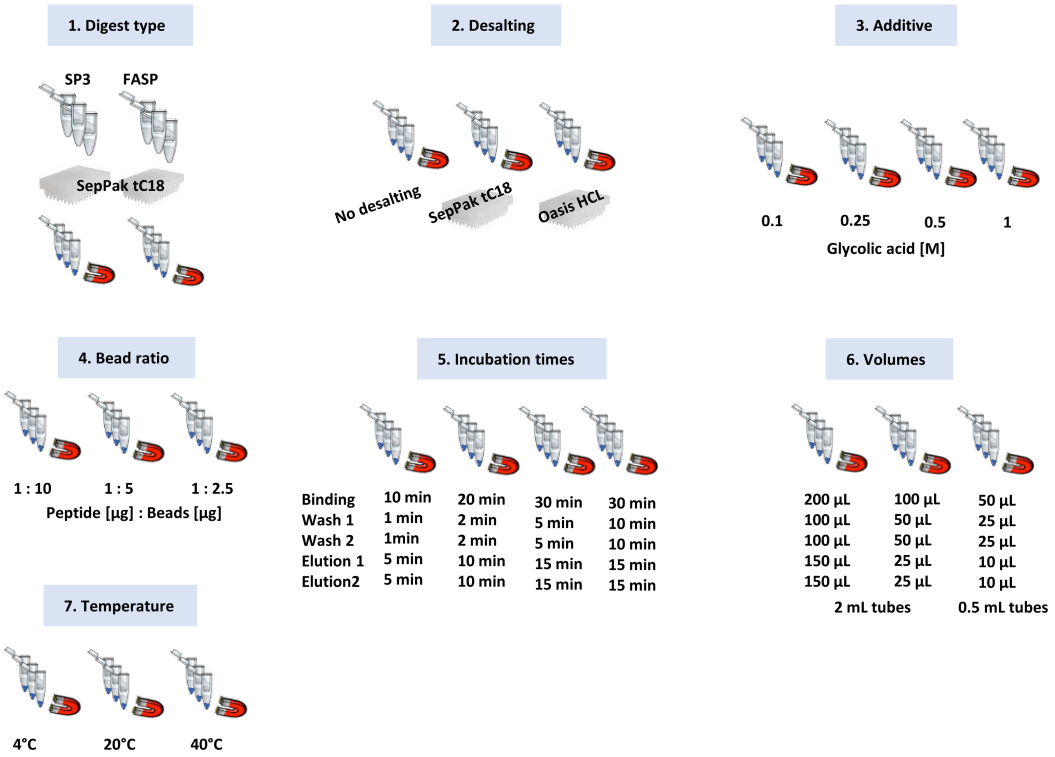

Figure 3.4: Parameters selected for initial phosphopeptide enrichment optimization

[135]. An increased concentration of such metal ions will consequently form complexes with phosphopeptides which subsequently will not be available for ionization. A possible solution for this problem might be a dedicated LC-MS/MS systems for plasma samples, so that as few LC-systems as possible are potentially influenced by plasma sample related LC-MS/MS issues. Furthermore, it has been shown that adding scavenging (chelating) reagents such as citric or medronic acid to the eluent improves ionization efficiency and chromatographic separation of phosphopeptides [136, 137]. To our knowledge it is unknown how a chelating additive affects a complex phosphopeptidome and unmodified peptidome chromatography performance, yet the long time impact of using a non-volatile additive, although in minute concentration. For such LC-MS/MS systems an increased maintenance effort might be necessary, charging of entrance plates and ion optics (incl. quadrupoles) have to be closely monitored and frequently cleaned.

Figure 3.5 shows an overview of the identified phosphorylated and unmodified peptides for each optimization parameter. Based on the first experiments, SP3 digest was selected as digest before phosphopeptide enrichment, as it yields twice as many phosphopeptides than the FASP digest while the enrichment efficiency is equally worse in both digest types compared to TiO_2 spin tips. Presumably, the SP3 digest is more effective in removing relevant contamination that interfere with successful phosphopeptide enrichment as discussed above. Next, the necessity of desalting after phosphopeptide enrichment was assessed by compared crude samples with desalted samples by a) SepPak tC18 an b) Oasis HLB. The hypothesis is, that desalting lead to increased ionization efficiency due to reduced ion suppression. Furthermore, it has been shown that the polarity of the SPE sorbent strongly affects the phosphopeptide recovery [138]. Thus, in addition to the widely applied C18 sorbent a more polar mixed-phase alternative, Oasis HLB, was assessed. The peptide recovery after desalting shows poor performance of both desalting strategies, nearly no phosphopeptides were identified in the desalted samples compared to non desalted samples. This finding can be explained by unspecific binding of the peptides to plasticware and containers which is of course more pronounced with minute sample amounts. Anyway, a solution to this is already provided in many LC systems: Loading of the sample onto pre-columns and subsequent elution by the gradient before the analytical column. Nevertheless, not all LC systems are equipped with this possibility and currently, only C18 material is used due to its broad specificity, thus the efficiency of on-line pre-column

based desalting is questionable. Other pre-column types might improve the phosphopeptide recovery. In addition to that, supplementary figure 6.1 shows appearing peaks in the desalted samples only. As the ordinate is equally scaled in all chromatogramms, it is obvious that the relative amount of additional peaks is high compared to the analytes in figure 6.1 A (no desalting). The base peak m/z is denoted on each peak and reveal the typical chromatographic pattern of Polyethylene glycol (PEG), which is a common contaminant leaching from plasticware and/or from cosmetic products. As two different elution plates after desalting and the same batch of eppendorf tubes and pipette tips were used for processing of the non-desalted samples, the origin of the contamination is unlikely to be derived from cosmetic product residuals *e.g.* from touching plasticware with bare hands after hand care or similar. More likely, the desalting plate material or the filter device of the desalting sorbents in the carrier plate serve as source of this contamination. Interestingly, the Oasis HLB sorbens was able to reduce the amount of a contaminating substance with m/z 1082.52 at RT 30.84 min that is also present in the non-desalted sample and thus is not caused by the additional sample preparation step. In all future assay, no desalting was applied to ensure maximum recovery of peptides after enrichment and preferably, a LC-MS/MS system with pre-columns enabled are preferred. The optimal concentration of glycolic acid as competing agent was assessed next. Here, a concentration of 0.5 M showed peak performance compared to lower and higher concentrations, which is plausible as very low concentrations should lead to increased unspecific binding and high concentrations might prevent optimal phosphopeptide binding by occupying binding capacities. In agreement with this, the enrichment efficiency increases from around 20 % in low competitor concentration to around 50 % in high competitor concentration. The investigation of optimal bead to peptide ratio is ambigous, as the number of phosphopeptides to compare is too low. This experiment has to be repeted for proper evaluation of the parameter, but it serves as another excellent example for the importance of clean LC systems. Shortly before measurement of those samples, a valve was changed during maintenance of the LC. The hypothesis is there, that the metal surfaces have not been passivated yet or due to the maintenance trace metal leaching occurred, which consequently leads to decreased phosphopeptide identification performance. Notably, the second replicate of each condition always shows much worse peptide IDs compared to replicates 1 and 3. When considering only replicates 1 and 3, a lower bead ratio of 1:5

or 1:2.5 seem to perform better, while the reproducibility was higher for the 1:2.5 ratio. Prolonged incubation times in all sample preparation steps have also shown to increase the phosphopeptide yield. A very crucial incubation time is the elution time, as the pH is increased and consequently alkaline hydrolysis of the phosphopeptides is possible, and has to be followed very accurately. The result for the optimization of incubation volumes is ambiguous, as medium and low volume incubation are improved compared to high volumes, but not clearly distinguishable. In favor of homogenization, medium incubation volumes offer a better performance (also regarding the median of phosphopeptide IDs), due to the round bottom shape of the used 2 mL incubation tubes. This geometry favors a homogeneous distribution of the beads in the solution while shaking. In contrast to that, low volume incubation on 0.5 mL tubes offer less surface contact and presumably less unspecific binding which results in an increased reproducibility for these samples. A potential improvement would be the combination of small volumes in 0.5 mL tubes while ensuring proper homogenization, which has to be tested separately. Furthermore, a higher temperature has been identified as beneficial for phosphopeptide identification with moderately higher enrichment efficiency. This finding is contrary to the hypothesis that the retention of phosphosites is thermodynamically favored compared to the unspecific retention of unmodified peptides, which is believed to be governed by acidic amino acid residues that mimic the binding of the phosphogroup. Apparently, the binding affinity is not solely governed by the acidic functional groups of the phospho- and unmodified peptides, but rather by the microenvironment of the whole peptide. If this is the case, the assumptions for unspecific enrichment becomes questionable. A possible explanation would be a HILIC-like retention mechanism, where an enrichment of an aqueous layer around the beads would lead to a liquid-liquid extraction superimposed to the phosphopeptide (and acidic residue) affinity towards the metal cation. To test this hypothesis, a very narrow titration of organic solvent in the loading buffer might reveal a dependence of the amino acid sequence of the unspecifically enriched peptides. If this behavior is true, the ratio of acidic amino acids will decrease with increasing acetonitrile content as the retention mechanism is more and more determined by the peptide-metal ion affinity rather than the HILIC retention.

Table 3.1 summarizes the outcome of the method optimization. Further validation of the method can be performed using synthetic, heavy and light labeled phosphopeptides that

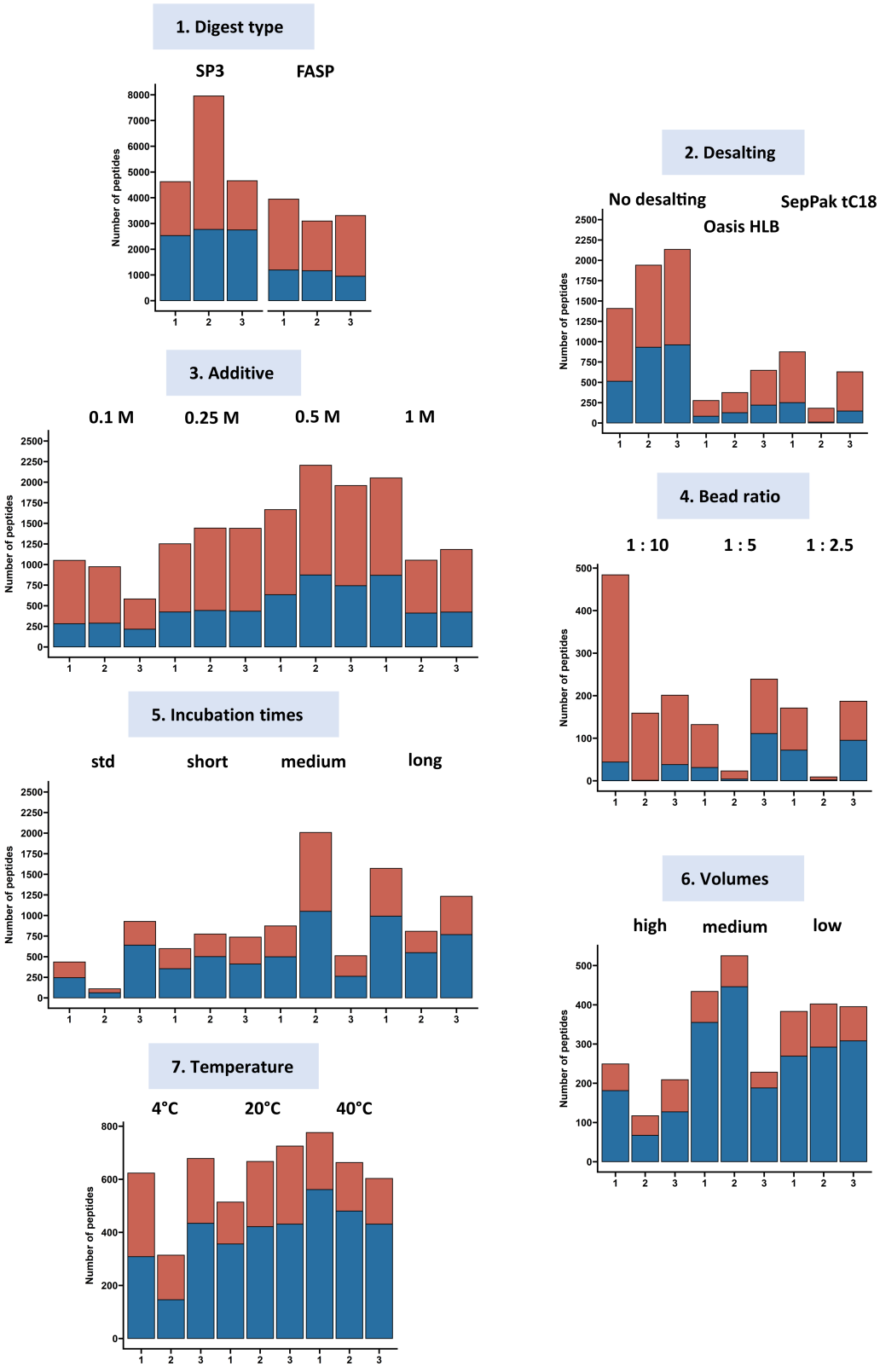

Figure 3.5: Peptide IDs of 25 µg mouse brain samples for the selected parameters measured on Orbitrap Exploris 480, except for experiment number 2 and 4, that were measured on the timsTOF Pro2. Both instruments were operated in DDA mode. Phosphopeptides in blue and unmodified peptides in red.

are commercially available. Different amounts of either labeled phosphopeptides can be spiked into sample matrix and the recovery can be calculated by spike-in of the complementary labeled peptide after enrichment. Furthermore, when adding both peptides in different ratios, the influence of internal standards on the linearity of the quantification can be assessed. The hypothesis is that the ratios become skewed at borderline ratios, but actually it is not known if that ratio is not linear at all. This would implicate that the use of internal standards during phosphopeptide enrichment will never yield exact absolute quantitative information, but rather qualitative, as lower endogenous peptide level will be reported as even lower as they actually are and higher values will be artificially increased. In extreme cases, this can lead to unwanted false positive identifications. To our knowledge, this kind of investigation has not been done before and opens the door for further investigation. Additional potential for improvement arises in the use of TRIS as buffer substance during digest as unpublished research suggests [139]. This assumption is reasonable, as TRIS provides an increased long term stability of the buffer compared to the volatile AMBIC. During over night digest, solvent evaporation and degradation of AMBIC into carbon dioxide, water and ammonia lead to a change in buffer capacity and would eventually result in alkaline conditions, that possibly hydrolyze phosphopeptides and thus result in reduced number of identifications and further decreased reproducibility.

Parameter	**Optimized value**	**Parameter**	**Optimized value**
Digest	SP3	Incubation times	long
Clean up	No desalting	Volumes	medium
Additive	0.5 M	Temperature	40 °C
Peptide:Beads ratio	1:2.5		

Table 3.1: Identified optimal values for phosphopeptide enrichment using Zr^{4+}-IMAC magnetic beads. SP3, yielding comparably clean peptide samples, served as optimal digestion protocol before phosphopeptide enrichment. Moderate amount of additive and a lowered peptide to beads ratio did show the optimal balance for low amount enrichment. Surprisingly, the phosphopeptide loss during desalting outweighted the positive effect during chromatography, thus no deslating after enrichment is advised.

Phosphopeptide enrichment as separate sample preparation step will always introduce not only variability but is also always biased towards certain phosphopeptide subpopula-

tions. A possible solution to this problem would be a selective chromatographic method, that separates the excess of unmodified peptides confidently from the desired phosphopeptides, so they can be immediately measured by the MS while eluting from the column without previous enrichment step. Electrostatic repulsion hydrophilic interaction liquid chromatography (ERLIC) serves as alternative separation mode to the classical C18 reversed phase separation [126]. Although not classified as selective enrichment method, reasonable separation of multi-phosphorylated peptide species has been achieved [140]. This method suits well for the fractionation of previously labeled peptide samples, that can be enriched either before or after ERLIC fractionation [40]. Nevertheless, HILIC method development is unconventional and not straight forward compared to reversed phase methods, as the retention mechanism and the parameters influencing the retention are not fully understood [141]. It is believed, that in HILIC multiple retention mechanisms are superimposed such as mass transfer, liquid-solid interaction with the stationary phase and liquid-liquid interaction with a aqueous rich layer around the functional groups of the stationary phase. When using a charged stationary phase, such as a weak anion exchanger phase, yet another retention (or repulsion) force is superimposed to the other HILIC mechanisms when separating charged analytes. ERLIC for phosphopeptides is always run under conditions where the phosphogroup retains the negative charge, while acidic amino acid residues and peptide termini are protonated and thus neutral or positively charged. This increases repulsion of unmodified peptides and thus aids selectivity. However, the electrostatic attraction of the negatively charged phosphogroup is not sufficient to counteract the repulsion of the positive charges with singly phosphorylated peptides and only becomes reasonable when two or more phosphogroups are present in the peptide. Ultimately, the selectivity of unmodified peptides and singly phosphorylated peptides is driven by the thermodynamics and kinetics of the liquid-liquid interaction and the repulsion from the stationary phase. Phosphopeptides experience an additional attraction force which presumably leads to a 'desorption' kinetic that is driven rather by the electrostatic force than the liquid-liquid extraction. Recent research has shown, that the retention of analytes can be influenced by the choice of a suitable counter ion in the elution buffer [142]. The hypothesis is that counter ions with higher capacity of water molecules in their hydration shell, such as magnesia compared to sodium, will lead to a size increase of the aqueous rich layer on the surface of the stationary phase and thus

increase the significance of the liquid-liquid interaction for the selectivity. in combination with a convex gradient, that utilizes the presumably faster liquid-liquid interaction desorption kinetics, the selectivity should be improved. Figure 3.6 shows the summary of the resulting phosphopeptide separations. For this, each 2000 µg of mouse brain tissue tryptic peptides were injected into an Äkta Pure 20 equipped with a weak anion exchange column operated at 16 °C separated using the stated conditions. The eluate was collected in 1 mL fractions, each desalted and analyzed with LC-MS/MS separately. As discussed before, the published original method by Alpert (2008) provides a good selectivity towards multiply phosphorylated peptides, whereas monophosphorylated peptides mostly coelute with the unmodified peptides. When changing the counter cation to Mg^{2+}, a broader distribution of all analytes over the chromatographic range is observed. In combination with a convex gradient a reasonable selectivity could be achieved. This feasibility study introduces the possibility to perform simultaneous enrichment and chromatographic separation of phosphopeptides, which has not been described in the field before. Nevertheless, further method development is necessary to replace the non-volatile reagents that are currently used with MS compatible additives. This method development was not pursued further, as the required peptide amount is larger than what can be routinely obtained, especially in large scale studies as required for the featured rapid adaptation study. Furthermore, a dedicated MS has to be reserved for a comparably long time, which is resource intensive.

In conclusion, phosphopeptide enrichment is still a challenging task and offers the potential for further improvement. A recently recognized capability of phosphopeptide enrichment is the co-enrichment of glycosilated peptides [143, 144, 145]. On the other hand, this circumstance can be used to further optimize the enrichment efficiency by investigating the impact of additional enzymatic deglycolization by PNGase F or similar, to remove such 'contamination' prior to the enrichment. Furthermore, no solution for the efficient enrichment of pY peptides is available. The enrichment is resource intensive (high amount of sample and costly antibodies for the enrichment are required). Affinity purification columns are not available and offer the possibility for investigation. So far, only polymer imprinted stationary phases have been developed as promising alternative for antibody based methods.

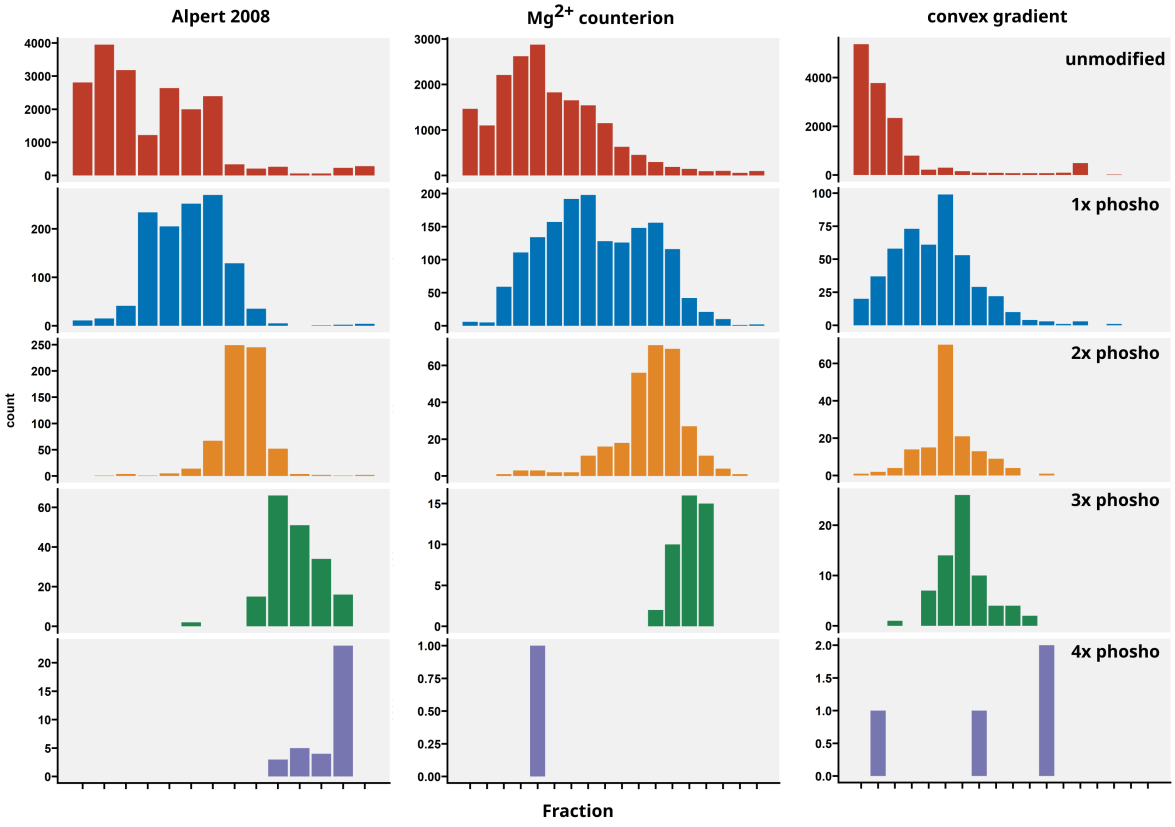

Figure 3.6: Optimization of ERLIC phosphopeptide selectivity of 2000 µg mouse brain peptides by alternative counterion and convex gradient. Peptide counts per collected fractions for unmodified peptide (red), singly phosphorylated (blue), double phosphorylated (yellow), triply phosphorylated (green) and four phosphosites per peptide (purple). While the original method published by Alpert in 2008 shows a high overlap of unmodified and phosphorylated peptides in the middle frations, using a Mg^{2+} counter ion and a convex gradient show a very low number of unmodified peptides was present while still containing a high number of phosphorylated peptides.

3.1.3 Comparison of DDA vs. DIA approach for phosphopeptide identification

A promising approach to gain more confidence in phosphopeptide data is the data independent acquisition (DIA) approach. Per definition, DIA generates MS2 spectra of higher complexity compared to DDA. Especially the identification of the phosphosites requires sophisticated bioinformatic methods that had not been available in the past. Recent implementations in proprietary software such as Spectronaut [97] and developments of open source software such as DIA-NN [84] in combination with affordable high performance computing resources made the analysis of phosphopeptides in DIA possible with sufficient confidence within a reasonable time frame. There are only few publications de-

scribing the use of DIA for phosphopeptides [97, 146, 98] and thus the knowledge about the differences in the data quality have not been reviewed yet, especially in the context of predicted spectral libraries. Furthermore, recent developments in coupling tandem ion mobility spectrometry to high resolution TOF instruments, leading to the commercialization of the timsTOF by Bruker Daltonics, promise a deeper understanding of proteomics datasets by adding an addi-tional identification feature and more confident identification by less complex MS spectra. To investigate the use of DIA for phosphoproteomics in general and especially the use of the Bruker timsTOF Pro 2, we took the opportunity of available measurement time and measured a dataset of three biological replicates of 1000 µg wildtype M. oryzae was measured in DDA with an Orbitrap Exploris 480 and in DIA with a Bruker nanoElute coupled to a timsTOF Pro 2, processed with PEAKS and DIA-NN, respectively and the results summarized in figure 3.7.

The number of identified phosphopeptides is very similar, while the number of unmodified peptides in the DIA samples is significantly higher. Consequently, the apparent enrichment efficiency decreases from around 80 % in DDA to 50 % in DIA as shown in figure 3.7 A. This observation is explained by the DIA scheme, as no criterion for fragmentation is applied, also unmodified peptides with low signal intensity are selected for fragmentation. Interestingly, all unique phosphopeptide identifications of the three replicates combined is roughly 10 % higher in DDA (7663 peptides) compared to DIA (7076 peptides) and the overlap of peptide IDs is small (23 %) as shown in figure 3.7 B. The overlap of peptide sequences without considering the phosphosite was slightly increased with 44 %, so roughly 20 % differ in the assigned phosphosite. It has also not been shown yet, to which extend the software DIA-NN actually provides false positive identifications. To exclude a higher false positive rate as reason for the low number of overlapping identifications, both datasets (DDA and DIA) were searched in either PEAKS or DIA-NN against a concatenated database of *Mus musculus* and *Magnaporthe oryzae* proteome. As the sample was generated from *M. oryzae*, the number of identified *Mus musculus* proteins is expected to be not more than the previously set up false-discovery rate of 1 %. For the DDA dataset, from 36006 total identifications were 112 peptides identified from *Mus musculus* (*i.e.* 0.3 %) and in the DIA dataset, from 40817 total identifications were only 65 identified from *Mus musculus* (*i.e.* 0.2 %). In conclusion, as the false identification rate can be excluded as a reason for the low overlap between the data acquisition strategies.

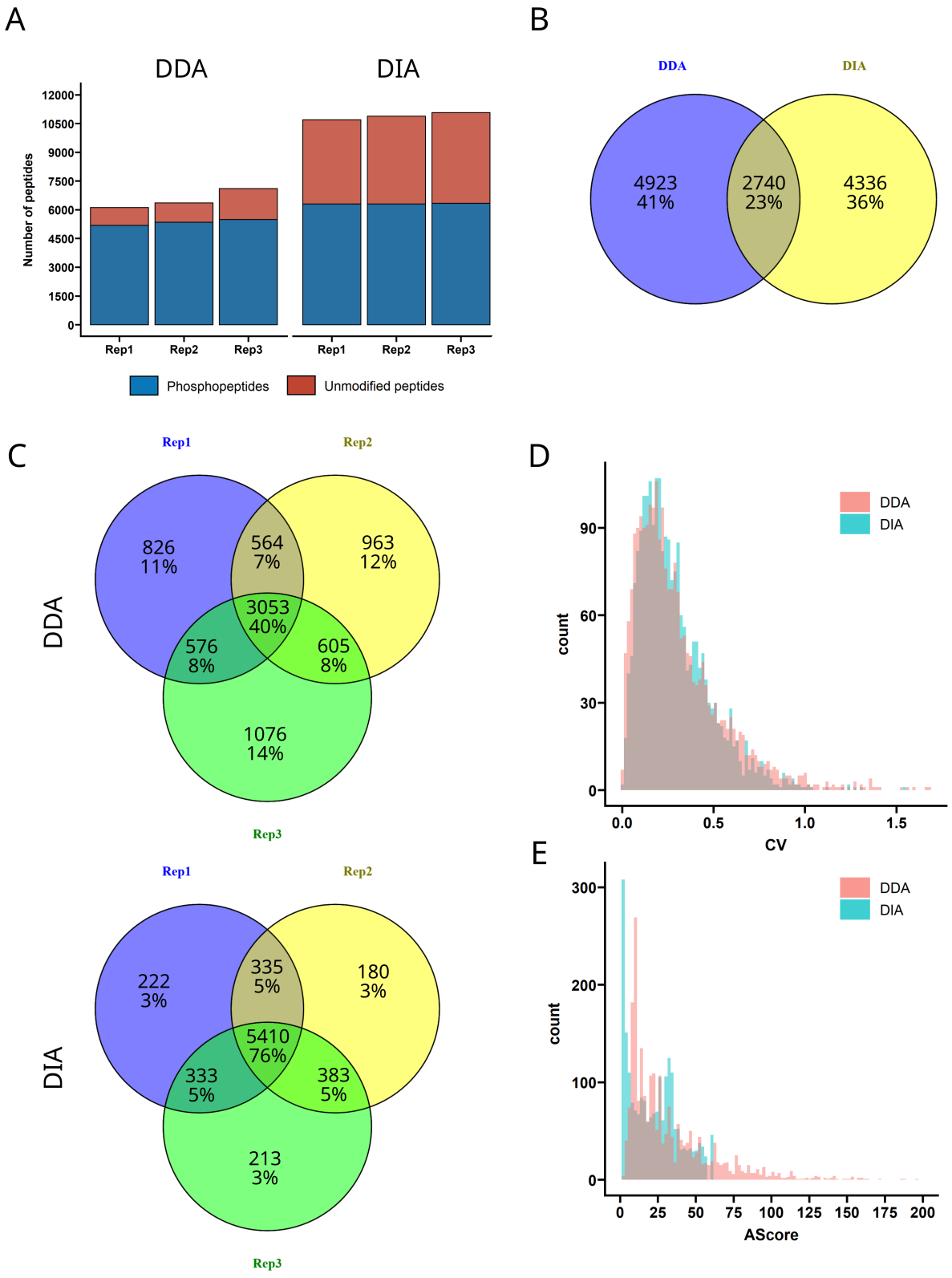

Figure 3.7: Performance comparison of three *M.oryzae* biological replicates, each 1000 µg, enriched for phosphopeptides measured in DDA (Orbitrap Exploris 480) and DIA (timsTOF Pro 2) regarding A) peptide counts B) overlap of identified phosphopeptides C) overlap withing DDA and DIA replicates D) precursor quantity reproducibility and E) phosphosite identification confidence

Comparing the intra-sample group overlap of the identifications within the replicate measurements in figure 3.7 C, reveal another possible reason for the difference in peptide numbers. DIA provides consistently more reproducible identifications, while the overlap for DDA measurements is much less. When accepting only peptides with at least two out of three identifications, the number of quantifiable peptides is 35 % higher in DIA (6461 peptides) compared to DDA (4798 peptides), while the number of complete peptide data (three out of three) is also increased in DIA measurement. Thus, not only the number of quantifiable peptides but also data completeness is increased.

For correctly picturing the biology in the samples, not only the number of quantifiable peptides is important, but also the reproducibility and quality. Therefore, the coefficients of variation (CVs) for every quantifiable peptide (at least two out of three replicates) have been calculated from the replicate measurements and plotted as histogram in figure 3.7 D. The difference between both datasets is not significant with median CVs around 25 %, which is reasonable due to technical variability in LC-MS/MS measurement. A beneficial effect of DIA on data quality has been shown on proteome level [147], which results from the higher number of peptides that are available for quantification.

A second important aspect in phosphopeptide identification is the correct localization of the phosphosite. Both approaches, DDA and DIA offer a confidence measure for the correct site. Nevertheless, even when no evidence for the correct phosphosite is present in the spectrum, the peptide still harbours a phosphogroup at some amino acid, otherwise the peptide precursor mass would not be correct. Thus, we can be confident due to common quality control measures (*e.g. false discovery rate calculation at peptide level*) that there is a phosphogroup somewhere in the peptide present, but the correct phosphosite identification can remain ambigous. Therefore, DIA-NN calculates a site localization probability and PEAKS provides the Ascore, which is calculated by multiplying the negative decadic logarithm of the p-value for incorrect identification by 10. Consequently, the higher the AScore the more confident is the identification with a maximum possible value of 1000. Typically, a confidence of at least 75 % (for calculation of AScore: 25 % probability of false localization) is desired [148]. Therefore, a common cut of value for the AScore is a value of 6, corresponding to 25 % false localization probability. In figure 3.7 E, the distribution of AScores obtained from both acquisition strategies is shown. It is obvious,

that DDA AScores peak around an value of 10, whereas DIA data seems to provide two different peaks, the first peaks with an AScore below 6 and the second peak with an AScore around 30, which equals a site confidence of 99.9 %. Thus, the median site confidence is roughly the same, due to the inhomogenious distribution of the DIA-NN confidences. The reason for this difference is presumably the higher complexity on MS2 in DIA data. There, confidence is only achieved in presence of strong fragment evidence, whereas the algorithm of PEAKS for processing DDA MS2 spectra seems to have a more refined algorithm to assign also calculate variances in probability with high sensitivity. Therefore, the assumption that DDA data provides more confidence in the site localization by higher quality spectra is only partly true. Nevertheless, in discovery phosphoproteomics, the correct phosphorylation site is anyway of less importance. More importantly, both algorithms provide equally high confidence that these peptides are phosphorylated (regardless the phosphosite). Conclusions about active/inactive pathways or protein phosphorylation with approximate protein sites can be drawn anyway.

In conclusion, the application of DIA is a promising strategy for the comprehensive description of a phosphoproteomics dataset. We have shown, that data completeness increases while the data quality remains at least equal. The downside of the DIA application are a resource intensive and time consuming bioinformatic processing and the lack for intuitive spectra visualization. A possible solution to this is provided by the proprietary software Spectronaut, that is able to vizualize XICs of precursors and fragments in a user friendly way [97]. Nevertheless, DIA-NN has been shown to provide superior identification performance utilizing neuronal networks while being open source at the same time. A direct benchmark of both software has not been described in the literature yet and would serve as interesting starting point for further bioinformatics research.

3.2 Rapid evolutionary events in *M.oryzae*

3.2.1 Proteome results

All samples as named in 2.1 were prepared as described in 2.2 and [26]. After rawfile processing in DIA-NN, an average protein number of around 5500 was identified in each sample across the dataset, as shown in figure 3.8. Across all measured samples, 6813 unique proteins could be identified, which equals 53 % of all 12 890 known *M.oryzae* proteins up to this date (16.05.2022). Although the proteomics quality and quantity of previously published *M.oryzae* proteomes show a large variety, with protein identifications ranging from 1600 to 4432 proteins [149, 150], this level of completeness has not been described before. A reason for the excellent coverage of the known proteome is the use of DIA in combination with a large number of samples. In consequence, this analysis benefits from higher chances of finding a good scoring peptide-spectrum match for refined analysis of the whole dataset with an experimental spectral library using the neuronal network strategy. Except wild type, all other samples are deletion mutants for the Hog1 MAP kinase genotype, but show different phenotypes. While wild type, reversibly adapted and irreversibly adapted (WT, REV and IRREV) show an excellent reproducibility in protein counts, the non adapted loss-of-function (LOF) phenotype shows a significant decrease of protein IDs in function of time, with the lowest protein counts reproducibly found in the 24 h samples. This observation reinforces the hypothesis stated before, that unspecific proteolysis occurs upon cell lysis (upon apoptosis in this case, or while sample preparation). This hypothesis could be further verified by creating an *in-silico* library including also unspecifically cleaved peptides, which is computationally intensive, due to the extremely long running times. In addition to that, the resulting predicted spectral library would contain a high number of precursors to potentially identify, which will lead to a drastic decrease in identifications due to the increased ambiguity of peptide-spectrum matches. A less resource intensive solution would be the re-measurement of the whole dataset in DDA and perform unspecific search on the resulting rawfiles. This way, data completeness and quantification performance will be reduced, but the number of identifications should be more equal between the sample types. Nevertheless, the Pearson correlation coefficient across all samples is excellent, though samples taken at 24 h seem to differ from the other time points. Apart from this observation, the correlation coefficient

is consistent within all sample types and differ between the sample types. In addition to that, the other sample types were prepared in parallel and if errors or deviations in sample preparation had occurred as well as an unsuitable (unrobust) sample preparation workflow was used, the irregularities would be randomly distributed across all sample groups. But good intra-sample group correlation and reproducible proteome results in other sample groups both indicate a good data quality within the sample types with apparent differences between them.

To further judge the data quality, the protein identification overlap of all biological samples from the wild type sample of the 24 h time point is shown in figure 3.9 A. 5380 proteins (86 %) have been found in all four replicates while 96 % where identified in two out of four measurements. Across the whole sample set, around 2800 proteins do not have a single missing value as shown in the data completeness plot in figure 3.9 B. The number of proteins increases over proportionally to the inflection point at around 5 missing values (out of 80 samples) with 4200 proteins. With increasing number of missing values from 5 to 79, the number of proteins increases steadily. To our knowledge, this is the most comprehensive proteomics dataset of *M.oryzae* that has been measured so far. In addition to that, not only on qualitative level (how many different proteins do we identify) but also quantitative level (variability in the dataset) we have an excellent dataset at hand, as shown in figure 3.9 C. The main assumption in many proteomics applications (except for *e.g.* pull down experiments) is, that the average level of most proteins does not change upon perturbation. Thus, either the sum of all raw peptide signal intensities (total ion current - TIC) or the sum of all calculated protein abundances (obtained from the processing software) should be within a reasonable range. Due to the large number of samples, no premeasurement was performed to adjust the final injection volume for minimal variance in TIC. Thus, the variability of the TIC is higher compared to the quality control HeLa, that were measured before, during and after the sample sets. Nevertheless, the wild type samples show a similar coefficient of variation of all 20 measured TICs (14.5 %) compared to the 12 measured QC HeLas (12.1 %). The absolute difference in TIC of HeLa compared to the samples is expected, as typically the sample load on column for the QC samples is only 50 ng, whereas all samples were loaded with approximately 100 ng. Thus, the observed difference fits to the expected values. In line with the observations in protein count, the not-adapted LOF phenotype shows the highest variation of around

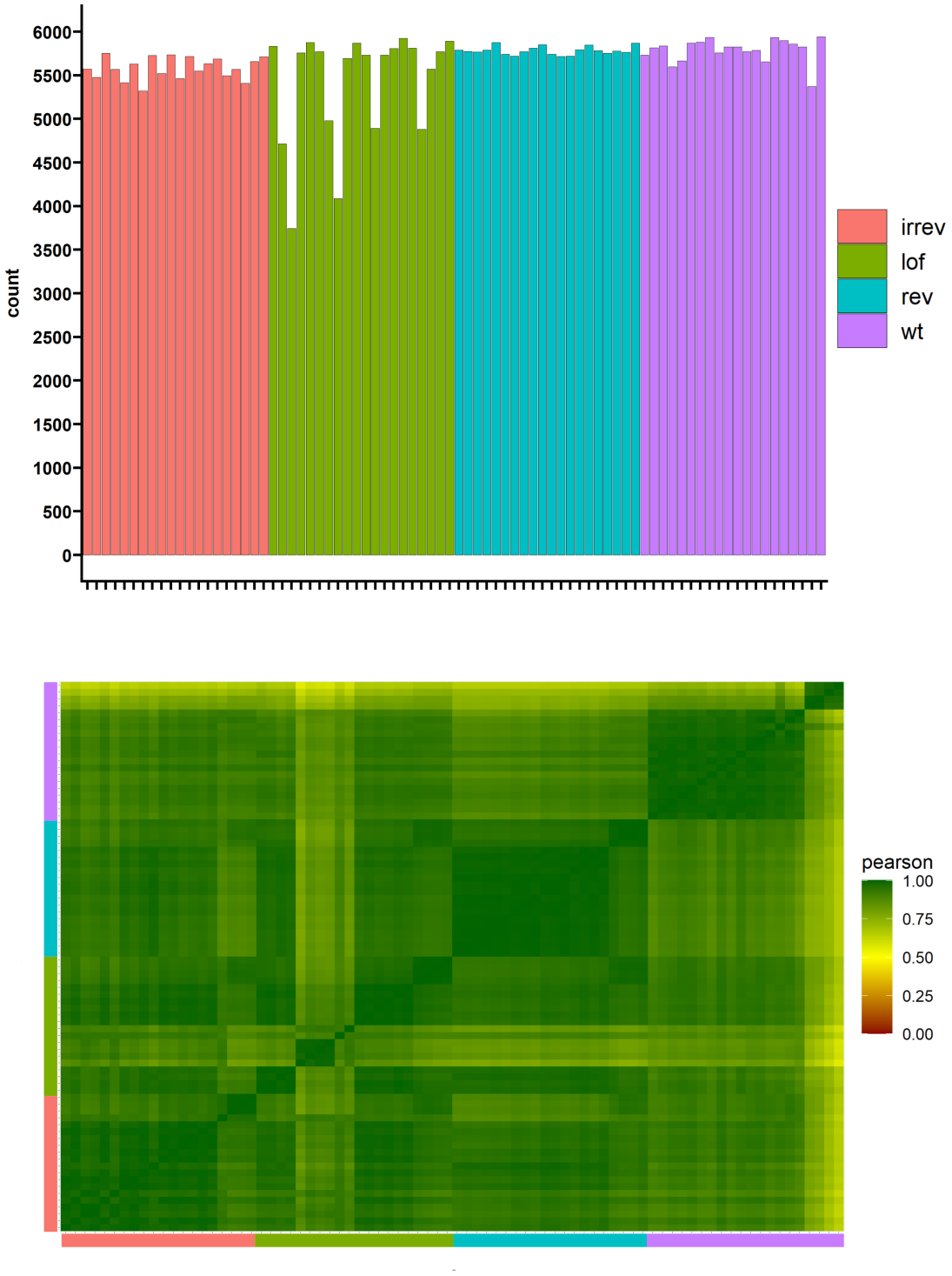

Figure 3.8: Protein identifications and pearson correlation of their quantitative abundance of each 1000 µg *M.oryzae* tryptic digest, measured on Orbitrap Exploris 480 in DIA mode. Each color represents one sample type of the following: irreversibly adapted (red), loss-of-function (green), reversibly adapted (blue) and wild type (purple). Each bar represents one measured sample of the biological quadruplicates side by side, increasing time points from left to right (0 min - control, 10 min, 60 min, 4 h, 24 h)

37 %. Interestingly, the TICs of all three pheotypes with Hog1 LOF genotype do show a significantly higher TIC compared to the wild type. A possible reason for this might be an occurring unspecific cleavage for these samples before protein digest during sample preparation. It is not known, how a high number of small peptides will influence the quantification result, but might falsify the outcome of the result which is used for concentration adjustment before LC-MS/MS measurement. If this is the case, the concentration of the LOF genotypes is higher than expected and would explain this discrepancy. On the other hand, consequently the not-adapted LOF phenotype should have higher TICs, which is not the case (although the CV ranks the highest). Therefore, other differences might also play a role such as batch effects from LC or MS/MS measurement. In general, for all samples the observed CVs are in a very reasonable range with only few highly differing samples. After processing including *in-silico* normalization in DIA-NN, the sum of protein abundances for each sample is calculated and the CVs within one sample group has been calculated. The observed differences and variability in TIC become less prominent when comparing the sum of protein abundances, with CVs ranging from 1.8 % to 13.4 % where the highest CV is consequently observed with the not-adapted LOF phenotype. In conclusion, the assumption could be supported with this results and a high data quality can be expected with a high number of valid protein quantification.

In order to identify significant changing proteins in the dataset, statistical methods such as t-test and linear models proved helpful in the past. While the gold standard in proteomics science is still the well known and easily applicable t-test in Excel or Graphpad Prism, more and more applications appear using linear models. Linear models have been demonstrated to perform superior to t-test in terms of sensitivity and accuracy as early as 2013 on proteome level [151], but recently similar advantages have been reported on peptide level [112]. In addition to that, handling of missing values and small datasets are improved [151]. The popular package *limma* (Linear Models for MicroArray data) was originally developed in 2005 by Smyth *et al.* [152] for comprehensive differential expression analysis of RNA microarray data. It also includes streamlined functions for common problems such as missing value imputation and data normalization and has a broad and active community available for help and advise on statistical problems. Other packages such as DSeq2 offer in principle the same capabilities for statistical testing, but limma has more features implemented and offers increased flexibility and is consequently

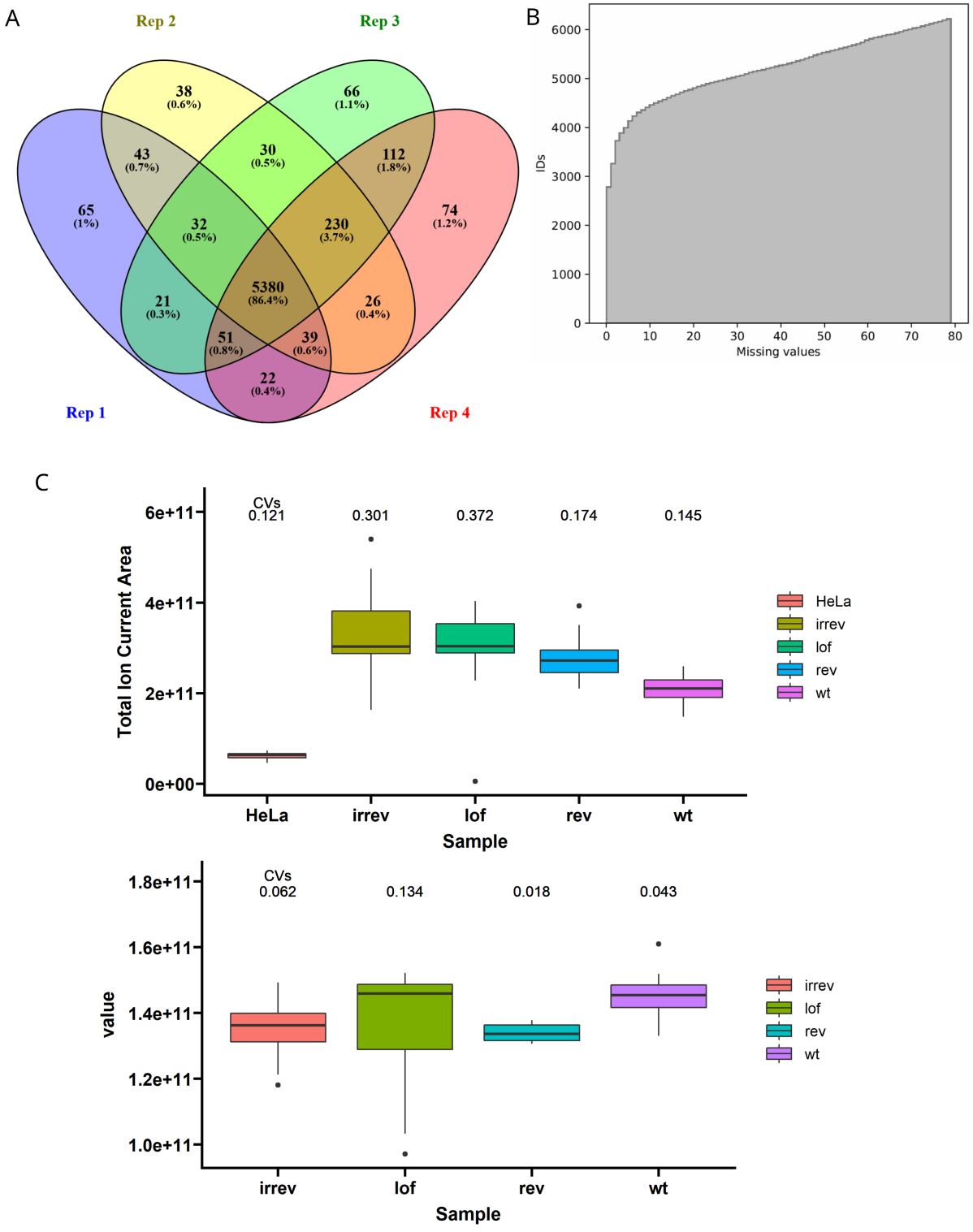

Figure 3.9: Measures of data quality from the data set presented in figure 3.8. A) Overlap of the four biological replicates from WT 1440 min time point B) data completeness and C) variability of TIC and sum of protein intensities (=value)

more popular and frequently used. Still, up to this date, the proteomics community is mostly relying on t-test and similar straight forward principles but is currently transitioning towards linear models. Not surprisingly, this is one of the very few published datasets reported that utilizes linear models for the analysis of the phosphoproteome [97, 153, 154, 155]. Thus, basic considerations for the statistical analysis are discussed for both, proteome and phosphopeptides, in the following.

A Gaussian distribution of the data population is a prerequisite for most statistical tests, including t-test as well as limma. Figure 3.10 shows exemplary the log2 transformed protein abundances of the wild type protein abundances of proteins with at least two identified peptides from DIA-NN. The processing software already applies a normalization strategy for precursors and proteins, consequently the transformed values of the raw protein abundances already satisfy the prerequisite for the statistical test. Therefore VSN, a commonly applied normalization strategy in proteomics, that has been proven superior for small datasets [100], does not influence the appearance of the abundance distribution. On proteome level, the number of missing values is low, nevertheless the following distributions show the influence of common imputation strategies in the data distribution. The single value decomposition method provides already a small skew of the distribution towards right, whereas k-nearest neighbors or limma built-in voom strategy replace the missing values with supposedly too high values, as it does not complement the assumed Gaussian distribution of the transformed protein abundances. The reason for that might be that proteome missing values are in many cases regarded as missing not at random (MNAR), and kNN as well as voom have been proven powerful to replace missing value of random origin (missing at random - MAR) [156]. Therefore, no missing value imputation for proteome data is applied before statistical test in the following to avoid a potential skew in protein abundances.

For the application of t-test, proteins with missing values of more than 40 % in both conditions have been filtered out while confidently appearing or disappearing proteins in one condition (all values NA in one condition, more than 60 % quantifications in the other condition) were assigned an arbitrary p-value of 0.0001 and fold change (FC) of 100 or 0.01 for appearing and disappearing respectively. The dataset was tested with two-sided hypothesis with assuming an equal variance in both conditions. Each sample

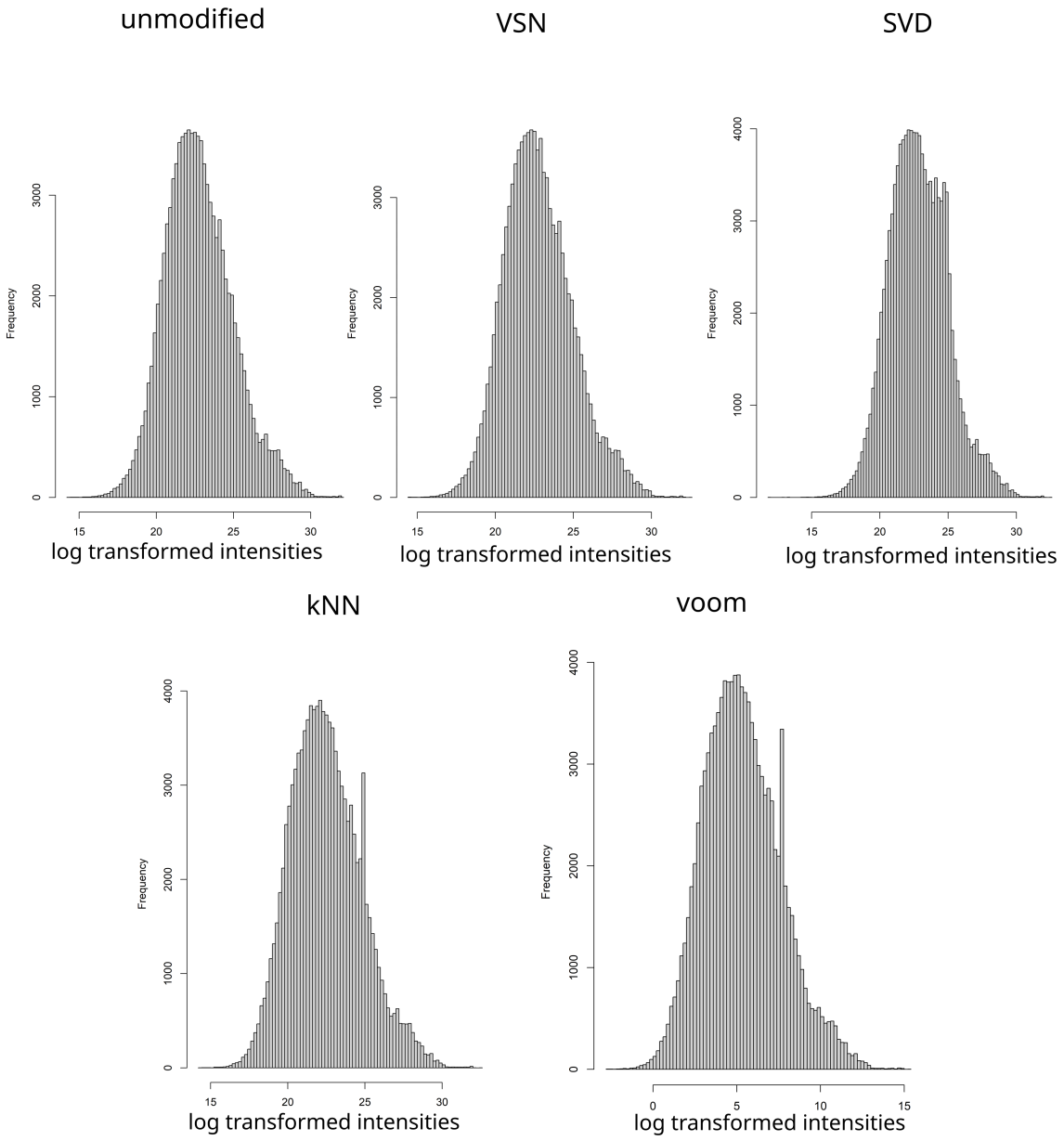

Figure 3.10: Evaluation of different normalization and missing value imputation strategies for protein level of the dataset presented in figure 3.8. All observed protein quantitative values were subjected to the stated normalization strategy and the resulting measured and imputed value are shown as histogram of the log transformed quantitative values.

was cultivated separately, thus increasing of statistical power by taking paired samples into account was not possible, the t-test was conducted as unpaired. Linear modelling using limma was performed using the standard parameters. The resulting p-values for the significance of a protein change were adjusted for multiple testing using the Benjamini-Hochberg False Discovery Rate (FDR) correction, which is called q-value in the following. An arbitrary fold change value of at least 2-fold or 0.5-fold and an adjusted p-value below 0.05 were defined as criteria for statistical significantly changing instances. Those threshold values have been defined arbitrary in the past and have historically been proven as good trade off values to identfy true positive changes from highly variable datasets, that introduce a variability due to sample preparation and measurement stability that was below a factor of 2 (*i.e.* 2-fold and 0.5-fold changes). This procedure was applied to compare each time point to the control. The results of the 24 h versus control time point from both statistical tests are shown in figure 3.11 A and B. In the depicted volcano plots, the negative decadic logarithm of the q-value is placed on the ordinate while the log2 transformed fold change is placed on the abscissae. This way, higher significant instances are found higher in the plot and fold changes are equidistant reflecting the factor of change. While the t-test provides a higher number of statistically significant changing instances, the distribution of the q-value and fold-change pairs is more condensed compared to the limma results. Presumably, the statistical power of the test is weaker due to considering only the standard deviations and means of the protein abundances between conditions instead of the quality parameters for linear models (residual sum of squares) and the assignment of an arbitrary p-value, which might skew the p-value adjustment towards more statistically significant q-values. Although the difference in shape is minimal, the shape of the limma volcano plot follows the expected distribution, as higher fold changes are more likely to reflect a true positive change while the t-test provides a volcano plot where also in less significant q-values higher fold-changes can be found. for the comparison of the biological outcome of the statistical test, all significant instances from t-test and limma across all time points were collected and the overlap was calculated as shown in figure 3.11 C. In general, the overlap is reasonably high with over 41 % of all significant instances. Interestingly, the number of exclusively identified instances is almost equal between t-test and limma. Following current opinions and own conclusions, limma has been used for the whole project to identify significant changes.

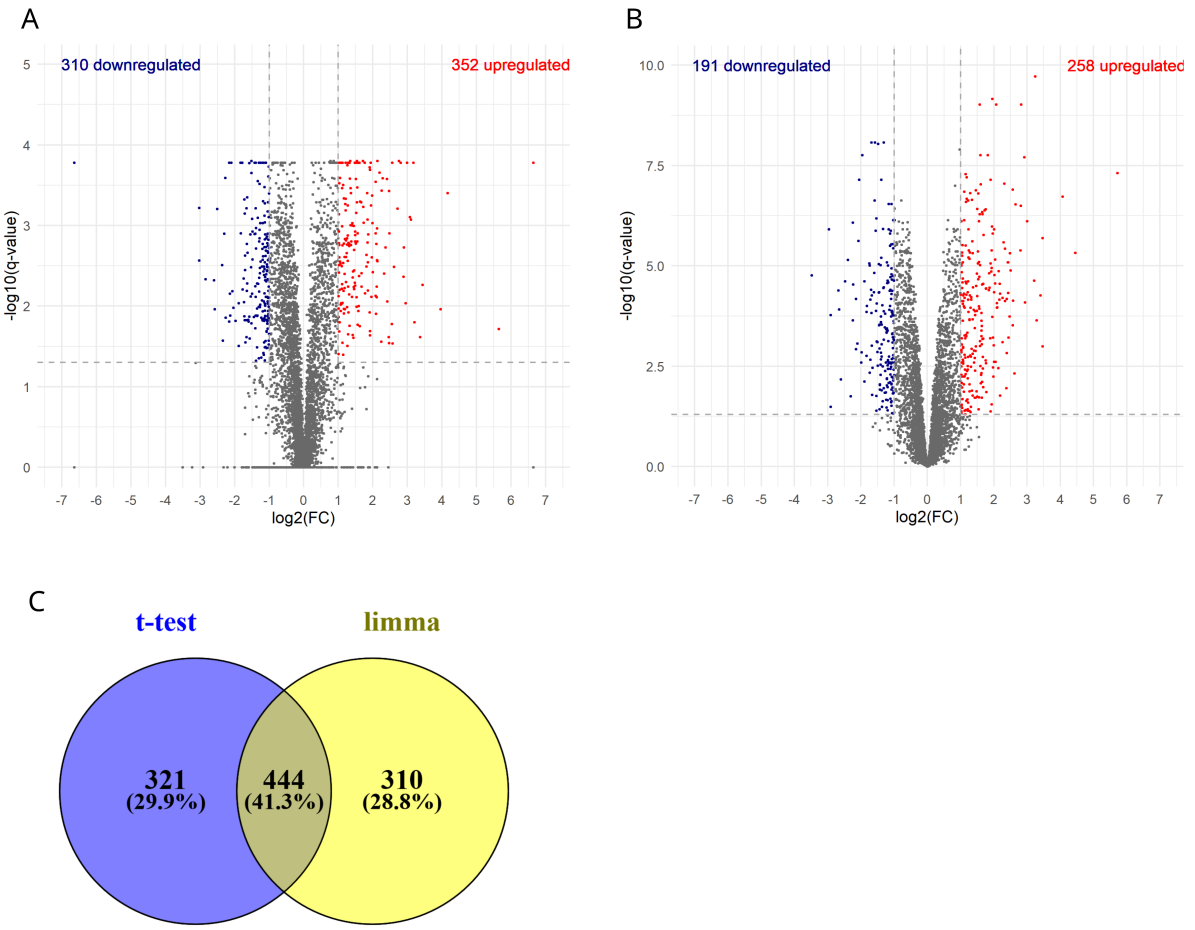

Figure 3.11: Comparison of statistical testing for the wild type proteome results presented in figure 3.8. For this comparison, all replicate samples from time point 24 h versus control time point were subjected to statistical testing. A) Student's t-test B) limma and C) the overlap of significantly changing instances from each test

Further ways for quality control and biological insight into the dataset is provided by principal component analysis and hierarchical clustering in sample level. For both approaches, a complete dataset is required without missing values. According to the previously conducted analysis taking various imputation strategies into account, SVD imputation has been applied, as the influence on the dataset is minimal compared all tested strategies. Figure 3.12 shows the result of these multivariate statistical methods. The principal component analysis explains with the first two principal components already 98 % of the observed variance in the dataset. Timepoints 0 min to 240 min cluster together, while the clustering of the biological replicates are mediocre but existing. Separated by principal component 2 that accounts for only 18 % of the variance is only the 1440 min time point, suggesting that the singificant instances from that time point are more likely to

drive the cellular response to osmostress compared to the other time points. Nevertheless, all biological replicates are separated by PC2 which adds confidence in the technical reproducibility of the dataset. The hierarchical clustering of proteins (rows) and samples (columns) is illustrated by z-score of the log transformed protein abundances in the heatmap. The dendogram for the hierarchical clustering of the samples is shown on top, while the clustering of the proteins is not shown as the resulting dendogram is too dense and crowded to extract meaningful information. The most valuable information from this heatmap is the good alignment of the biological replicates into the same clusters. Only one replicate from time point 60 min is switched with one replicate from the control (time point 0 min). In respect of the general quality attributes and the good alignment of the really differing samples, a technical cause for this discrepancy in similarity is unlikely (*i.e.* a wrong sample and rawfile mapping, accidentally switching samples during processing and testing *et cetera*).

3.2.2 Proteomic response in wild type upon KCl stress

To prove the applicability of the approach to identify meaningful biological processes, the proteome and phosphoproteome response of the wild type upon KCl osmotic stress are analyzed in detail without prior targeted data review. It is expected to identify the previously described osmostress response of yeast [157], which should be MAP Kinase signaling by HOG that will be validated by specifically interrogating the dataset about involved proteins in this pathway. Figure 3.13 shows the volcano plots for the protein abundances of each time point compared to the control of the wild type. After 10 min of KCl stress situation, no significant changes were observed, while the number only slightly increases after 60 min. This observation is expected and additionally underlines the data quality, as the environmental sensing as well as the transcription and translation of genes to proteins requires time in the range of hours rather than minutes. Consequently, the number of changing proteins increases significantly after 4 h of treatment and peaks for the 24 h osmostress samples, while the increase between 60 min and 4 h is more pronounced than the increase between during the next 20 h on KCl stress medium. Interestingly, some of the proteins that are differentially expressed after 4 h revert back to their control level expression, which indicates a regulatoray role while proteins exclusively differentially expressed after 24 h rather fulfill a homeostasis related role. In addition to the explainable

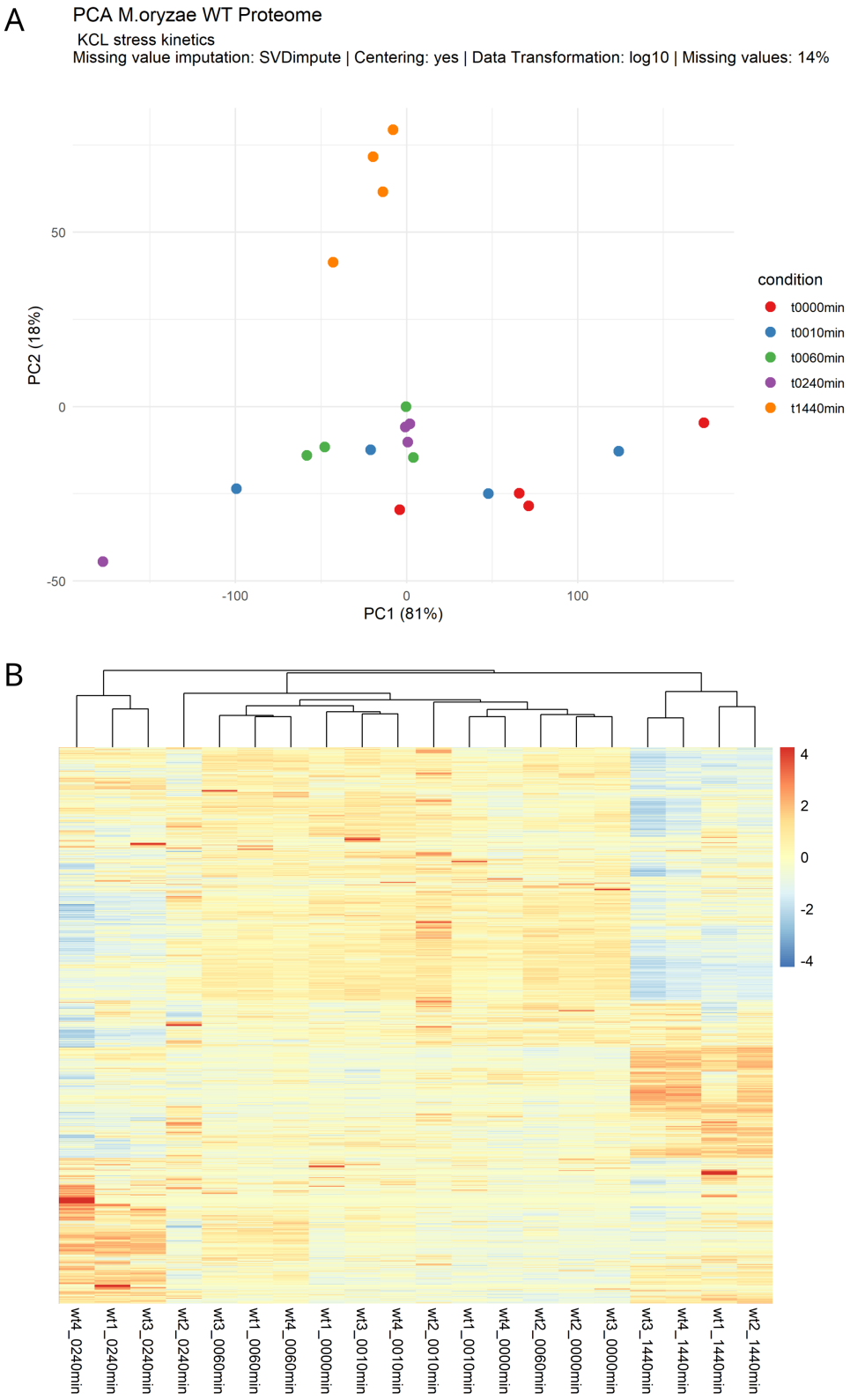

Figure 3.12: Multivariate statistical analysis of the dataset presented in figure 3.8. A) Principal component analysis B) heatmap for sample clustering as quality control of all wild type proteome results combined

(expected) number of significantly differentially expressed proteins, no skew or bias in fold changes or adjusted p-values could be observed which indicates the validity of this approach.

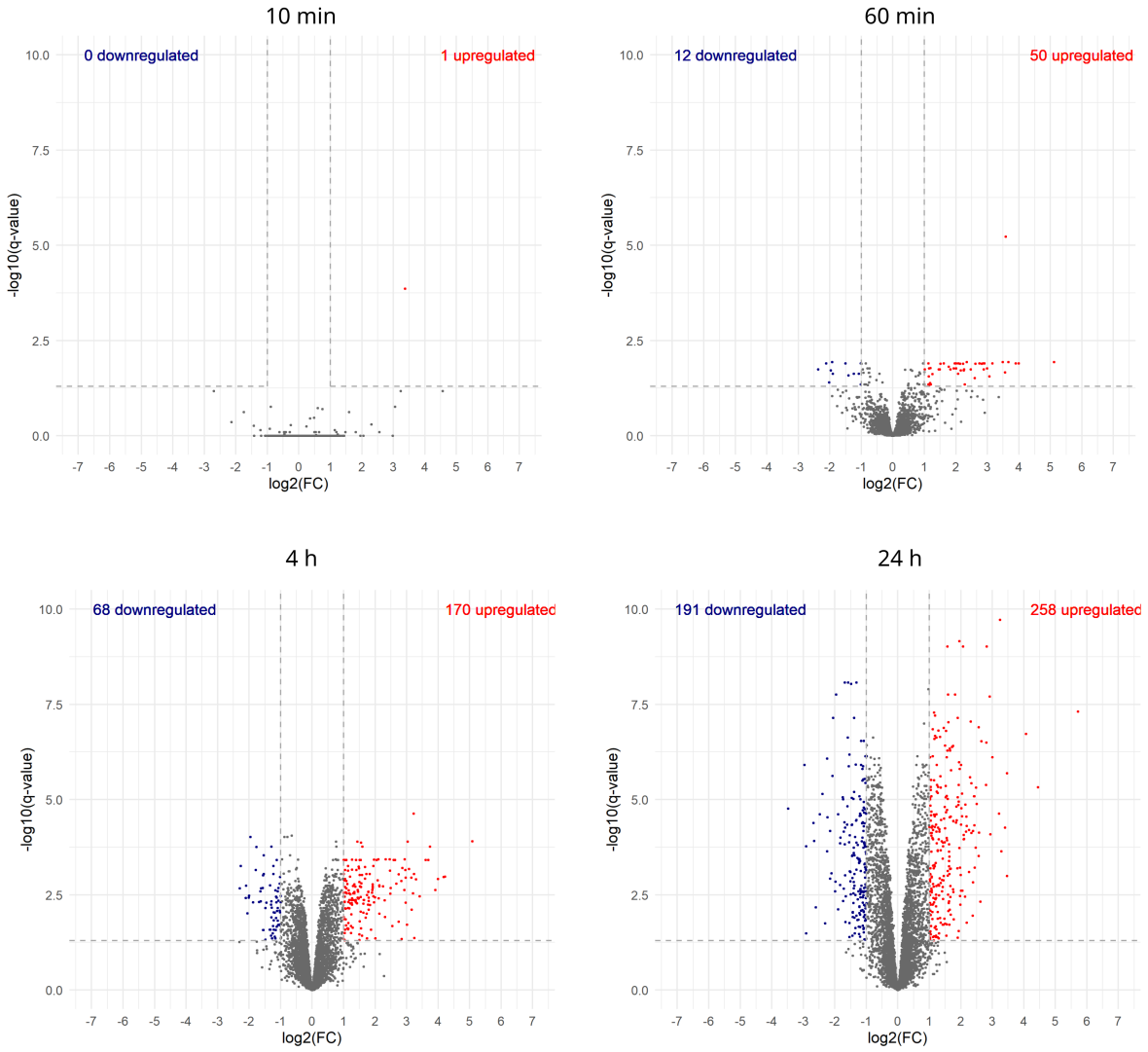

Figure 3.13: Volcano plots of changing protein levels for the dataset presented in figure 3.8 as fold changes and q-values for each time point compared to the control of *M. oryzae* wild type samples upon osmotic stress.

Gene ontology clustering analysis in Cytoscape ClueGO with standard parameters is shown in figure 3.14. Already this simple analysis reveals osmoregulation associated terms such as *carbohydrate metabolic process*. As complementary method for gene ontology enrichment also STRING-DB was used and the obtained significant GO terms are shown an negative decadic logarithm of the adjusted p-value. In agreement with the ClueGO analysis, KEGG pathway enrichment indicates the expected proteomic response as *Metabolic*

pathways and *Biosynthesis of secondary metabolites*. Nevertheless, the full potential of the dataset in not yet used, as temporal resolution of the proteome response is given.

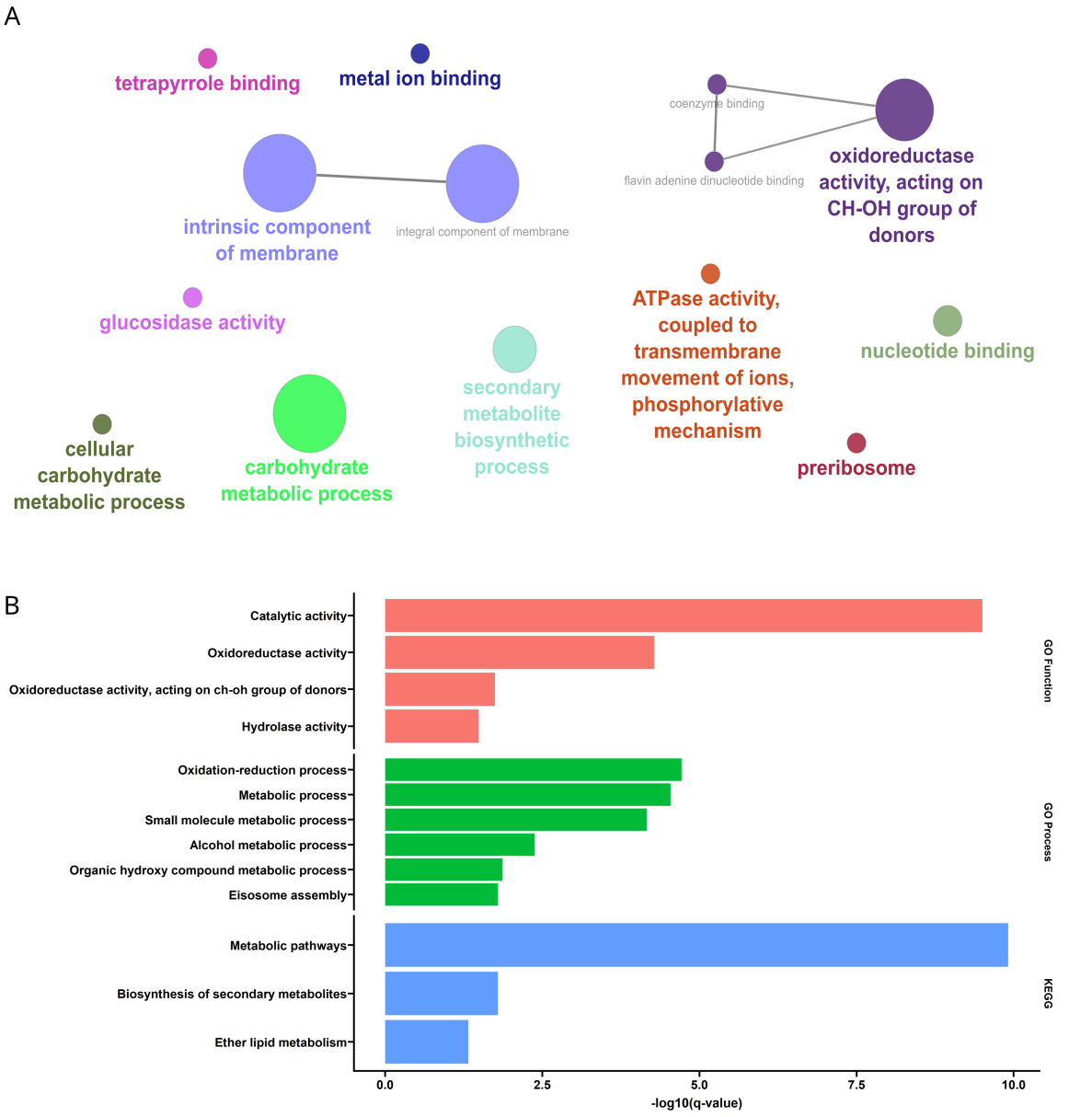

Figure 3.14: Gene ontology analysis of all signifiantly changing proteins of *M.oryzae* of the dataset presented in figure 3.8 upon osmotic stress during 24 h. A) in ClueGO and B) from STRING-DB

Time course experiments enable conclusions about temporal regulatory roles of proteins, but the identification of such is not straight forward and resource intensive. Furthermore, the large number of over 600 differentially abundant proteins could be identified across all time points. This large number might obscure underlying effects and hinder the effective and precise identification of relevant biological meanings. Thus, data reduction is key

to understand the responses in more detail. In order to do so, clustering of protomic changes should be applied to the calculated fold-changes to reduce the number of intersting proteins that might follow a similar temporal response pattern. As data completeness is required for clustering in general, it was intentionally not performed using the raw protein abundances, as it is likely that missing value imputation will have a larger impact on the outcome compared to the use of fold-changes (*i.e.* the number of missing values is higher). For this dataset, clustering by k-means, dbscan and hierarchical clustering have been evaluated and the results of the diagnostic analysis is shown in figure 3.15 to 3.17 respectively.

For k-means clustering, a number of clusters has to be defined as parameter for the clustering [158]. The within-ness plot and the shilouette plot serve as indicators for the number of clusters k to choose. As shown in figure 3.15, both diagnostic plots suggest the use of k=2. This number is obvious by visual inspection of the principal component analysis of the proteins, where two explicit clusters are visible. Not surprisingly, the corresponding spaghetti plot in figure 3.18 reveals the nature of these clusters increasing and decreasing in fold changes over time.

The dbscan clustering algorithm requires a distance ϵ and the minimum number of entries as input parameters [159]. The used R package provides the 5-Nearest-Neigbour plot as diagnostic for the identification of a senseful ϵ, which is shown in figure 3.16. Based on this, an epsilon of 1 was chosen for the first analysis and in accordance with the k-means clustering diagnostics, the same two clusters are identified by the dbscan algorithm, but also excluding potential outliners from the analysis (because they do not match the clustering criteria such as minimum number of neighbors within the chosen ϵ) as shown in the PCA plot. Not surprisingly, independent from the chosen distance and the minimum number of neighbors, the number of clusters remained two, differing only in the number and position of outliners. This underlines the robustness of the algorithm for unsupervised learning, but also shows its limitation as in this case a higher flexibility is required. Consequently, the desired aim to reduce the complexity of the results can not be efficiently realized using dbscan.

Hierarchical clustering requires a starting point and the desired tree-height cut-off value as parameters [160]. The algorithm yields very inhomogeneous distribution of entries, with

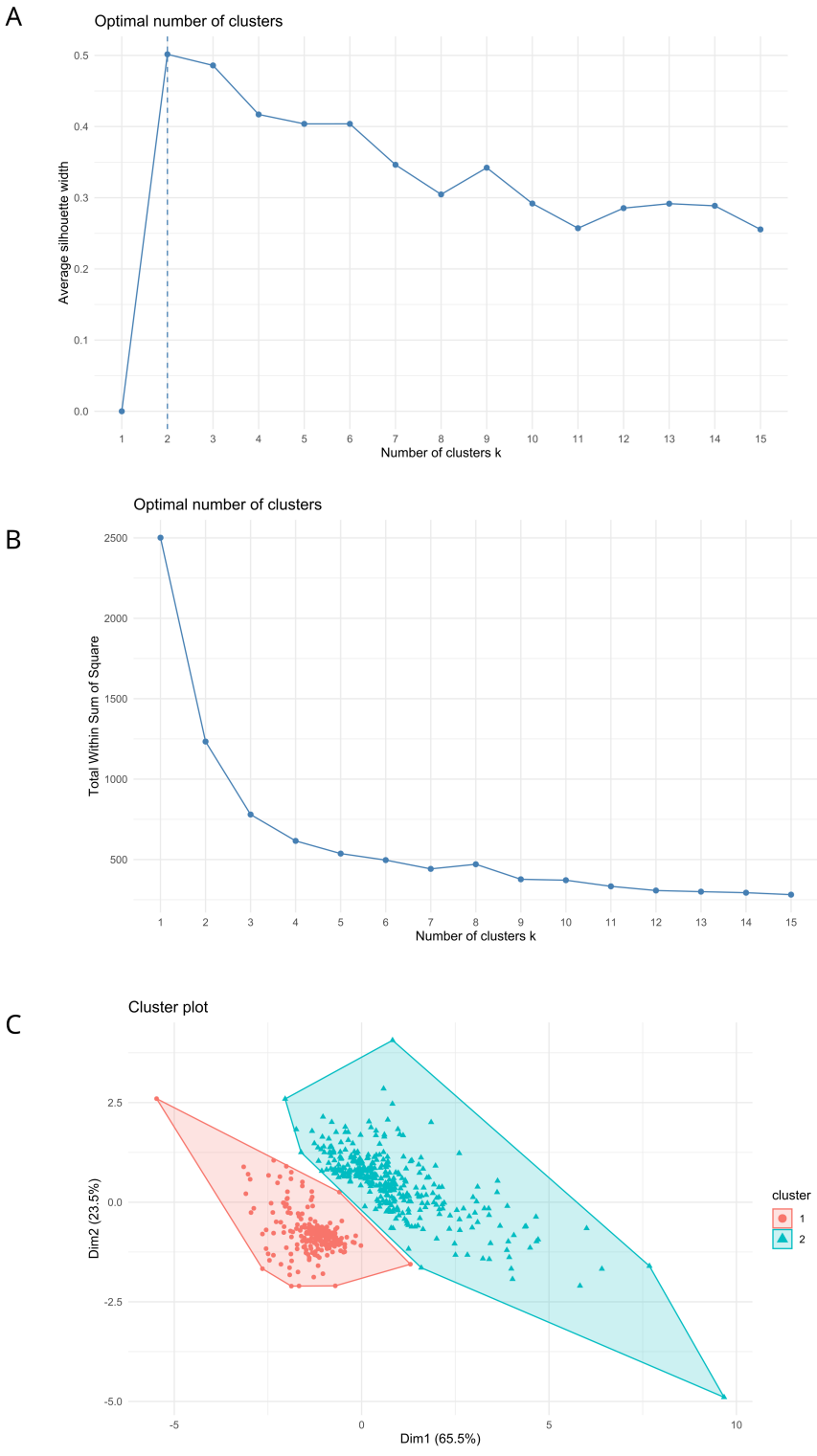

Figure 3.15: Diagnostics for clustering according to the k-means clustering algorithm applied on observed proteome quantities of *M. oryzae* of the dataset presented in figure 3.8 upon osmotic stress in the time course of 24 h. A) Silhouette plot B) Withinness plot C) PCA with color code of the two suggested groups in red and turquoise

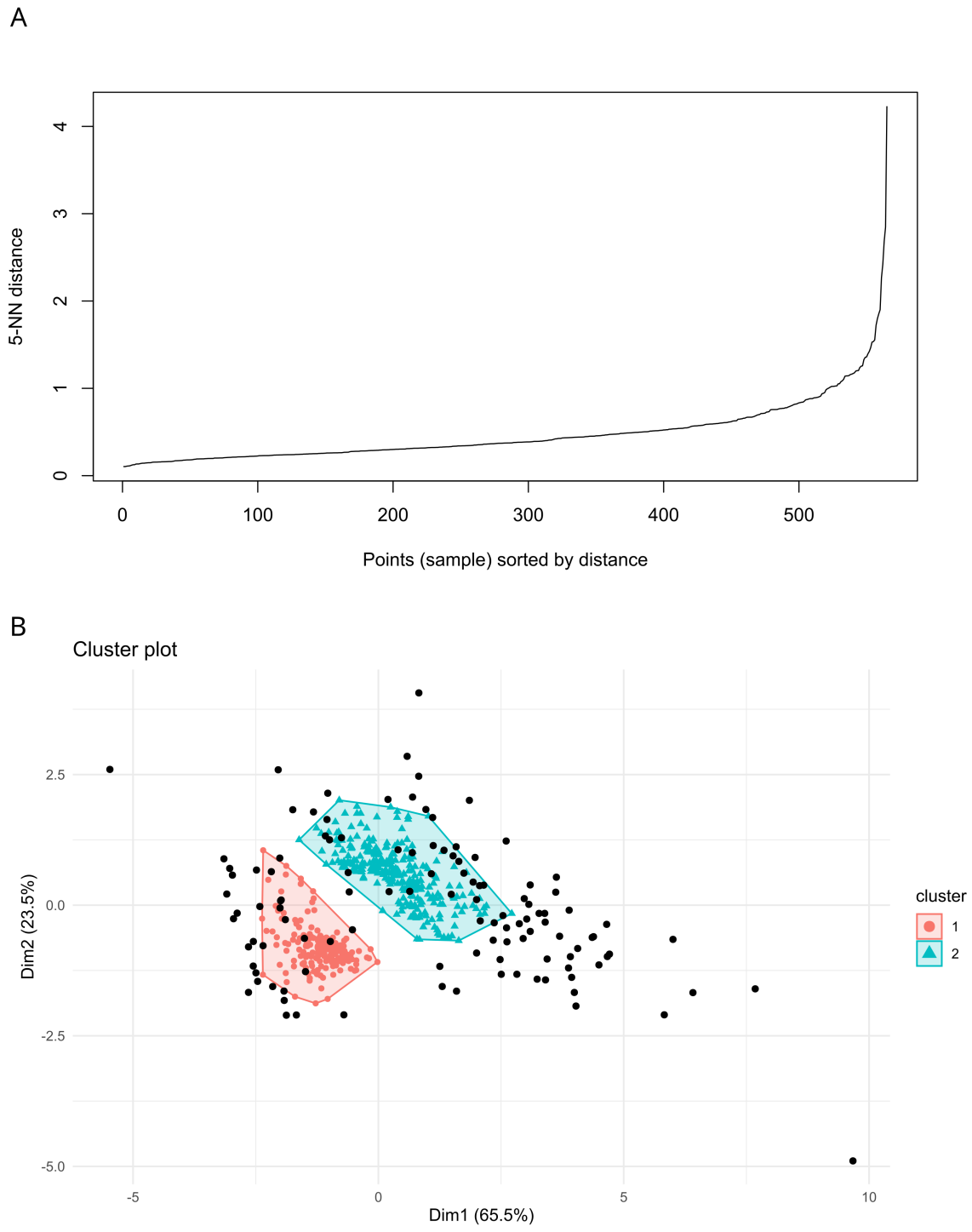

Figure 3.16: Diagnostics for clustering according to the dbscan clustering algorithm applied on observed proteome quantities of *M.oryzae* of the dataset presented in figure 3.8 upon osmotic stress in the time course of 24 h. A) Distance plot B) PCA with color code of the two suggested groups in red and turquoise. Black dots represent outlines, that were excluded by the algorithm and not assigned to one of the groups

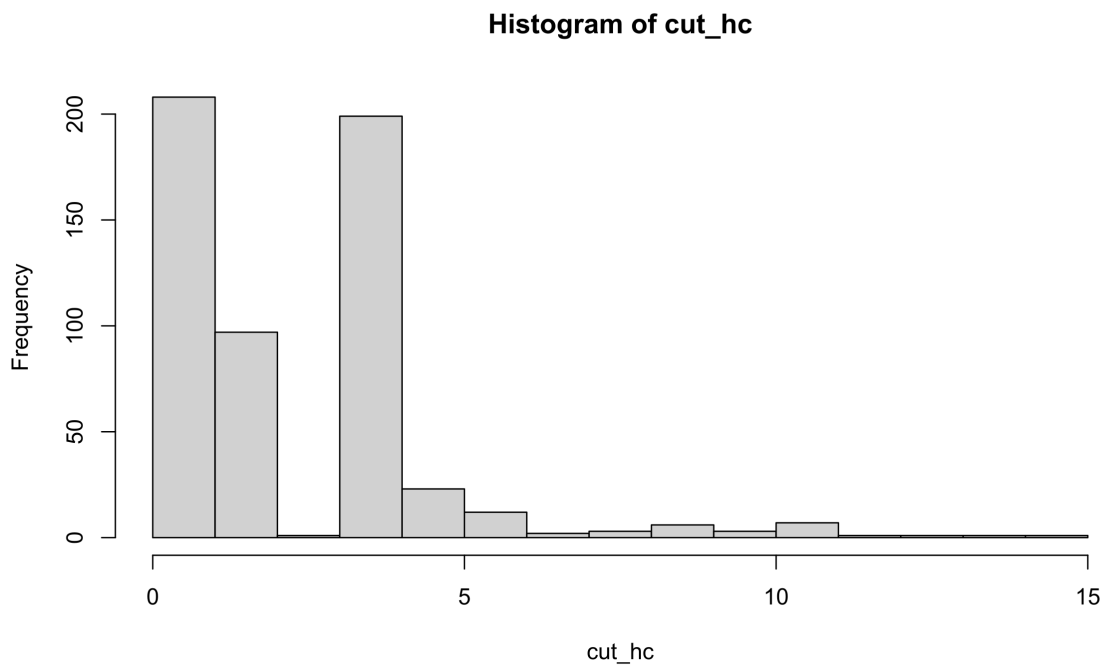

Figure 3.17: Diagnostics for clustering according to the hierarchical clustering algorithm applied on observed proteome quantities of *M. oryzae* of the dataset presented in figure 3.8 upon osmotic stress in the time course of 24 h.

most entries distributing always between two clusters as shown in the histogram in figure 3.17. Here, an exemplary number of 15 clusters were chosen as cut-off, but choosing a higher or lower number of cluster cut-off does not yield in a more homogeneous distribution of entries. Thus, also with hierarchical clustering, the aim can not be met.

In conclusion, k-means clustering is the only strategy offering the required flexibility to divide the dataset into the desired smaller parts, that can be analyzed in the required detail with reasonable use of resources. For this, arbitrary values for k were tested and visually inspected in order to balance between the number of proteins in each cluster and the biological meaning as summarized in figure 3.18. With k=5, a reasonable number of clusters could be identified, while already with k=10 the redundancy of similarly changing proteins is increased. Naturally, when dividing the dataset into k=15 clusters, the redundancy is even more increased, with no obvious benefit in differentiating between biologically different temporal responses. Based on these observations, clustering with k=5 was used for further analysis. The results of the clustering and the corresponding ClueGO

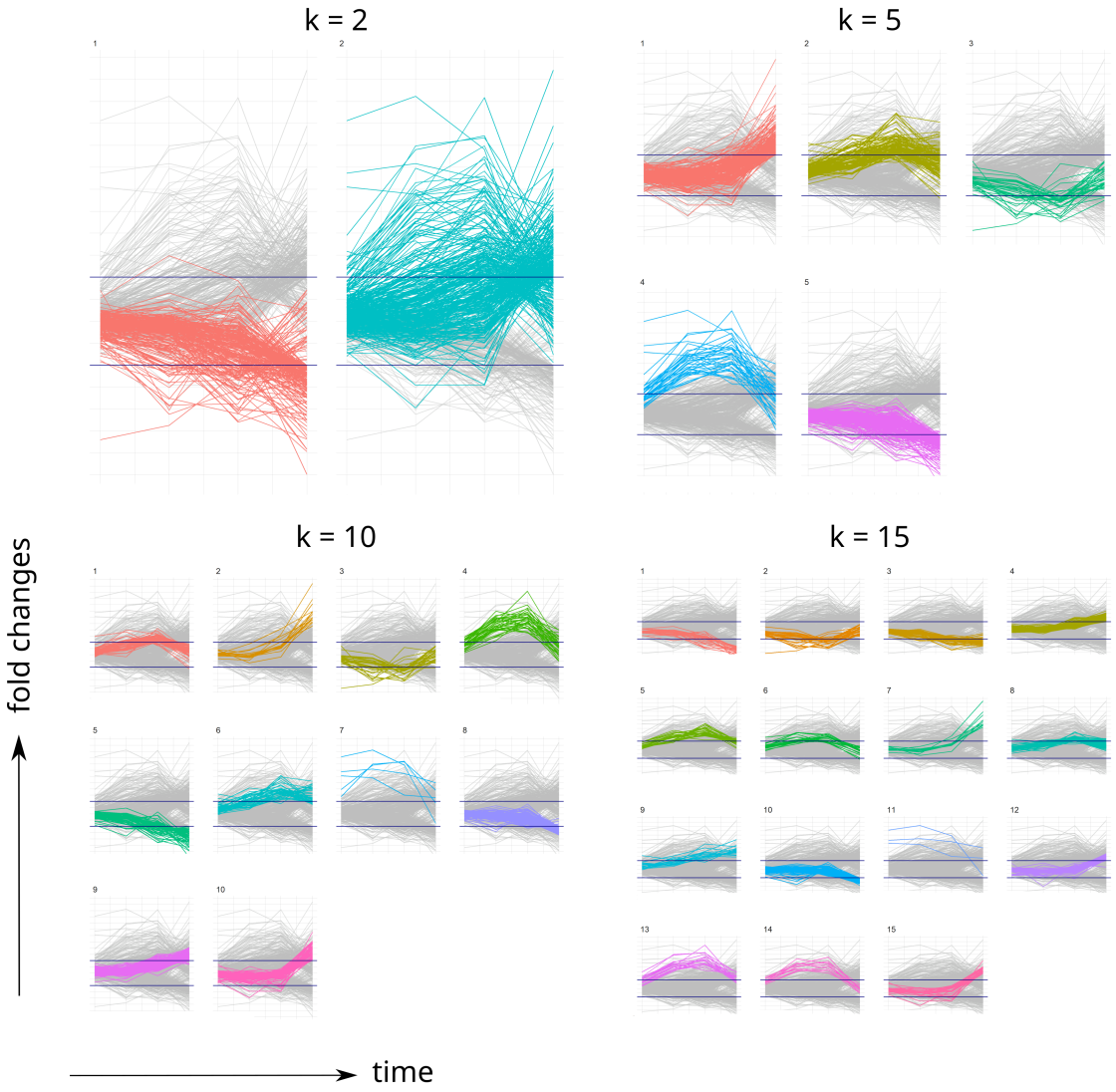

Figure 3.18: Empirical clustering by k-means with k = 2 / 5 / 10 / 15 applied on observed proteome quantities of *M.oryzae* of the dataset presented in figure 3.8 upon osmotic stress in the time course of 24 h. The commonly used threshold for signifcantly changing instances log2(fold change) = 1 and -1 are marked with blue horizontal lines for better identifcation of signifcant response clusters.

gene ontology enrichment / clustering analysis is shown in supplementary figure 6.2 in sufficient detail. In these cases, ClueGO analysis was performed with relaxed settings, collecting all GO terms regardless their statistical significance. This way, a very detailed functional analysis can be conducted that is not convoluted by the reduced number of proteins. In cluster 1, that represents proteins that become significant only at the 24 h time point, terms that realize homeostasis could be identified, which includes sugar metabolism, glucosidase activity and active transmembrane transporter activity. This is in line with previously uncovered findings that transporters such as Stl1 play a role in osmostress or as general stress response. In the second cluster, proteins are included that are upregulated short term after stress (60 min and 240 min) and tend to fall back to their control level or stay upregulated in homeostasis at 24 h. Frequent terms related to phosphotidylinositol also in combination with Torc2 signaling have been observed. Torc2 inhibition has ben shown in yeast to inhibit glyercol efflux by closing the aquaporin Fps1 and thus aid the accumulation of glycerol as intracellular osmolyte, which is completely independent by Hog1 activation [161, 162]. This observation is surprising, as evidence exists that Hog1 deficient yeast in general is sensitive towards hyperosmotic conditions. Thus, such an HOG independent alternative osmostress response would have been expected in the adapted Hog1 lof mutants to compensate the missing Hog1 functionality. Nevertheless, intracellular glycerol accumulation has also been observed in wild type upon osmotic stress, but to a much lesser extend than the accumulation of arabitol [4]. The role of Hog1 in the accumulation of arabitol as main osmolyte is still unclear and thus, Torc2 signaling might be key for the adaptation process in the adapted Hog1 deletion mutants. In cluster three, mainly downregulated proteins can be found, that recover their level from the control samples in homeostasis at 24 h. Most of the identified GO terms involve the regulation of transcription of cell cycle related proteins. This is in line with the observation, that the wild type *M.oryzae* is downregulating the cell cycle process to focus the resources on establishing a stable response to the osmotic stress. Interestingly, sphingolipid and glycerolipid metabolism is also present in this cluster, which might be connected to the previously identified Torc2 signaling. It has been described that sphingolipid metabolism is involved in cell wall remodelling as response to environmental changes [163]. In cluster 4 , proteins are included that show a very strongly upregulated response in early time points that decrease in intensity at homeostasis. Gene ontology terms of this cluster in-

clude pentose phosphate pathway and pentose / glucuronate interconversions which both have been shown to be involved in the production of the pentose arabitol [164]. This explains the observed phenotype of *M.oryzae* wild type that accumulates arabitol as intracellular osmolyte upon hyperosmotic stress. The last cluster contains proteins with steadily decreasing abundance. Most obtained gene ontology terms are related to either rRNA metabolic process, ribosomal activity and small molecule biosynthetic process. In contrast to upregulated transcriptional activity in cluster 3 (*i.e.* decrease of negative transcriptional regulation) during earlier time points, proteins in cluster 5 seem to regulate this process and decrease and counteract this activity after 24 h.

In conclusion, we would show that the developed approach is capable of deciphering the proteomic response in previously unreached detail and can in principle be used for the detailed analysis of the following scientific questions regarding the wild type phosphoproteome and especially for the elucidation of the response in the adapted phenotype of the Hog1 lof mutants.

3.2.3 Phosphopeptide results

From the proteolytic digest that was used for proteome analysis, an aliquot was used for phosphopeptide analysis as described in [26]. As shown in figure 3.19, on average a high number of phosphopeptides of over 10 000 in most samples could be identified. Across all samples, 29 494 unique phosphopeptides could be identified. Similar to the proteome results, the reproducibility of the phosphopeptide counts is reduced for the not adapted LOF Hog1 deletion mutant. It also underlines, that this observation was not a technical issue, but rather sample related. The Pearson correlation of the phosphopeptides only show a generally lower value compared to proteome, as on protein level multiple peptide level information is merged and abundance values are leveled out across the samples. Thus, on peptide level the variability is increased. Still, the correlation within the sample groups is high with obvious differences to the other sample groups. In the most recent phosphoproteomics study of *M.oryzae* from 2015 by W.L. Franck *et al.* in the group of R.A. Dean [165], the number of phosphopeptide was not reported but the number of phosphosites identified was 4894. In our dataset, we were able to identify 45 291 phosphorylated sites. In addition to that, the gradient length for the LC-MS/MS analysis

was 3 h long compared to 1 h in our study. Although our workflow requires 4-fold more protein material for the enrichment, the overall data completeness is presumably higher as DIA instead of DDA was used for data acquisition. In conclusion, although no further quality measures were reposted in the study of 2015, the presented data in our study is not only competitive to the published results but exceed the information wealth by far.

For in depth assessment of the quality attributes of the phosphopeptide results the overlap, data completeness as well as the TIC reproducibility have been calculated as shown in figure 3.20 A, B and C respectively. Exemplary for the WT sample at 1440 min, the overlap for all four biological replicates was around 40 %. Considering also peptides that were identified at least twice (*i.e.* in 50 % of the samples), the overlap increases up to 78 %. The data completeness plot in figure 3.20 B confirms this observation on precursor level, also including non-modified peptides. Compared to the proteome data completeness plot, where protein level is shown, the slope of the missing value / identification relation is increased, indicating a larger number of missing values. The reproducibility of the data is thus of greater importance. A measure to describe the reproducibility of the sample sets, is the observed variance of the TIC, which is shown in figure 3.20 C. While the quality control HeLa runs show a CV of 12 % on the calculated TIC Areas from Skyline, the samples sets show a very similar range of CV such as 15 % to 21 % for WT, irreverisbly adapted and reversibly adapted while the loss of function mutant suffers from high variability of around 40 %. As discussed before, this variability is likely caused by the unspecific protease activity upon cell death. In general, the values are in good agreement compared to the QC runs, which reflects the technical variability to be expected. As no injection volume adjustment has been done, the expexted addition to this variability is nearly indistinguishable from the technical variability, which underlines the robustness of the developed sample preparation procedure in general.

Nevertheless, missing values still cause problems in further statistical testing, thus two strategies have been evaluated to reduce their influence on the result. As a prerequisite, it is expected that on average the overall peptide abundance does not change, thus the sum of all peptide abundances should be constant throughout all samples. Their variability serves as second measure in addition to the TIC area for the evaluation, as summarized in figure 3.21. While in the first step, all appearing phosphopeptide abundances (regard-

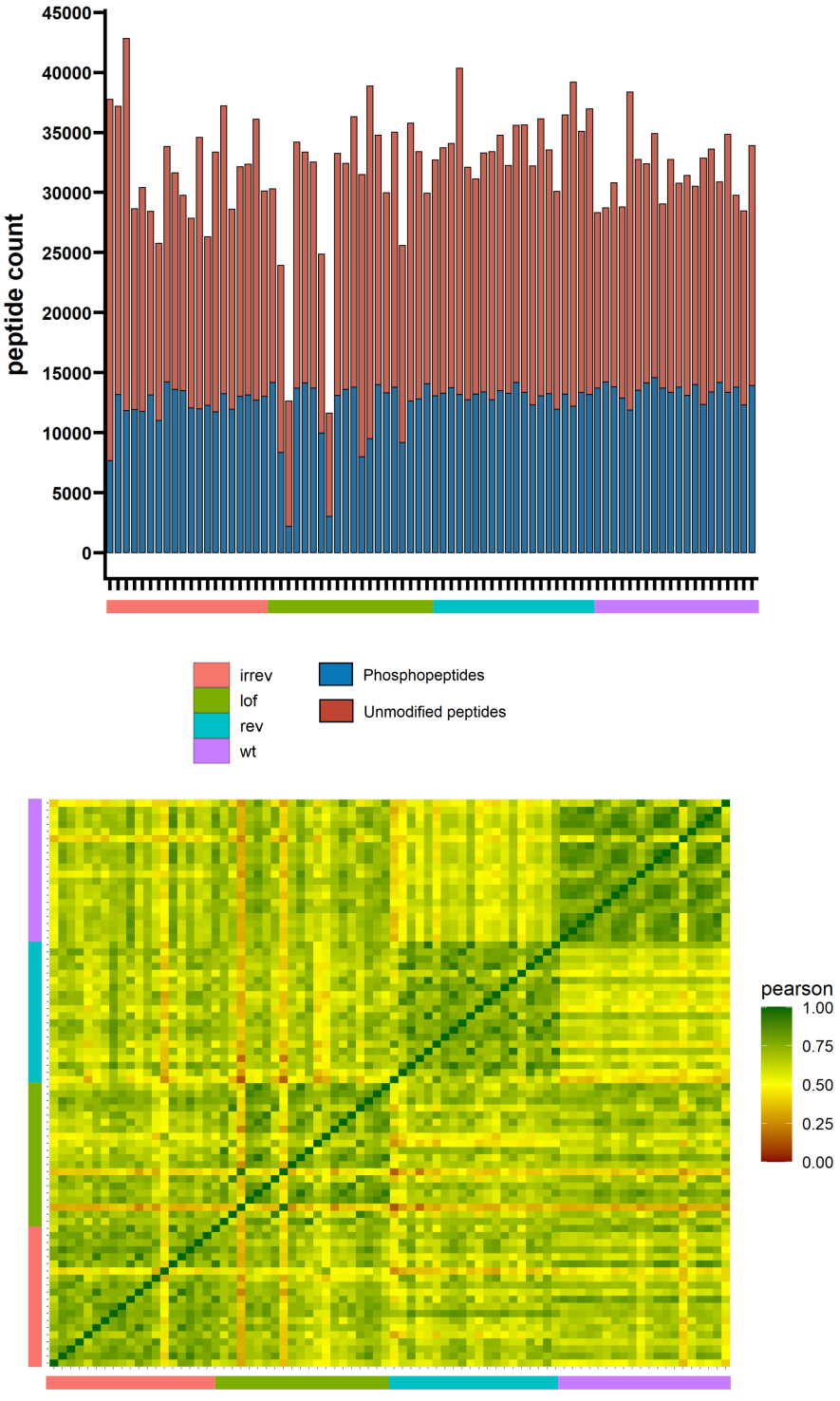

Figure 3.19: Phosphopeptide identifications and Pearson correlation of their quantitative abundance of *M. oryzae* of the sample set presented in figure 3.8. The number of phosphopepties are shown in blue, not modified peptides in red. Each color code at the bottom represents one sample type of the following: irreversibly adapted (red), loss-of-function (green), reversibly adapted (blue) and wild type (purple). Each bar represents one measured sample of the biological quadruplicates side by side, increasing time points from left to right (0 min - control, 10 min, 60 min, 4 h, 24 h)

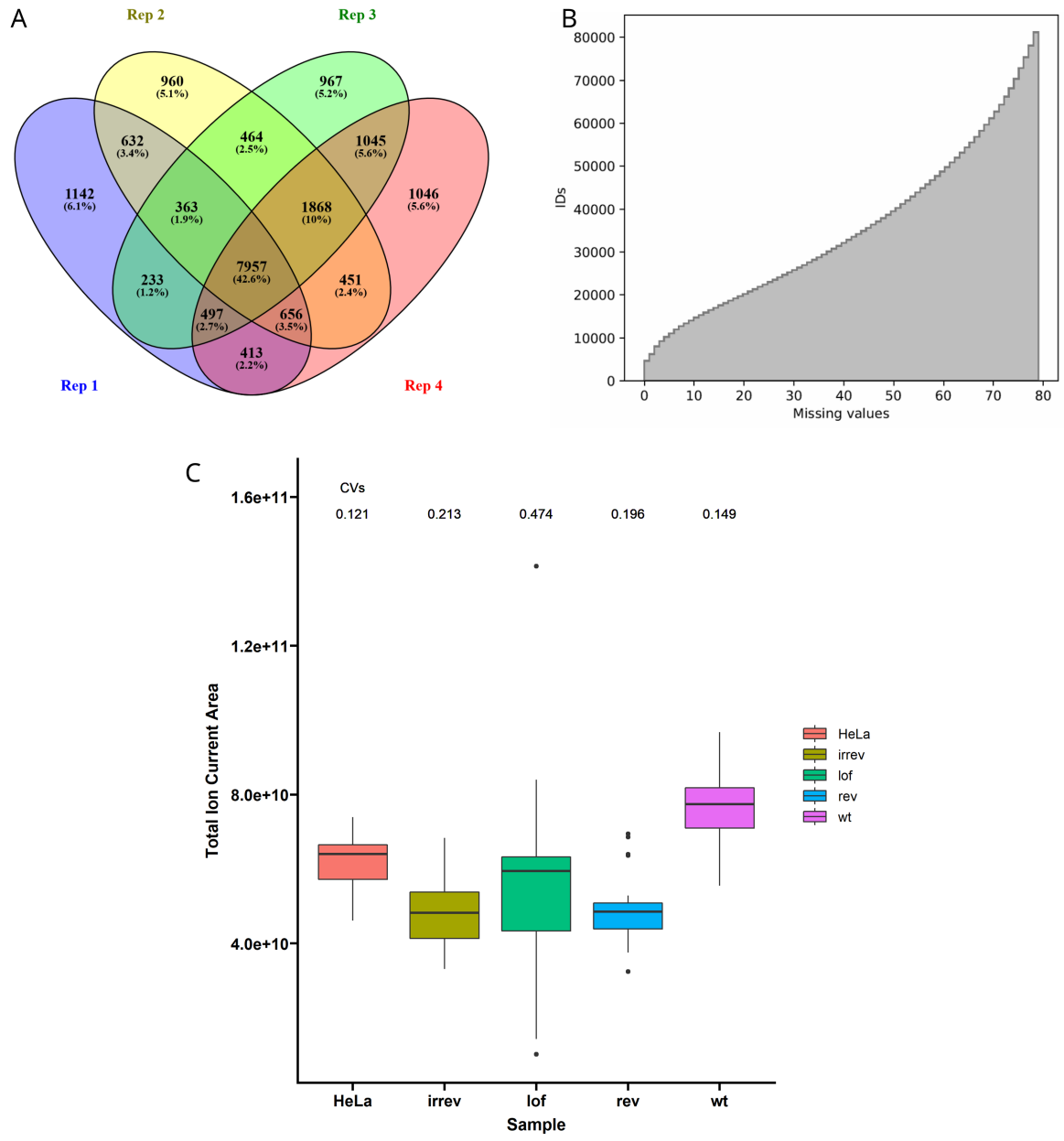

Figure 3.20: Measures of data quality of the sample set presented in figure 3.19. A) Overlap of the four biological replicates from WT 1440 min time point B) data completeness and C) variability of TIC

less their number of appearance wihtin the four biological replicates) were summed for each sample and the CV for each sample group was calculated. Then, the effect of VSN normalization and the combination of kNN imputation and VSN normalization was calculated. Compared to the raw values, VSN normalization reduces the variability by a factor of around 1.5 for the irreversibly adapted type up to 6 for the reversibly adapted type. The absolute value decreases by a factor of 10^5, as VSN normalization introduces a log2 transformation, that is anyway advised for statistical analysis by limma. with kNN imputation before normalization, this value can be further reduced to 1 % and less, but the absolute value increases by 2-fold. That indicates a relevant impact of the high number of missing values, as the imputed values contribute to 50 % of the evaluated values. As they are chosen by similarity, which is reduced in such a highly incomplete dataset, they will get imputed as very similar values. Thus, the observed variation in indeed a artificially introduced equalness of the data. Therefore, no missing value imputation is used before further statistical testing with. Accepting only peptides that appreared at least two times out of four biological replicates, surprisingly the obtained measures for the reproducibility are similar to the unfiltered dataset. But, as the number of missing values is decreased, the impact on the absolute sum of abundances of kNN imputation is reduced. The influence of missing value imputation on the CV is still high. Thus, the risk of artificially skewing the dataset is increased and therefore kNN imputation will not be used before statistical analysis. Nevertheless, the effect of kNN after VSN normalization is not shown here, as it is expected to introduce more noise and will skew the statistical testing with even increased impact.

For the subset of the wild type samples, the impact of missing value imputation on the abundance values has been investigated. For this, all peptides with at least 12 out of 20 identifications were selected which represents 60 % of all measurements. This way, the number of missing values is reduced and a possible bias in missing value imputation is potentially reduced. Indeed, the results in figure 3.22 show that the impact of the missing value imputation on the distribution of the peptide intensities is minimal. Two conclusions can be drawn here: First, the normalization of the processing software DIA-NN works already well, therefore VSN normalization does not change the distribution. It remains questionable, if this normalization provided by software is the optimal approach. But the likelihood that this approach is valid is high, as previous analysis (collaboration

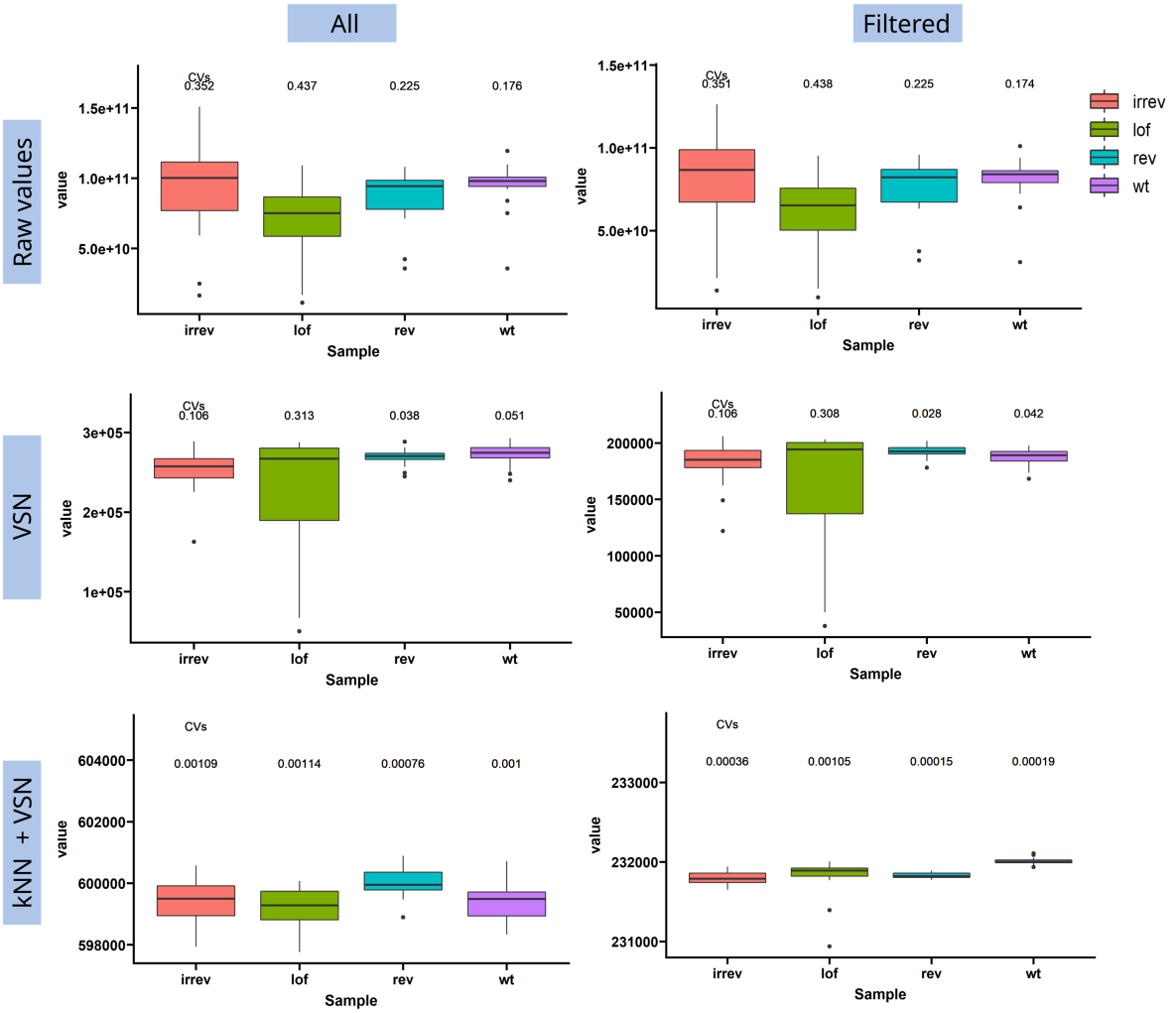

Figure 3.21: The sum of all phosphopeptide quantification values of the sample set presented in figure 3.19 is calculated for each phenotype dataset. Different data preprocessing strategies, such as VSN and the combination of kNN imputation and VSN are able to reduce the intra- and inter-dataset variability. On the left, the results for all observed values are included (All), no the right only peptide that were present in two out of four replicates (Filtered).

projects - data not shown) and the results from this study suggest the correct biological outcome, in case what has been published in the literature previously is correct. Second, the missing value imputation by SVD and kNN yields a similar shape for peptide intensity distribution. Thus, based solely on the shape of the distributions both methods are likely to be suitable for missing value imputation in this dataset.

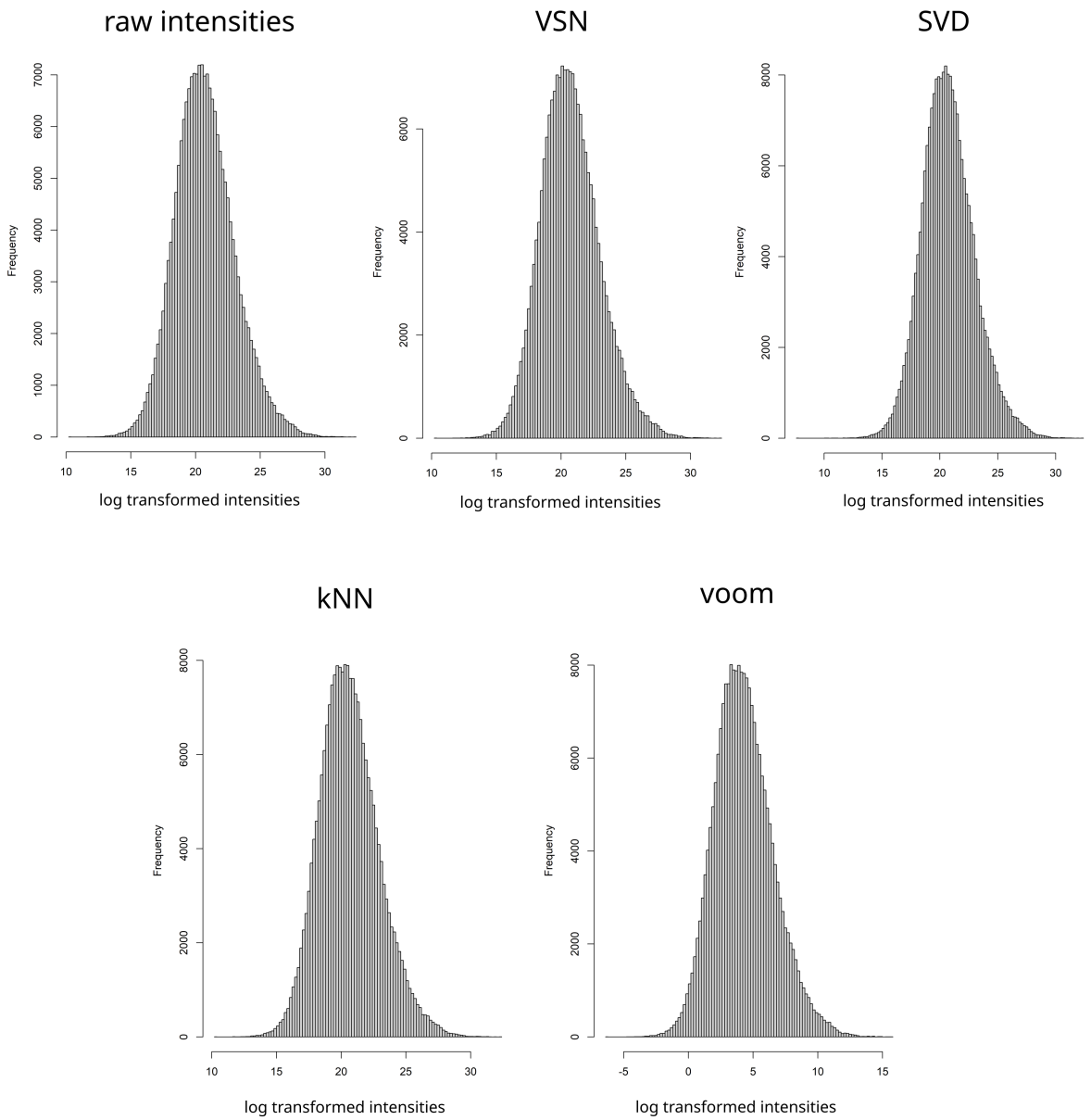

Figure 3.22: Histogram of phosphopeptide intensity values within the wild type samples only of the sample set presented in figure **??**. Evaluation of VSN, SVD, kNN and voom normalization and missing value imputation strategies do not significantly change the distribution of the intensity values compared to the raw intensities.

We could show here, that although the impact of missing value imputation on the peptide

intensity distribution is very low, the influence on the inter-sample variability is high shown by the CVs of sample specific sum of peptide intensities. Therefore, a high risk of artificially biasing the underlying dataset is still given and based on this the statistical testing is performed on a not imputed dataset. Nevertheless, many advanced multivariate statistical methods require data completeness and in case of phosphopeptides this has to be considered in the result interpretation and the value of the interpretation has to be verified by other means.

For the homogeneous proteome dataset, statistical testing by linear models using limma has been used. For phosphopeptide datasets, limma is only evaluated in one review publication [151]. Therefore, t-test and limma as described for the proteome data were both evaluated for statistical testing of the wild type 24 h samples compared to the control samples on phosphopeptide level and the results are shown in figure 3.23 A and B. The number of identified significant differentially abundant phosphopeptides defined by a fold-change of more than 2 or less than 0.5 in combination with an adjusted p-value of lower than 0.05 is lower in t-test (1312) compared to limma (1978). Interestingly, the calculated fold-changes do not seem to correlate well with the obtained adjusted p-value in case of t-test. The reason for this might be related to the assumption of equal variances in combination with a higher number of missing values, which influences p-value and fold-change respectively. In contrast to that, limma shows a typical volcano plot shape where a fold -changes and adjusted p-values show a correlation. This observation is in agreement with the aforementioned review [151], as the authors have shown that especially with a low number of replicates (or high number of missing values) linear models provide a better understanding of the dataset. In consequence, the overlap of significant instances is only around 37 %, but including already 70 % of all identifications from t-test. Considering these observations, limma was also selected for statistical testing of phosphopeptides.

PCA and hierarchical clustering was performed with equal conditions as for the proteome dataset, the results are shown in figure 3.24. Similar to the proteome samples, samples of time point 1440 min are fairly separated by principal component 1, whereas all other time points are homogeneously distributed. Furthermore, both principal components contribute to 50 % of the variability in the dataset, but also in combination with PC3 in two- and three-dimensional representation, no reasonable clustering of the samples can

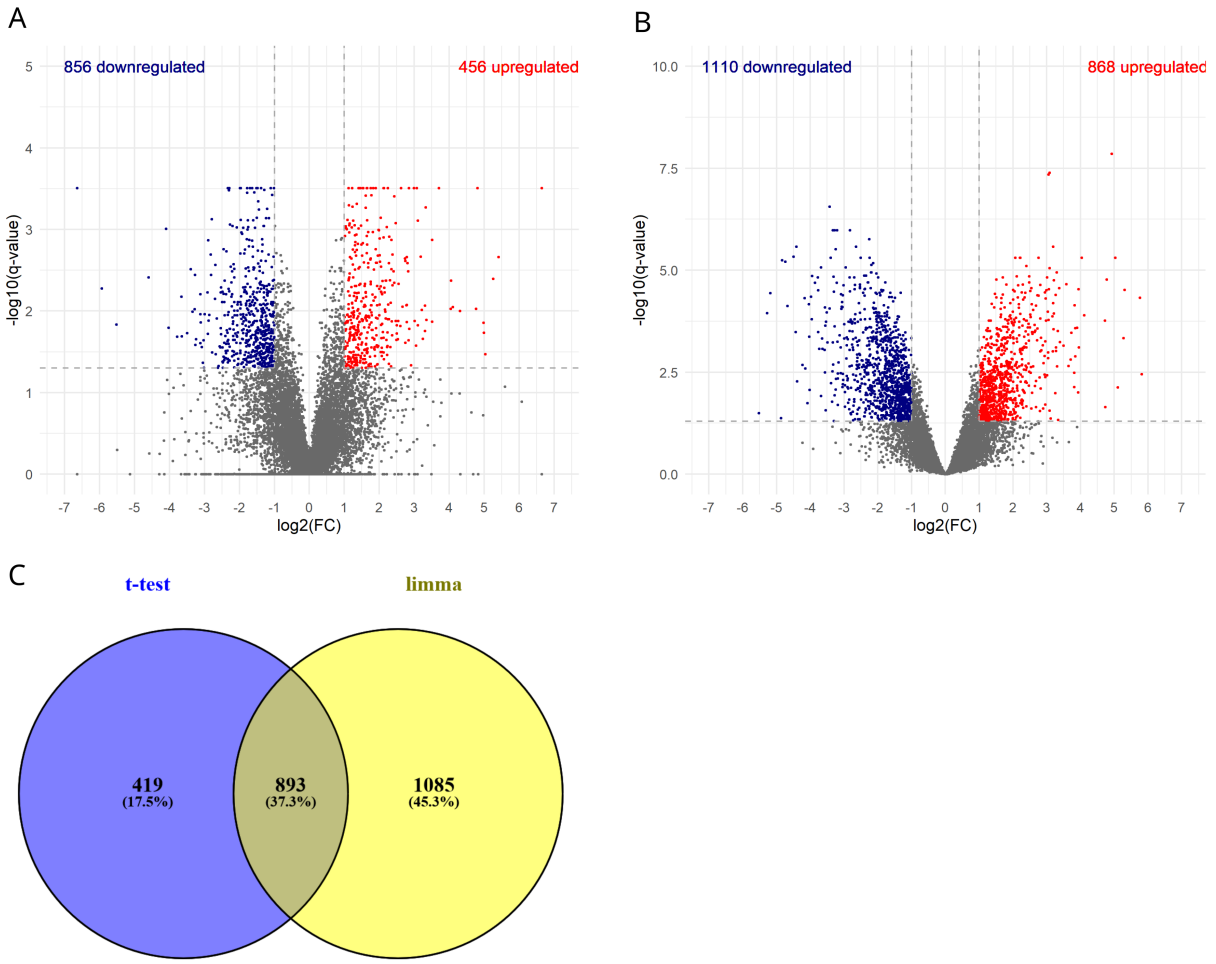

Figure 3.23: Comparison of statistical testing for the wild type phosphopeptide results 24h versus control time point of the sample set presented in figure 3.19. A) Student's t-test B) limma and C) the overlap of significantly changing instances from each test

be achieved. On the other hand, unsupervised hierarchical clustering indicates a good degree of similarity between the 240 min and 1440 min can be observed, although one replicate of the 1440 min sample group seems to suffer from a systematic deviation of all values. No technical reason could be identified, to exclude this sample from the analysis, thus it remained included in the analysis. In the earlier time point samples, three out of four replicates cluster well and one replicate respectively is clustered in the wrong sample group. Across the whole dataset, the clustering works reasonable and no obvious justification could be concluded to exclude samples from the analysis. Although single samples seem to correlate less with the sample groups, it is expected that the statistical testing using linear models is able to compensate such deviating abundance levels.

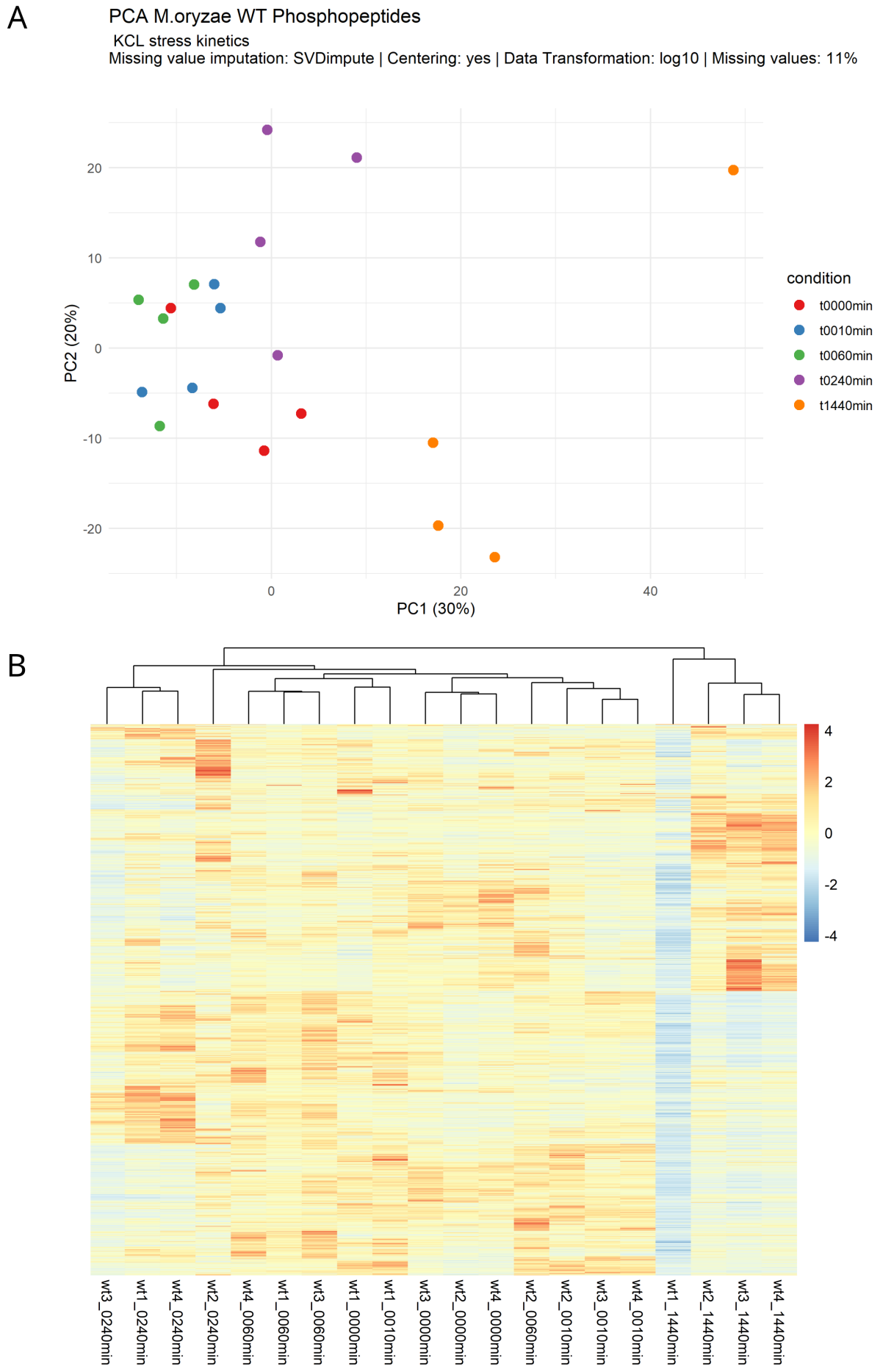

Figure 3.24: Multivariate statistical analysis of the sample set presented in figure 3.19. A) Principal component analysis B) heatmap for sample clustering as quality control of all wild type proteome results combined

3.2.4 Phosphosignaling in wild type upon KCl stress

The results of the statistical test is summarized in figure 3.25. Surprisingly, the change in significant phosphopeptides remains moderate for the first two timepoints, 10 min and 60min, and becomes reasonable for the later time points, 240 min and 1440 min. A rapid signaling response by phosphorylation has been shown by numerous publications [166, 167]. Therefore, most significant changes were expected in the early time points. As we can exclude mix up of samples for reasons already discussed, the cause for this observation is apparently a high variability between the reported phosphopeptide abundances, although normalization by VSN has been performed. As an example for such a case serves the phosphorylation of Hog1, shown in figure 3.26. Visual inspection would obviously identify a significant increase in signal intensity at 10 min, that is rapidly decreasing already at 60 min and nerly neglegible after several hours, which is in agreement with the commonly expected phosphorylation signaling patterns. limma suggests a fold change of 2.7 and p-value of 0.099 and adjusted p-value of 0.17, while the t-test provides 2.9, 0.011 and 0.26 respectively. Both results are very similar, which strengthens the hypothesis that both methods are suitable for the robust identification of unambiguous changing events. In both approaches, when considering only the p-value for identification of statistical relevance, this phosphorylation event would have been included in the positive hit list. Due to the FDR adjustment, it will not be considered at the first time point as significant, but for 240 min and 1440 min the results become statistically significant in both cases. In consequence, especially for small phosphoproteomics datasets, it remains questionable if FDR adjustment is beneficial. In the featured *M.oryzae* dataset, the absolute number of identifications across all samples is high, thus an FDR adjustment is required to prevent a substantial identification of false positives. This would lead to misleading results in downstream analysis, such as gene ontology enrichment. Based on this observation, for small datasets we recommend to consider the p-value and accept a possible increase in false positives, especially in discovery settings. In these cases, quality control and downstream analysis ensure to filter for less likely significant instances. In contrast to that, any case of proteomics experiments are typically more reproducible and robust as multiple peptides can be considered for abundance calculation. Thus, the distribution of p-values will be more precise and FDR adjustment clearly aids to prevent false positive identifications while maintaining a reasonable number of significant hits.

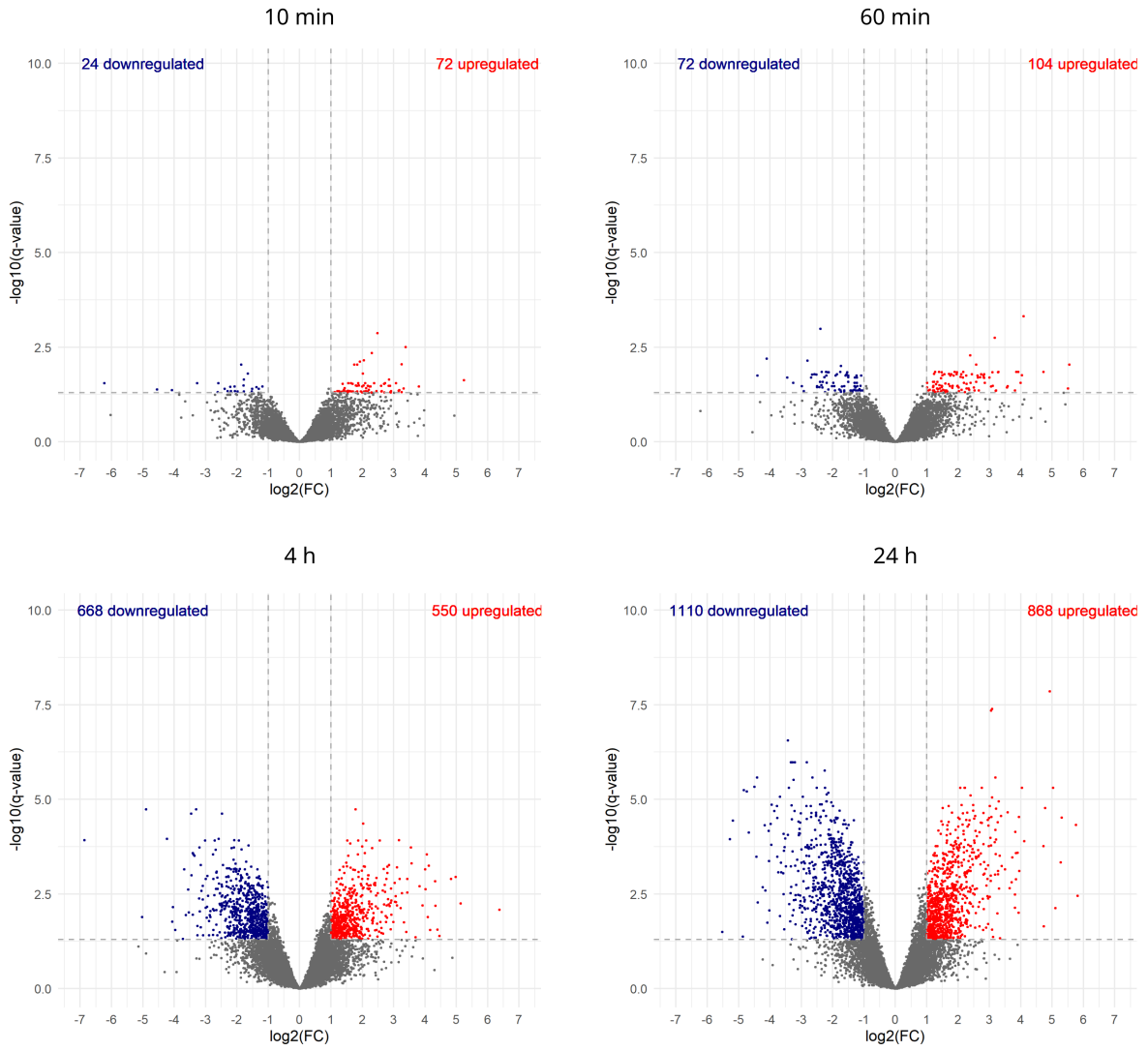

Figure 3.25: Volcano plots of changing phosphopeptide levels of *M.oryzae* during 24 h after osmotic stress of the sample set presented in figure 3.19 as obtained fold changes and q-values for each time point compared to the control

For an overview of signaling processes regardless their temporal role, all significant instances across all samples were submitted in ClueGO with standard settings considering the databases GO biological processes and KEGG for gene ontology enrichment and clustering resulting in a GO network shown in figure 3.27. Not suprisingly, protein phosphorylation is among the enriched GO terms, indicating that this approach yields reasonable terms. The GO term network reveals, that a large number of terms are related to cell cycle and its regulation which is reflected by the reduced growth of the fungus upon osmotic stress. Furthermore, the GO terms *phosphorelay signal transduction*, the large cluster of *signal transduction* and *regulation of cellular process* pinpoint to the well known MAP-

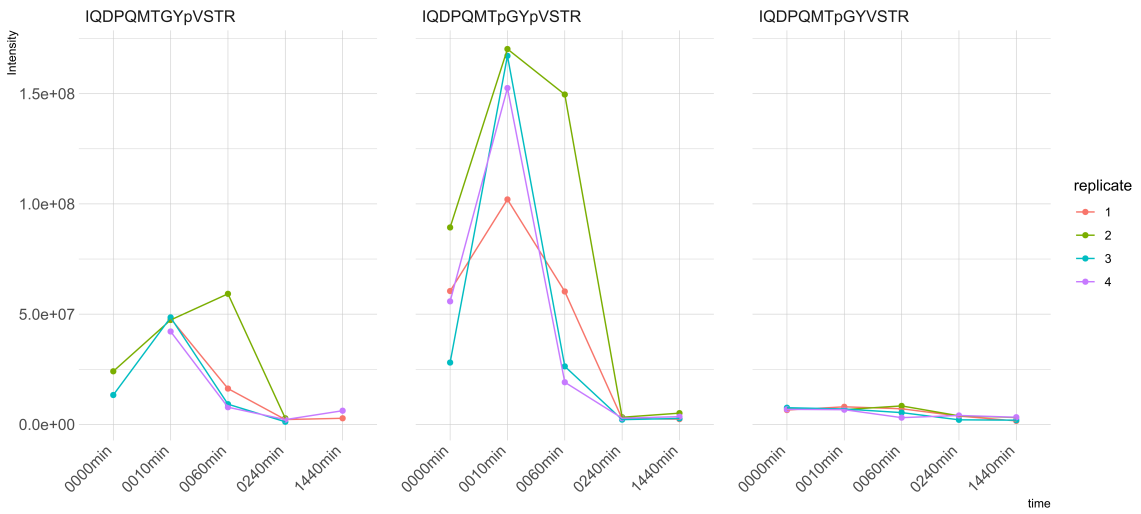

Figure 3.26: Temporal change of the dual phosphosites pY, pT and pY+pT of MoHog1 in four biological replicates upon salt stress of the sample set presented in figure 3.19 confirms the immediate response of the dual phosphorylation pY+pT. The pY only phosphosite also shows an upregulation, but with lower fold-change as the dual phosphosite pY+pT. The phosphosite pT does not show any immediate response, but decreases with low fold change at alter time points of more than 240 min.

Kinase pathways that are orchestrating cellular responses to extracellular stress, including the HOG pathway with proteins like MoSln1 (MGG_07312) and MoHIK1 (MGG_11174). To validate these findings, gene ontology enrichment analysis using STRING DB has been performed, that provided six KEGG pathways shown in figure 3.28 including also MAPK signaling pathway. As osmsostress response by MAPK kinase signaling has been identified in both approaches, the analytical approach has proven a suitable tool to identify/validate results from literature regarding osmostress response in *M. oryzae*.

3.2.5 Temporal changes in wild type upon KCl stress

Unsupervised clustering of the fold-changes has been applied to the dataset to obtain functional groups with temporal resolution. The use of different clustering algorithms has been already discussed for the proteome dataset, therefore k-means with an arbitrary value for has been applied to the phosphopeptide dataset. By visual inspection a k = 10 was identified to provide a reasonable trade-off between similar temporal shapes and the number of protein accessions for the phosphopeptides within the groups, as shown in figure 3.29. For each cluster, GO term enrichment with ClueGO has been analyzed and the GO network results are shown in supplementary figure 6.8 and following. The first

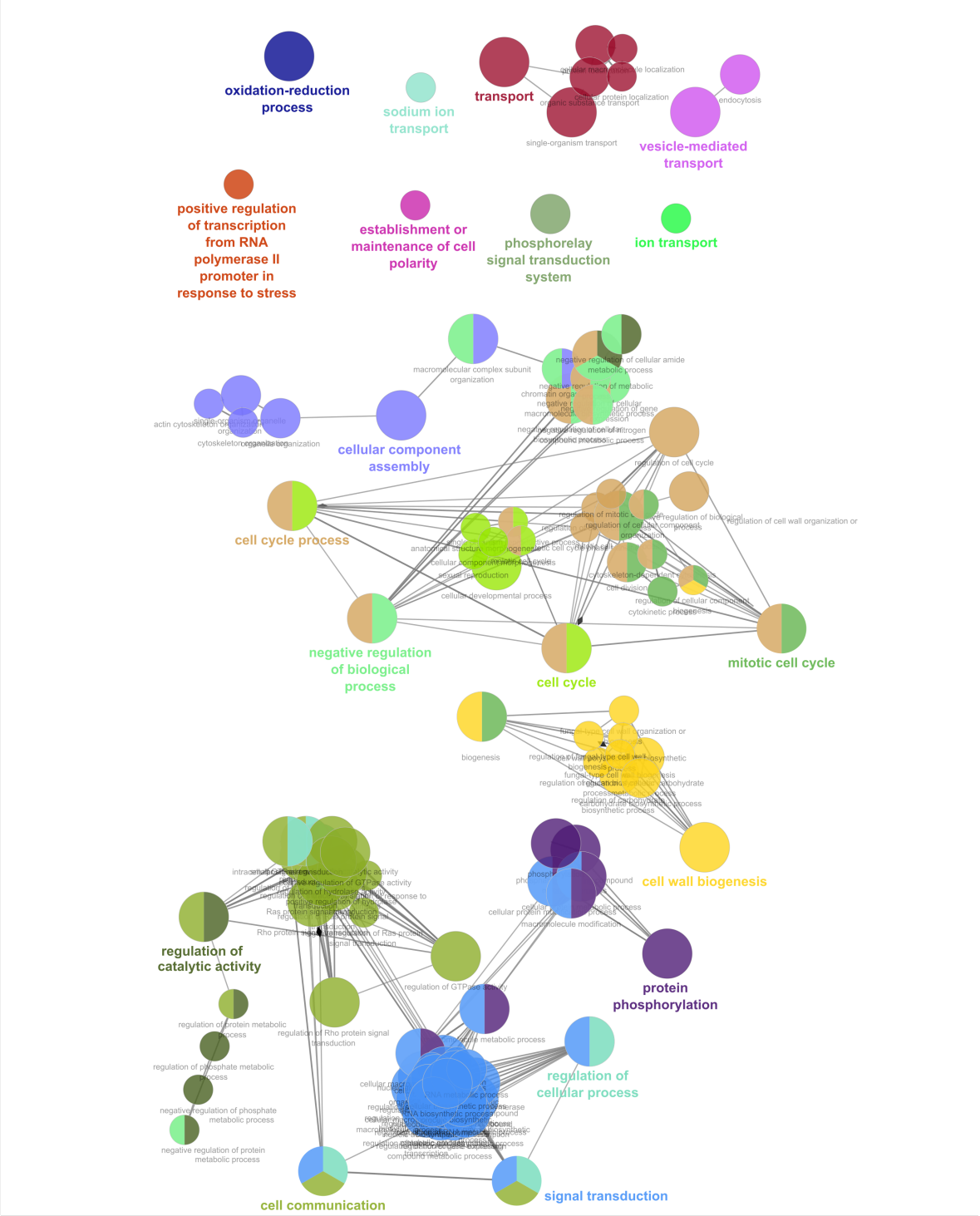

Figure 3.27: ClueGO network for all significantly changing phosphopeptide instances in wild type *M.oryzae* upon salt stress of the sample set presented in figure 3.19. Predominant processes include cell cycle, protein phosphorylation and signal transduction.

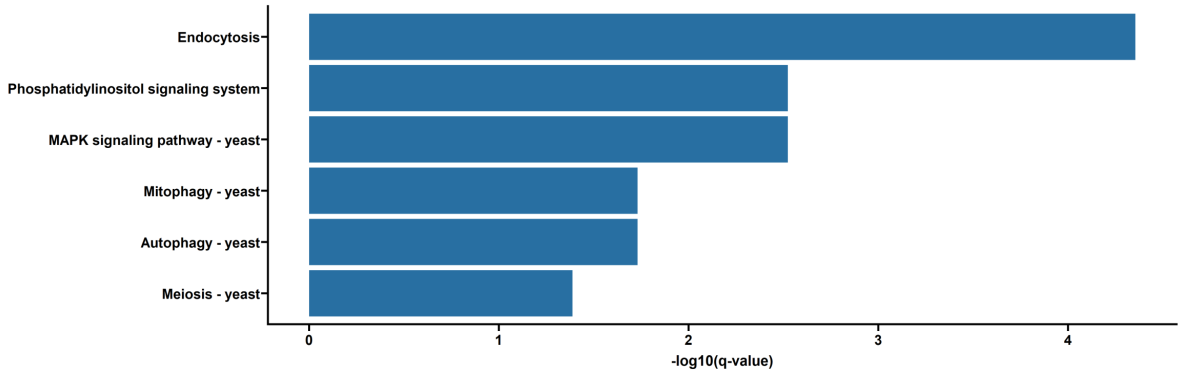

Figure 3.28: STRING DB GO term enrichment of KEGG pathways for all significantly changing phosphopeptide instances in wild type *M.oryzae* upon salt stress of the sample set presented in figure 3.19. Among the significantly enriched terms, MAPK signaling pathway is present, also including the HOG pathway. This finding validates our findings in general as they are in line with previously published and confirmed data.

cluster, with moderately upregulated entries, consists of phosphopeptides from proteins mainly involved in the cell cycle, also *negative regulation of cellular process*, which is in agreement with the observed phenotype. Also, proteins involved in *polyol biosynthetic process* in combination with *negative regulation of lipid biosynthetic process* can be found in this cluster. In addition to that, increased *positive regulation of hydrolase activity* is observed. Cluster two and three follow a similar temporal shape of proteins only upregulated in the last two time points 240 min and 1440 min. Most promising GO terms to explain the observed phenotype include *intracellular signal transduction* and *cellular carbohydrate biosynthetic process*. Interestingly, *lipid metabolic processes* is among the terms, which is in contrast to the observation in the first cluster. Also, *Ras protein signal transduction* appears in cluster three with three uncharacterized proteins (MGG_00928, MGG_03048, MGG_07310). Although it is known, that Ras signaling is coordinating stress response in yeast [168], so far it has been only descibed in the context of appressorium formation and plant infection in *M.oryzae* [169]. Thus, the role of these uncharacterized proteins in late osmostress signaling and response remains unclear. Ras signling reappears with three GO terms also in cluster four, where initial signaling is represented. This is also reflected by the identification of the term *intracellular signal transduction* which includes HOG pathway proteins such as Hik1 (MGG_11174) but also Ras signaling related proteins (MGG_09531, MGG_11425). Cluster five consists of phosphoproteins that are involved in *ion transmembrane transport*, that remain active compared to the control over all time

points. In cluster six only two proteins with no functional relation are clustered. The first is Glycogen synthase kinase 1 (Gsk1, MGG_12122) that is involved downstream of the MAP kinase Mps1 in the regulation of growth among others [170]. Its downregulation indicates, together with the enriched GO terms in the first cluster, a reduced growth agreeing with the observed phenotype. The second protein in this cluster is NADP-dependent malic enzyme (MGG_08173), which is presumably involved in the peroxisomal fatty acid metabolism. This observation is somewhat surprising, as fatty acid metabolism has been identified in previous clusters as upregulated over time. Cluster seven is the complementary part to cluster two and three, consisting of downregulated phosphoproteins only in the 1440 min time point. three major types of GO terms are clustered, first *regulation of GTPase activity* which includes *Rho-* and *Ras protein signal transduction*, second *organic substance biosynthetic process* and third *cell wall organization or biogenesis*, which is overlapping to a large degree with the results from cluster 4. Cluster five includes protein with an interesting temporal shape, they show a downregulation mainly at time point 240 min, which indicates a regulatory role. An indeed, proteins involved in regulation of *cell cycle*, *positive regulation of cellular process* and *positive regulation of GTPase activity* have been identified. In cluster nine, the temporal shape is increasing at first, but continuously decreasing over time. This indicates a role in immediate stress response, which is validated by the identification of HOG related proteins (*MAPK signaling pathway*), but also interestingly *Ras protein signal transduction* is also included. The last cluster is comprised of constantly downregulated phosphoproteins and include the well known *regulation of cell cycle* from previous cluster, but also interestingly *regulation of catalytic activity* that is in the same network with *Rho-* and *Ras protein signal transduction*.

In conclusion, the presented approach for proteome and phosphoproteome analysis has proven suitable for the identification of underlying processes in osmotic stress response. In proteome response, significant changes in pentose catabolic processes could be identified that explain the observed phenotype. In contrast to the well known and described HOG pathway in other yeast, such as *S.cerevisiae*, the accumulated osmolyte is arabitol instead of glycerol in *M.oryzae*. To our knowledge, this relation has not been shown in literature yet and might open the door for further upstream investigations for regulators and/or transcription factors that influence the expression of the proteins involved in pentose catabolic processes. Furthermore, it has been show many times that the HOG pathway

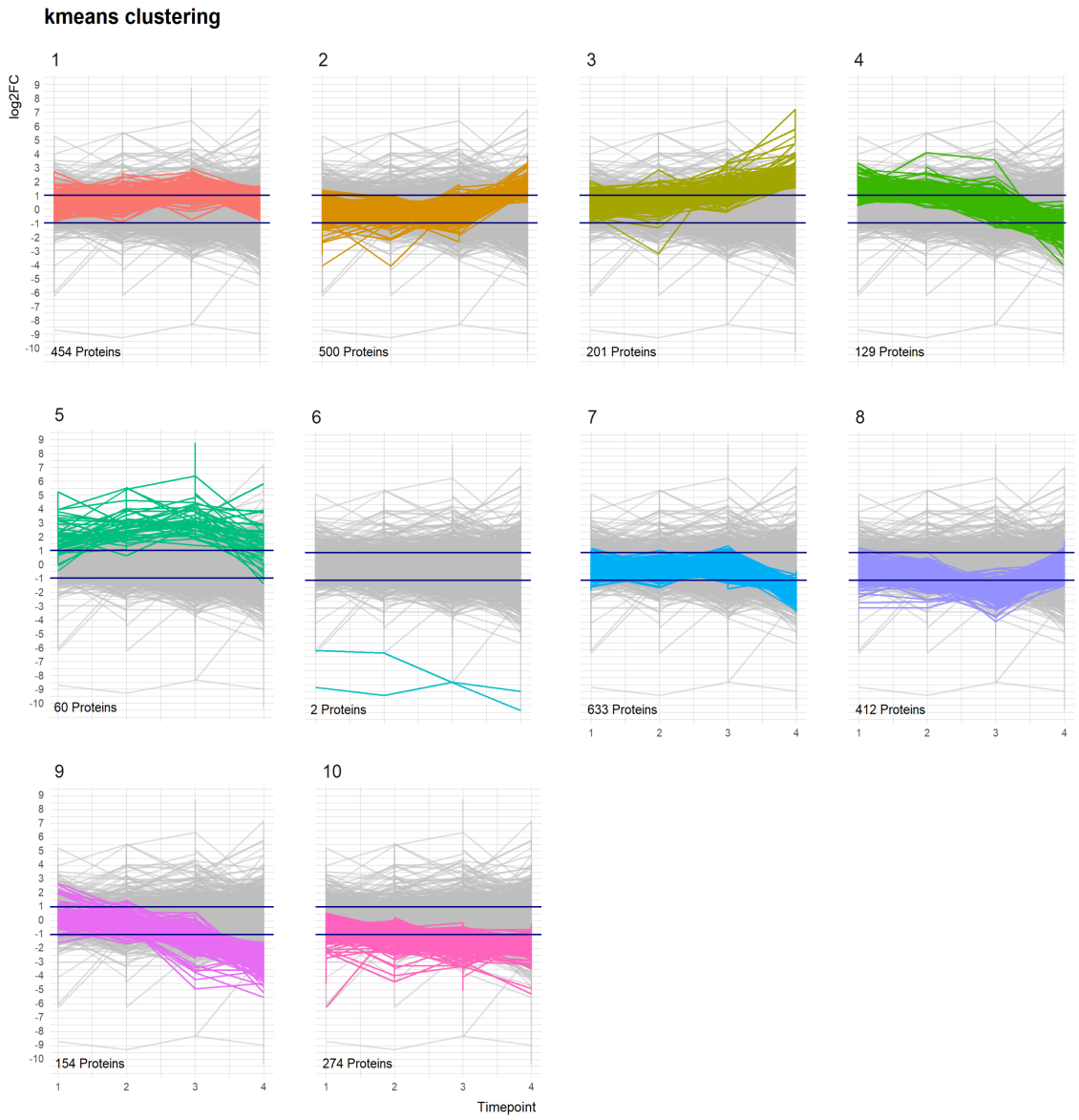

Figure 3.29: Temporal changes of phosphopeptide intensities of wild type *M.oryzae* of the sample set presented in figure 3.19 clustered by k-means clustering algorithm with arbitrary chosen k-cluster, displayed in k groups as spaghetti plot. On the y-axis the log2(fold change) is shown, while the x-axis shows each of the four time points as vertical lines. The commonly used threshold for significantly changing instances log2(fold change) = 1 and -1 are marked with blue horizontal lines for better identification of significant response clusters. Immediate responding phosphosites can be found in clusters 4 and 9, putatively regulating phosphosites in cluster 5 and phosphosites with functions in maintaining homeostatsis can maybe found in cluster 2 and 3.

is active during osmostress, but how this is connected to the production of arabitol is not comprehensively understood. The adaptability of HOG defective mutants suggest alternative pathways targeting the accumulation of glycerol might be present as alternative osmostress sensors/signaling. Interestingly, Rho- and Ras related signaling appears in many GO term enrichment analysis as up- and downregulated in several instances. To our knowledge, a contribution of such environmental signaling pathways have not been described in *M.oryzae* yet, but are known in other fungi to play roles in environmental sensing [171]. Thus, our phosphoproteomic study did not only validate the suitability of our approach, but also pinpoints to other active signaling pathways, that might be worth investigating with molecular biology strategies such as loss of function mutants and phenotype screening. In addition to that, previously undescribed lipid metabolism apprears to play a role in osmostress. Putatively, either as alternative source of energy as sugar components are required for osmolyte catabolism as has been shown in [172, 173], or as source itself for the catabolism of osmolytes. It has been reported in *C.albicans*, that Hog1 deficient mutants lead to an accumulation of lipid droplets under hyperosmotic stress, and with further experiments the authors concluded a crucial role of Hog1 in lipid homeostasis during salt stress [174]. In this publication, the authors also show that osmotic stress triggers the accumulation of peroxisomes in a Hog1 independent manner, namely through the GTPases Dnm1 and Vps1. This is well in agreement with our observation of Rho- and Ras dependent signaling. Having a comprehensive dataset at hand for the cellular response of *M.oryzae* to osmotic stress, differential analysis with Hog1 deficient genotypes is facilitated.

3.2.6 Altered protein and phosphopeptide response of wild type versus adapted Hog1 deletion mutants

The most interesting question in this study is related to the irreversibly (*i.e.* stable) adapted phenotype of the Hog1 loss-of-function genotype. What is general osmostress response in comparison with the wild type? And how does signaling change in comparison to the wild type?

The general response is represented by changes in the proteome. To validate the genotype, evidence for Hog1 could not be identified in this dataset, which indicates a successful

deletion of the gene from the genome. The proteomics results were analyzed using the same strategy as for the wild type samples and the results can be inspected in supplementary figures 6.5 and following. Similar to the wild type, proteome response is minimal until the last time point at 1440 min, where statistical testing by limma identified 399 downregulated and 475 upregulated proteins. The following k-means clustering and GO term enrichment analysis revealed similar terms, such as *glycan degradation* including *carbohydrate catabolic process* and also *Fatty acid degradation* in the network with *Peroxisome*. Furthermore, *Glycerolipid metabolism* appears, which has not been observed in the wild type. Interestingly, involved proteins in the GO term *Carbohydrate metabolic process* are related to enzymatic reactions with glycerol, which is the observed osmolyte in the irreversibly adapted phenotype. Foremost, the identified upregulated Dihydroxyacetone kinase (MGG_04014) is a promising candidate to link the observed phenotype to the proteome results.

On phosphopeptide level, a reasonable singificant change is observed already after 60 min, with no evidence of any phosphorylated Hog1 present. The same clustering procedure as for the wild type phosphopeptides has been applied. With this, in the first cluster that represents downregulated phosphoproteins over time, the GO term *phosphorelay signal transduction* is observed, which is representing the typical sensing of the osmotic stress. Interestingly, *MAPK signaling* and *carbohydrate metabolic process* are also found in this cluster. The involved proteins for MAPK signaling include flbA (MGG_14517) and Pmp1 (MGG_15140), which is a phosphatase that is known to play a role in appressorium formation through regulation of Pmk1 [175]. Pmk1 controls the glycerol accumulation for turgor generation to penetrate the plant surface during invasion. In this study, Thines *et al.* could also show that this process is Hog1 independent. Thus, an activation of this pathway serves as possible explanation of the phenotype through a possible downregulation of the phosphatase Pmp1. FlbA has also been shown to play a role in regulation of G protein coupled signaling during conidiation and appressorium formation [176, 13]. Based on this, a possibly involved Ras GTPase activating protein Smo1 might play a significant role, as it has been shown that it is essential for appressoria formation [177] through activation of the Ras2 protein. In addition to that, the deubiquitinase Ubp3 has also been shown to be crucial for appressoria formation [178]. In cluster three, comprising immediately and constantly upregulated phosphoproteins, ontologies involved in *regulation of*

hydrolase activity are found, but interestingly also *histone modification*, which includes the transcription initiation factor TFIID subunit 1 (MGG_01207). This transcription factor has been shown in *S.cerevisiae* to orchestrate cellular responses with identified interactors, which also include Ubp3 [179]. Remarkably, it has been shown that Ubp3 is a substrate for Hog1 and is essential for osmostress response [180]. Completing this picture, in cluster four to eight, that cover various diverse temporal profiles, two GO terms are prominently found in each cluster: *Ras protein signal transduction* and *GTPase activity*. Among many uncharacterized proteins assigned to the GO terms, Arf GTPase-activating protein (MGG_01472) serves as potential indicator towards involved pathways. In this protein familiy, Arl1, Cin4 and Gga1 have been shown to be involved in regulation of host penetration and invasive growth, which presumably requires glycerol as intracellular osmolyte [181].

Comparing all possible obtainable gene ontology terms of wild type osmostress response and irreversibly adapted Hog1 defective mutant, with no restriction in regard to statistical enrichment, the overlap is olny around 44 %. Interestingly, many GO terms that have been described in more detail, are overlapping between the wild type and adapted mutant, but in more detail they differ by their assigned protein entries. Thus, although Ras protein and GTPase activity related signaling is found in both genotypes, the identified proteins for the adapted mutant are more elusive to explain the phenotype.

3.2.7 Proteome and phosphoproteome analysis of not adapted Hog1 deletion mutants

It is known, that MoHog1 deletion mutants are sensitive to osmotic stress and are not able to grow under these conditions. Due to the high extracellular osmolarity, we hypothesize that intracellular water will eventually permeate the cell membrane or will be actively transported through it to establish osmotic homeostasis. Thus, the cells dry out and undergo any form of cell death. Eventually, intracellular proteins will be released, such as proteases, that alter the correct identification and quantification of proteins using LC-MS/MS based proteomics. In this dataset we could observe that the number of identified proteins significantly decreases in later time points, although very harsh lysis conditions were applied. This observation is in agreement with the aforementioned hypothesis.

Nevertheless, it is not known how Hog1 defective *M.oryzae* responds on cellular level upon this stress. Therefore, the proteomic analysis for the measured samples was performed ans summarized in supplementary figure 6.10 and following. The volcano plots show already a very different picture compared to all other sample types. The majority of significant changes is already observed at the 10 min time point, whereas the time points 60 min and 240 min do not shown a reasonable number of changes (< 100). The last time point again shows a very high number of significantly changing proteins. This observation is not in line with the expected time frame of proteomic response, that usually requires more than 10 min for transcription and translation. Thus, a possible mislabeling or misassignment of the samples has been checked and can be excluded as the reason for the unusual identified response. In addition to that, all other sample types show a reasonable and expected response on proteome and phosphoproteome level, which makes a sample mix-up more unlikely. Consequently, a similar result is observed for the phosphopeptide response. Gene ontology enrichment analysis reveals *Peroxisome* as one of the major GO terms on proteome level. This agrees very much with previously described metabolic response in this dataset for both, wild type and adapted mutant, and in the literature. Consequently, no processes related to carbohydrate metabolism or pentose phosphate pathways could be identified. On phosphopeptide level, GO terms related to *Cell cycle* and *MAPK signaling* were identified. In total, 21 phosphoproteins were included in the *MAPK* term, of which 14 phosphoproteins have also been found in the wild type samples as significantly changing. The remaining 7 proteins are comprised from two uncharacterized proteins, Sho1 and Its3, which haven been shown related to osmotic stress [182], and three proteins not obviously related to osmostress (MGG_01816, MGG_05207 and MGG_04325).

In conclusion, our data suggest that Hog1 defective not adapted *M.oryzae* mutants undergo cell death and release unspecific proteases to the extracellular matrix. Furthermore, no reasonable proteomic response could be identified, as the identifications are superimposed to the effects of cell and proteolysis. Nevertheless, peroxisome activity could also been shown in these samples, which is in line with previously descibed results. In addition to that, phosphoproteomics overlays very well with the response observed in the wild type. Interestingly, no gene ontology term specifically related to the induction of apoptosis could be found, rather the broad term *Cell cycle*. Still, apoptotic processes might be included in this term, but they do not appear significant among other cell cycle

related processes. But as the wild type response also includes cell cycle related processes, it seems unlikely that *M.oryzae* actively undergoes apoptosis.

The third observed phenotype for the Hog1 defective genotype has been demonstrated to regain osmostress regulation and effectively grow on stress medium, similar to the irreversibly and stably adapted phenotype, but looses this capability when re-cultivating it on stress medium after a cultivation period in unstressed condition. Interestingly, on proteome level no obvious protease activity could be observed, as shown in figure 3.8. This observation suggests, that the reversible adapted phenotype is more successful in preserving cellular integrity. In contrast to that, the response on proteome and phosphoproteome level is similar to the not adapted phenotype, as shown in supplementary figures 6.14. Although the gene ontology terms for proteome also include *carbohydrate metabolic process* and *glycan degradation*, which is rather similar to the wild type, the phosphoproteomic response is untypical and also the gene ontology terms are not informative, as shown in supplementary figure 6.16 and following. As well as for the not adapted phenotype, *MAPK signaling* is among the enriched gene ontology terms. For the reversibly adapted phenotype, this term is comprised of 20 proteins in total, of which all have also been identified in the not adapted phenotype.

Comparing both osmostress sensitive phenotypes, not adapted and reversibly adapted, the main difference seems to be the maintained cell integrity in the reversibly adapted phenotype. While the cellular proteome response of the reversibly adapted phenotype is similar to the wild type response, it seems to be mostly random in the not adapted phenotype. Interestingly, the phosphopeptide response is similar in numbers and gene ontology enrichment for both osmosensitive phenotypes. Although MAPK signaling has been identified, it does not show the characteristics of the adapted phenotype. Presumably, the presented proteomics data is not well suitable for the elucidation of ongoing processes, due to ongoing unspecific protease activity and unknown cellular processes caused by osmotic stress which might convolute the actual relevant response, especially in the reversibly adapted phenotype. Therefore, the reason why and how the adaptation processes can be reversed, remains unclear. In this case, shotgun proteomics might not be the optimal method to shed light on these questions. Apart from protein phosphorylation, numerous PTMs can govern cellular processes. As possible solution we suggest the investigation of

epigenetic processes in the nucleus, such as the analysis of histone modifications.

3.3 Phosphoproteomic profiling of HOS cell culture and clinical samples

3.3.1 Statistical analysis of the phosphoproteome

In the past, the separation of isobaric positional phosphoisomers by ion mobility spectrometry has been demonstrated [28]. To our knowledge, a comprehensive analysis of the relevance of ion mobility separation for phosphopeptides in a challenging dataset has not been shown so far. Therefore, a osteosarcoma sample set was measured with trapped ion mobility spectrometry separation before MS analysis, which is both included in the Bruker timsTOF Pro 2. Three osteosarcoma cell culture samples treated with Ceritinib were compared to three control samples, phosphopeptides enriched from 25 µg peptides by the optimized Zr^{4+}-IMAC protocol and measured in triplicate LC-MS/MS runs in DIA. Figure 3.30 shows the reproducible identification of around 4500 phosphopeptides while reaching an enrichment efficiency of around 50 %. The excellent reproducibility is also underlined by the phosphopeptide correlation as shown. The subsequent statistical testing revealed 30 upregulated and 69 downregulated phosphopeptides upon treatment identified by t-test with FDR adjusted p-value of below 0.05 and a log2 fold change of the median peptide quantities of less than -1 or at least 1. The following principal component analysis using SVD missing value imputation, centering and log10 transformation reveals a separation of the Ceritinib treated group compared to the control group with higher importance of PC1 (93 % variance explained) and minor, but complementing importance of PC2 (5 % variance explained) as shown in figure 3.31 A. This observation is exceptional, as phosphoproteomics datasets usualy suffer from high variabilities and high number of missing values (typically around 70 %), which both have been successfuly addressed in this dataset by an optimized phosphopeptide enrichment procedure and the use of DIA. Presumably, SVD based imputations seems to work well, although it has been reported to tolerate only up to 10 % missing values [183]. As principal component 1 provides reasonable explanation of the dataset, the top 10 absolute loadings from PC1 and 2 were extracted and summarized in table 3.3. StringDB protein-protein interaction analysis of these 10 top loading phosphoproteins allowing 1st and 2nd shell interactors as shown in figure 3.31 B) reveal an interconnection of many proteins through 2nd shell interactors. Gene ontology analysis in StringDB of these proteins and their interactors reveal the en-

richment of a potential role of TGFb signaling via the proteins NUP214, UBE2I, PIAS1 and SPTBN1, of which SPTBN is identified as significantly downregulated in our dataset which indicates a possible downregulation of immunosuppressive functions of the tumor cell. Furthermore, the top 10 loading proteins are involved in SUMOlyation processes as well as translocation, transcription and translation which indicates upregulation of apoptotic processes and cellular stress response.

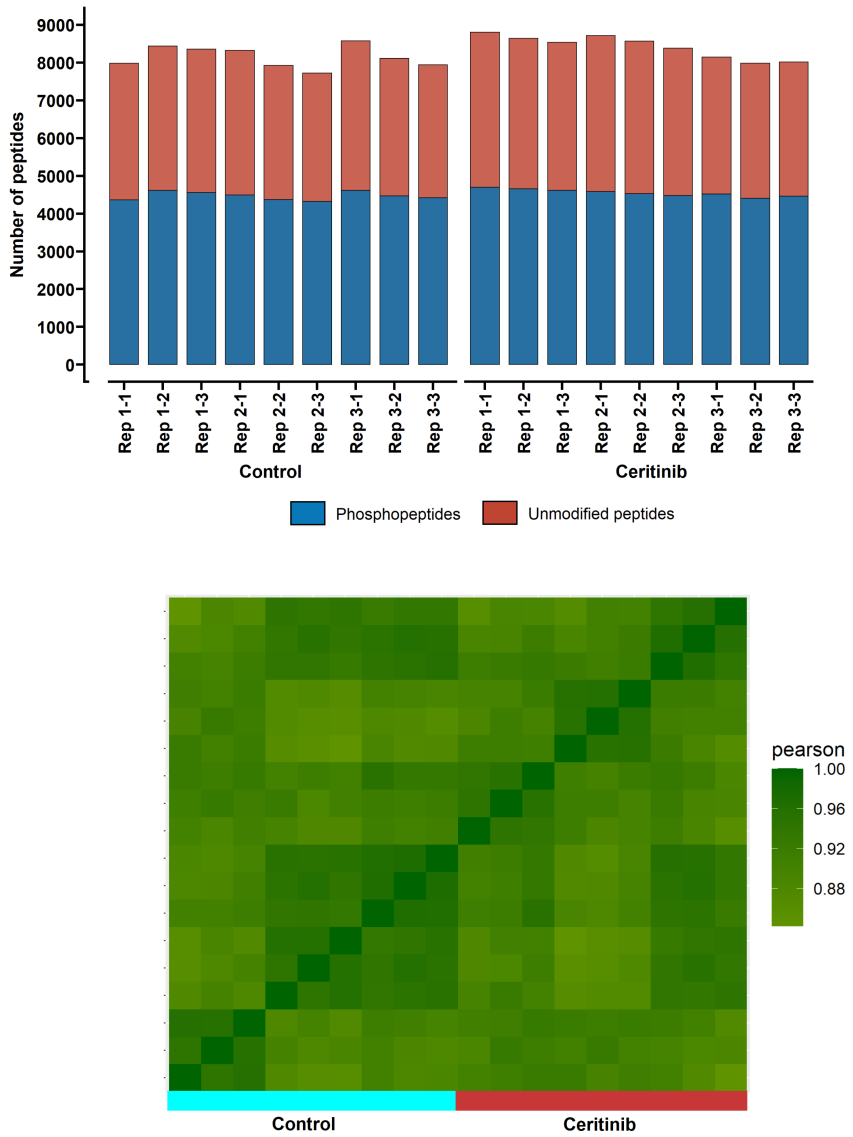

Figure 3.30: Three biological replicates of control group and Ceritinib treated cultured HOS cells. Phosphopeptides enriched from 25 µg trypsin digested protein. Each sample measured in triplicates on timsTOF Pro 2 in DIA mode. The number of identified phosphopeptides in blue and non phosphorylated peptides in red show at least 4500 identified phosphopeptides with a enrichment efficiency of around 50 %. The pearson correlation of phosphopeptide abundance is shown right, with good intra-sample reproducibility and no obvious outliner sample.

Phosphoprotein	Loading PC1	Uniprot description (excerpt)
NCBP1_HUMAN	-0.8315	Nuclear cap-binding protein subunit 1. Component of the cap-binding complex (CBC). NCBP1/CBP80 is required for cell growth and viability
UB2J1_HUMAN	0.5276	Ubiquitin-conjugating enzyme E2 J1. Catalyzes the covalent attachment of ubiquitin to other proteins. Part of recovery from ER stress. Plays a role in MAPKAPK2 dependent translational control of TNF-alpha synthesis.
RBP2_HUMAN	0.1637	E3 SUMO-protein ligase RanBP2. Recruits BICD2 to the nuclear envelope and cytoplasmic stacks of nuclear pore complex known as annulate lamellae during G2 phase of cell cycle.
SPTB2_HUMAN	-0.0434	Spectrin beta chain, non-erythrocytic 1. Candidate for the calcium-dependent movement of the cytoskeleton at the membrane.
HJURP_HUMAN	-0.0347	Holliday junction recognition protein. Incorporation and maintenance of histone H3-like variant CENPA at centromeres.
MILK1_HUMAN	-0.00901	MICAL-like protein 1. May be involved in a late step of receptor-mediated endocytosis regulating for instance endocytosed-EGF receptor trafficking.
SLIRP_HUMAN	-0.0064	SRA stem-loop-interacting RNA-binding protein, mitochondrial. RNA-binding protein that acts as a nuclear receptor co-repressor. Also able to repress glucocorticoid (GR), androgen (AR), thyroid (TR) and VDR-mediated transactivation.
LS14A_HUMAN	0.0061	Protein LSM14 homolog A. Essential for formation of P-bodies, cytoplasmic structures that provide storage sites for translationally inactive mRNAs and protect them from degradation. Acts as a repressor of mRNA translation.
BEND3_HUMAN	0.0043	BEN domain-containing protein 3. Transcriptional repressor which associates with the NoRC (nucleolar remodeling complex) complex and plays a key role in repressing rDNA transcription. The sumoylated form modulates the stability of the NoRC complex component BAZ2A/TIP5.
TGON2_HUMAN	0.0037	Trans-Golgi network integral membrane protein 2. May be involved in regulating membrane traffic to and from trans-Golgi network.

Table 3.3: Summary of top 10 absolute PC1 loadings in HOS cells treated with ceritinib. Proteins involved in TNFa sysnthesis, RNA processing and EGF receptor trafficking anre included in the list, indicating changes in relevant cancer related processes.

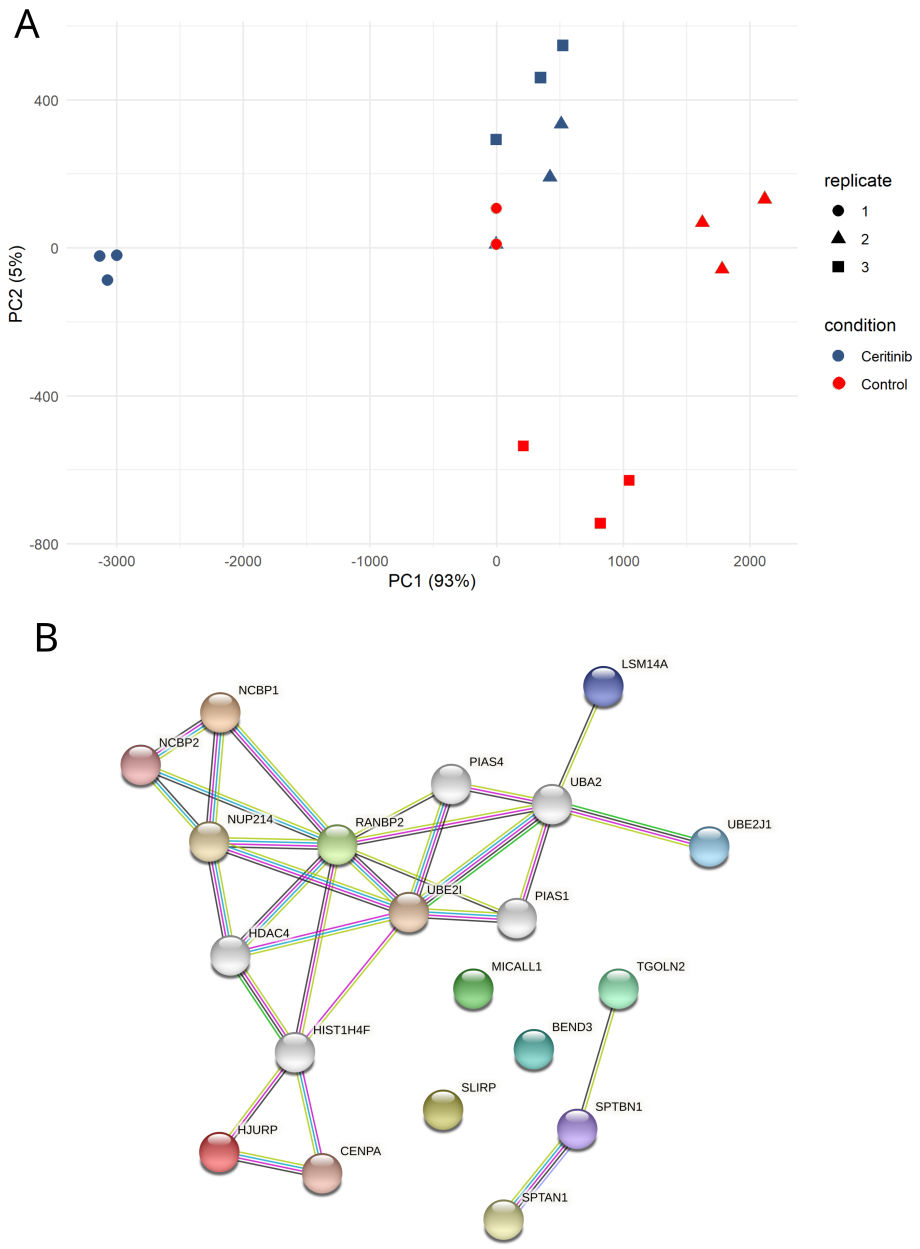

Figure 3.31: Analysis of phosphopeptides from Ceritinib treated versus control HOS cell culture samples of the sample set presented in figure 3.30: A) Multivariate analysis by principal components. Control samples and treated samples cluster together, separated over a combination of PC1 and PC2, which together account for 98 % of the observed variances. B) STRING protein interaction analysis of the top 10 loadings of PC1 and PC2 including first and second shell interactors shows not only separated proteins, but also proteins with known interaction. This indicates a functional relationship of those caused by the treatement.

It has been shown that Ceritinib acts (off-target) through inhibition of the insulin growth factor receptor (IGF1R) [184]. Indirect evidence for the inhibitory effect is provided by the significant downregulation of the phosphoprotein insulin receptor substrate 2 (IRS2). Irs2 is known to interact with p85 [185], that is also involved in the aforementioned TGFb signaling [186]. Thus, both pathways together might orchestrate the cellular response, a fact that, to our knowledge, has not been described before as identified per literature review and StringDB textmining, which opens the door for further hypotheses and research.

Gene ontology enrichment analysis of the protein accessions for the significantly differential phosphopeptides was performed including the GO databases Reactome [187], Wikipathways [188] and KEGG [189] and the results are summarized in figure 3.32 A. The Cytoscape plugin ClueGO with standard settings has been used for the enrichment analysis. Top terms include apoptotic processes and Rho GTPase cycle, with phosphopeptides involved in RhoB and RhoC GTPase cycle are mostly downregulated whereas phosphopeptides involved in RhoBTB are exclusively upregulated upon Ceritinib treatment. It has been demonstrated that identified proteins involved in RhoB and C GTPase cycle such as TJP1 and TJP2 are required for successful regulation of cell migration [190]. The upregulated RhoBTB proteins such as ACTN1, RBBP6 and RBMX have been shown to regulate cell motility [191], play a role as negative regulators of cell growth and are involved in apoptosis [192] and act as tumor suppressor by promoting the expression of TXNIP [193].

Furthermore, kinase and substrate enrichment analysis (KSEA) was perfomed using KSEApp. In principle, evidence for phosphorylated instances downstream and upstream of a specific kinase is collected and compared to the statistically expected number of evidence. Thus a p-value and q-value can be provided for hypo- and hyperactivity of these kinases, described by a z-score as shown in figure 3.32 B. As background sources for this analysis either curated entries only (from PhosphositePlus) or including also predicted kinase/substrate relations (NetworKIN) can be used. Here, both curated and predicted kinase-substrate relations were used as background for the knowledge extraction and statistical test with a q-value cut off of 0.05. Strong evidence could be identified for hyperactivity of CLK1 upon Ceritinib treatment, whereas six kinases show string evidence for hypoacitivity including cancer relevant kinases such as AKT1 and GSK3A and B. The downregulated activity of

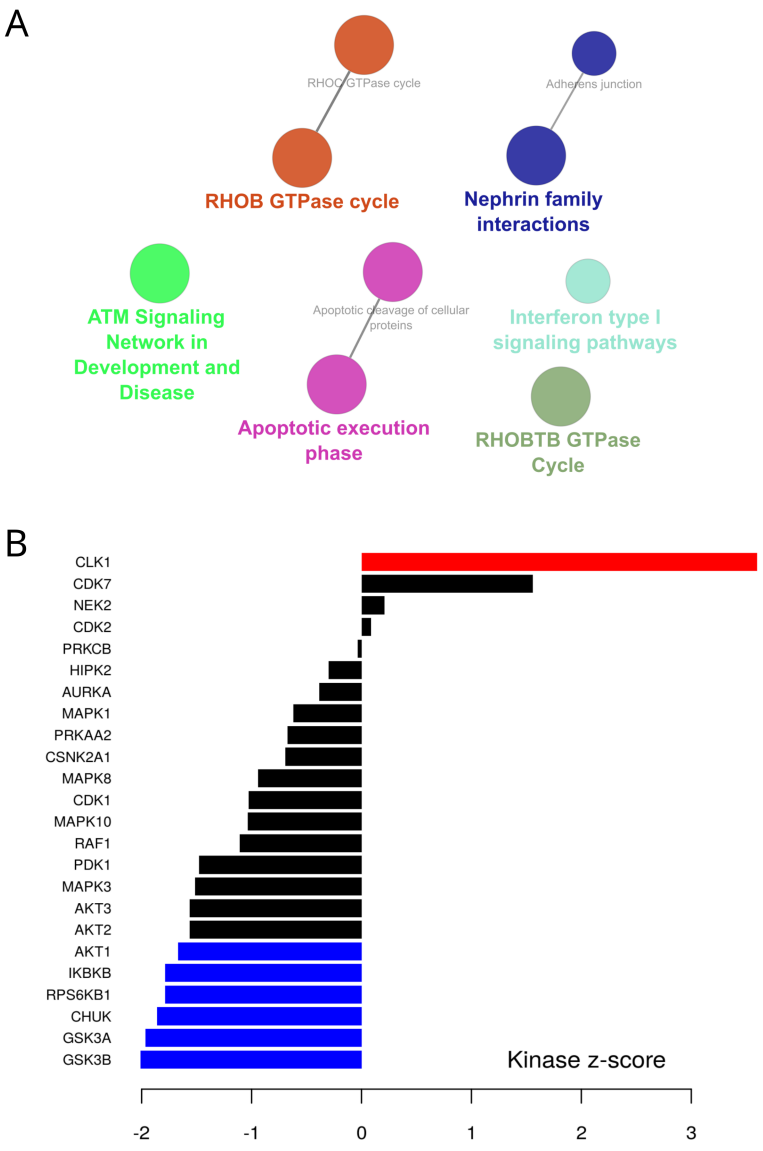

Figure 3.32: A) ClueGO analysis of all significantly changing phosphoproteins after Ceritinib treatment in HOS cell culture compared to not treated control samples of the sample set presented in figure 3.30. B) Kinase substrate enrichment analysis

AKT and subsequently the downstream GSKs has been described and discussed already in the literature [194], which validates our proteomics approach and KSEA as powerful tool for the correct identification of biological effects. On the contrary, CLK1 has been described in a review analysis to be effectively inhibited by Ceritinib treatment [195]. Although referring to secondary literature [196], there and in the featured review no evidence for this statement is given and further literature research did not provide more reliable information about the inhibition or activation of CLK1. Thus, their finding remains questionable. It has been shown, that CLK1 activity is required for the activation of PTP-1B [197] which itself is known to activate SRC induced oncogenic phenotype [198]. This is in line with the observed clinical phenotype at the Children's Hospital at the University Medical Center in Mainz, where a combination treatment of Ceritinib (as IGF1R inhibitor) and Dasartinib (as SRC inhibitor) proved a valid concept in the successful treatment of osteosarcoma in cell culture and in the clinic [22].

3.3.2 Successful validation of Ceritinib effects from low amount phosphoproteomics

The aim of Ceritinib treatment is the induction of apoptosis and the suppression of a migratory phenotype of the osteosarcoma cell. The featured phosphoproteomic analysis from as low as 25 µg starting material proves that both desired responses are achieved upon treatment. Although PCA analysis is often performed in proteomic studies, its capabilities are often not fully utilized. Here, we could show that the main loadings of the PC1 in this datasets aid the actualy discovery purpose of proteomics experiments. We identified TGFb signaling as putatively involved pathway, that shares components (p85) with the previously described off-target use for Ceritinib by inhibition of the IGF1R signaling. Thus, the use of the PCA loadings proved as another useful tool in the toolbox of bioinformatic analysis that is not widely used in the field. Further validation with small controlled experiments (*i.e.* known outcomes) have to be conducted to proof accuracy and sensitivity of this approach, such as stimulation of cancer cell lines with growth factors. In addition to that, we could validate our approach by confirming already described cellular responses of Ceritinib treatment. Therefore, we are confident to use this approach for profiling of individual clinical samples to characterize the responses to drug treatment and counteract bypass mechanisms (such as Src activation through CLK1 hyperactivity),

which serves as a step further in personalized cancer medicine.

3.3.3 Characerization of isobaric phosphopeptide separation by IMS

It has been shown in the past, that isobaric phosphopeptide isomers can be separated ion mobility due to conformational changes in the gas phase introduced by the different phosphorylated sites [199]. So far, a large scale evaluation of the beneficial effect for separation of isobaric has not been discussed in the field. Therefore, all identified phosphopeptide pairs of the osteosarcoma dataset were filtered and chromatographic resolution was calculated for the ion mobility separation. DIA-NN unfortunately does not provide the actual IMS peak width, so that an average ion mobility peak width of 0.2 1/k0 was assumed, based on previous DDA runs from unambiguously identified phosphopeptides and used for the calculation of resolution. Out of 5909 identified phosphopeptides across all samples, 2936 unique phosphopeptide sequences were found to form 1639 co-eluting positional isomer pairs, defined as a difference in retention time less than the average peak width of 30 s. In separation sciences in general , a resolution of at least 1.5 is considered as baseline separated. But even with resolution below 1.5 a reasonable separation is provided, that is sufficient for the deconvolution and identification of positional phosphoisomers. Therefore, a arbitrary value of 0.5 as minimum resolution was used for filtering. In total, 316 co-eluting isomer pairs were successfuly separated by ion mobility, the relation summarized in figure 3.33 A, an example of the separation power is shown in figure 3.33 B. As obvious from the XIC, no separation of the phosphoisomer pair was achieved, while the ion mobility provides two distinct isobaric peaks, that have been identified as two phosphopeptide isomers.

Ogata *et al.* have claimed, that the conformational changes in gas phase influencing the ion mobility is caused by intramolecular interactions between the phosphogroup and amino acid with basic or acidic character [200]. In order to validate their observations and transfer it to the problem of isobaric co-eluting phosphopeptide isomers, peptide properties of all phosphoisomer pairs, regardless their chromatographic elution, were calculated using the *Peptide* package in R. In supplementary figure 6.18 the relative amount of amino acids with certain character (tiny, aromatic, basic, acidic *et cetera*) in the plain sequence

of the phosphopeptide pair is plotted as histogram for all co-eluting pairs with overlaid histograms for IMS separated and not separated pairs. Statistical significant correlations could not be observed for any peptide property, but a general trend is visible for some peptide properties. Taking all trends into account, a higher probability for sufficient IMS separation is given with a peptide of following amino acid properties: 10 % basic / low number of acidic residues / 15 % charged / 60 % polar / 40 % non-polar / no aromatic residues / high number of small residues / GRAVY index around -1. These findings would underline the hypothesis stated by Ogata *et al.*, but the correlations in the osteosarcoma dataset appears not to be significant. In fact, Ogata *et al.* also investigate the relative position of the phosphopeptide and proximity of the phosphogroup towards basic and acidic residues but the effect on ion mobility is rather marginal in their case and also, no statistical significance is provided. A similar analysis on the osteosarcoma dataset is not expected to provide deep insight into the underlying reasons for conformational changes of phosphopeptide isomers.

Rather, as possible outlook we suggest to also include a larger dataset for this analysis and calculate a model, that also takes the combinatorial effect of peptide properties into account. This way, robust statistical significance of the influence of peptide properties on the collisional cross section can be provided. Some properties, such as amino acid size (small/aromatic) have not been described yet to have an influence on the collisional cross section of phosphopeptides. Presumably, the relative influence of the phosphogroups on ion mobility is increased in presence of small amino acid residues. Consequently, in presence of bulky amino acid residues (*e.g.* aromatic residues) the resolution decreases as the collisional cross section might be mainly influenced by the size of the amino acids rather than the conformational changes by the phosphosite. With similar absolute effect of the phosphosite to the CCS, the relative difference decreases with increasing molecule size. Furthermore, the degrees of freedom for intramolecular movements will be decreased with an increased number of large amino acids in the peptide sequence. In addition to that, further improvements of data acquisition such as longer ramp time during the ion mobility separation will lead to increased sensitivity and resolution. This way, number of identifications and their confidence can be possibly boosted and serve as optimal acquisition for further low amount osteosarcoma studies also with clinical samples.

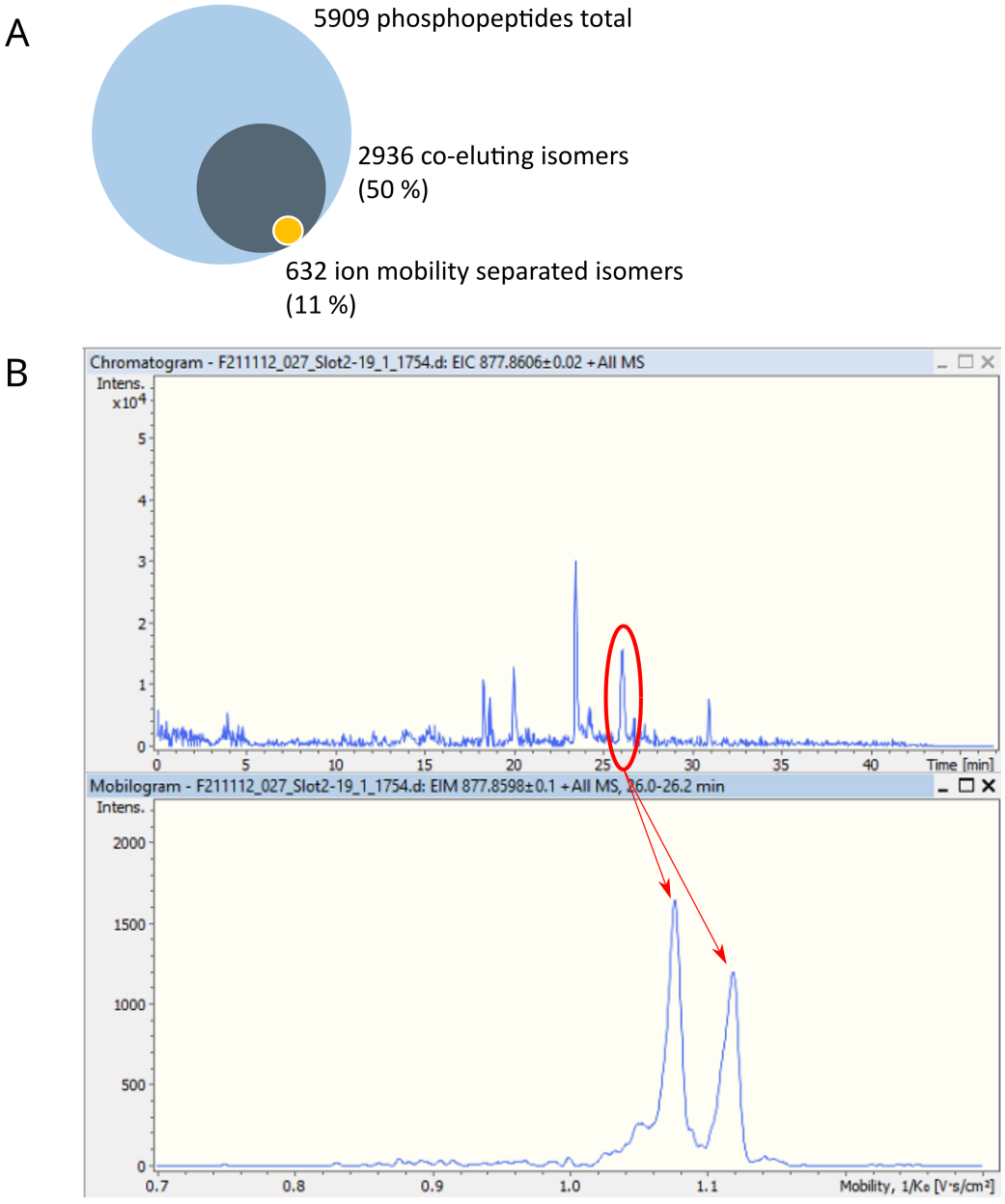

Figure 3.33: IMS enables separation on co-elution isobaric phosphopeptides of the sample set presented in figure 3.30 A) Number of identified co-eluting and ion mobilty separated isobaric phosphopeptide isomers with B) example of a chromatogram and corresponding ion mobilogram

3.4 Downscaling of mouse Th17 phosphoproteomics

3.4.1 Statistical analysis of the phosphoproteome

Downscaling of phospoproteomics workflows enables sample specific analysis of samples with very low amount of sample material available, such as tissues and sorted cells. Typically, phosphoproteomic analysis from *in-vivo* experiments (*e.g.* from mouse), does require to pool several biological replicates. This way, the desired protein and phosphoprotein abundances are homogenized and due to a expected high biological variability the desired effects become ambiguous. With the presented workflow it is now possible to analyze the phosphoproteomic response mouse specific, without pooling the samples prior to phosphopeptide enrichment. Nevertheless, it has not been shown so far in the literature, that the use of low input amount would lead to the very same biological conclusion as a typical phosphoproteomics experiment. Thus, isolated naïve T cells from mice with either Casein Kinase II (CKII) knock-out or wild type (WT) were differentiated into Th17 cells. The cells were cultivated to obtain at least 1000 µg protein and were enriched by TiO_2. Sample preparation, raw data acquisition and processing as well as statistical testing was performed as part of the core-facility work, that will not be discussed here. Result tables with fold-changes and FDR adjusted p-values were provided for comparison in this analysis.

In the context of low amount phosphopeptide enrichment optimization, supernatant of the cell lysate corresponding to 25 µg were used for tryptic digest and subsequent phosphopeptide enrichment with the presented protocol. The three biological replicates were measured in triplicates applying DIA mode and around 6500 phosphopeptides were identified per sample with an average enrichment efficiency of around 50 % for the entire sample set, while the treated samples consistently showed a sligthly higher phosphopeptide count, as shown in 3.34. This low enrichment efficiency is related to the use of DIA, as discussed before. The phosphopeptide correlation revealed a deviation of one biological replicate of the control group, compared to all other samples. A technical reason for this observation can be excluded, as the replicate measurements are consistent for each sample and the phosphopeptide identifications are reproducible for each sample, thus the reason might be either sample or sample preparation related. the fact, that all other samples show a good correlation, makes a failed sample preparation rather unlikely. Still, the number of

phosphopeptides is lower for all technical replicates of that specific control sample. The low correlation might also indicate a difference in cell differentiation, as no correlation could be observed with neither control nor treated sample group. Nevertheless, no obvious reason could be identified to exclude this sample from the analysis, thus it was still included in the dataset for statistical testing. Using t-test for the identification of statistically significant changes in phosphopeptide abundance, 137 up- and 193 downregulated instances could be identified, illustrated in figure 3.35 A. Surprisingly, also this dataset did show a reasonable clustering of the treated sample group separated to the control sample group based on principal component 1, accounting with 89 % to a high degree to the variability in the dataset, as shown in figure 3.35 B. The low level biological conclusion of the experiment is currently in peer review publication process and will be exhaustively discussed in the published literature. Here, only the high level outcome of the 1000 µg experiment is compared to the results from 25 µg.

3.4.2 Overlap with standard phosphoproteomics workflows

The overlap of the identified peptides (including modifications, *i.e.* phosphosite position) that were selected for statistical testing is 2185 peptides (14 %) while 6081 peptides (40 %) uniquely identified from 1000 µg and 6910 peptides (46 %) uniquely identified from 25 µg. The low overlap of peptide sequences has already been discussed in a previous dataset and the reason for this is the difference in acquisition method. Furthermore, the number of relevant phosphopeptides for statistical testing seems to be lower for the 1000 µg sample set, but the reason is simply a different filtering strategy that has been applied before testing. Still, the low overlap is not unexpected, as TiO_2 and Zr^{4+}-IMAC can have different phosphopeptide species affinities, so both methods preferably enrich peptides with differing physicochemical properties. Gene ontology enrichment of statistically significant changing phosphoproteins obtained from both enrichment strategies has been performed by ClueGO and summarized in figure 3.36. Gene ontology terms with no relevant clustering were differing for both enrichment strategies, GO terms related to SUMOylation. The main clusters in both experiments is related to cell cycle, which is in line with the expected biological outcome of the treatment. This underlines that although the nature of enriched phosphopeptides is different, changing phosphopeptides from the same related proteins can be enriched. The biological conclusion was identical in both

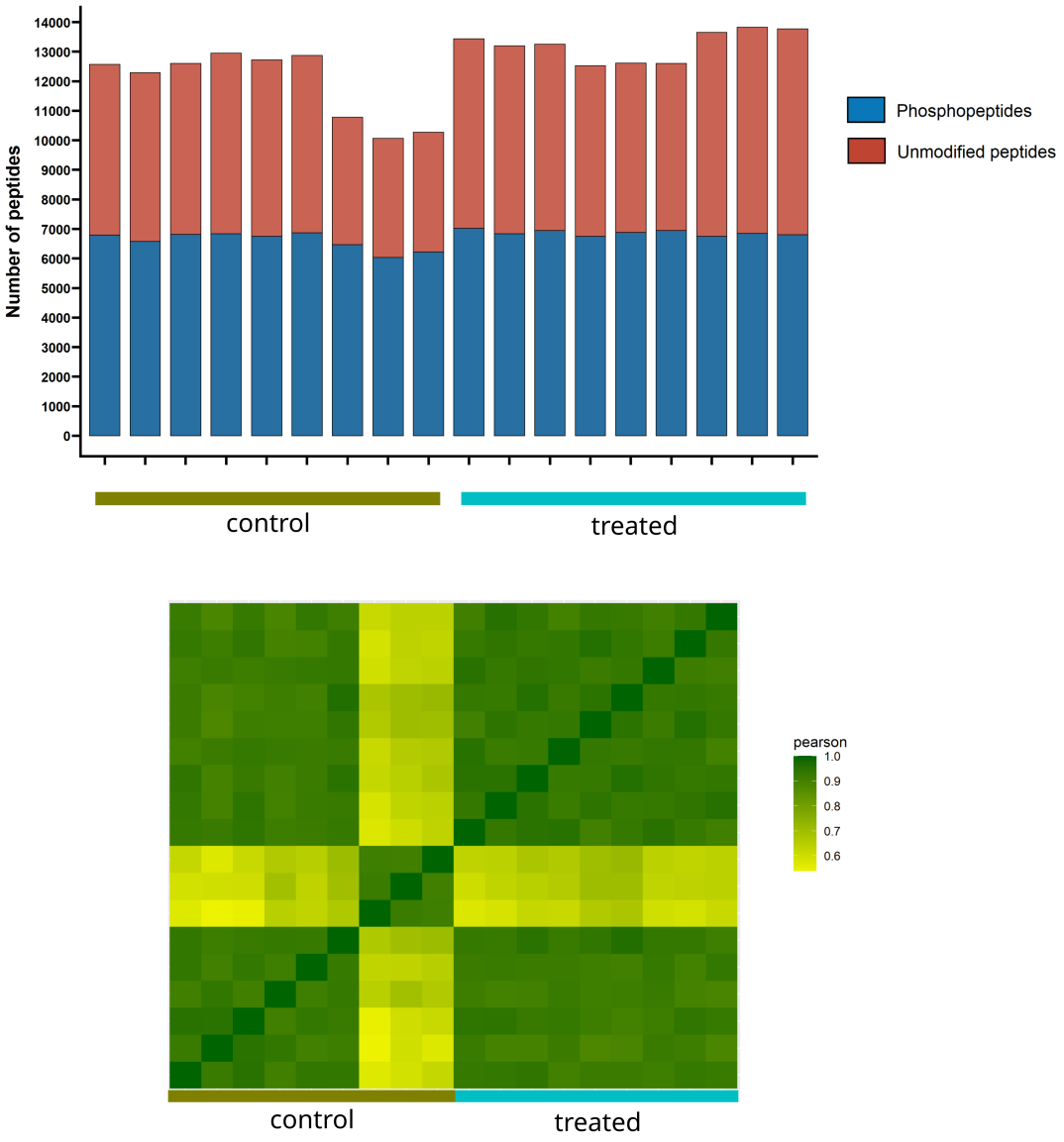

Figure 3.34: Three biological replicates of control group and treated cultured murine Th17 cells. Phosphopeptides enriched from 25 µg trypsin digested protein. Each sample measured in triplicates on timsTOF SCP in DIA mode. The number of identified phosphopeptides in blue and non phosphorylated peptides in red show at least 6500 identified phosphopeptides with a enrichment efficiency of around 50 %. The pearson correlation of phosphopeptide abundance is shown right, with good intra-sample reproducibility, except for sample 3 of the control group.

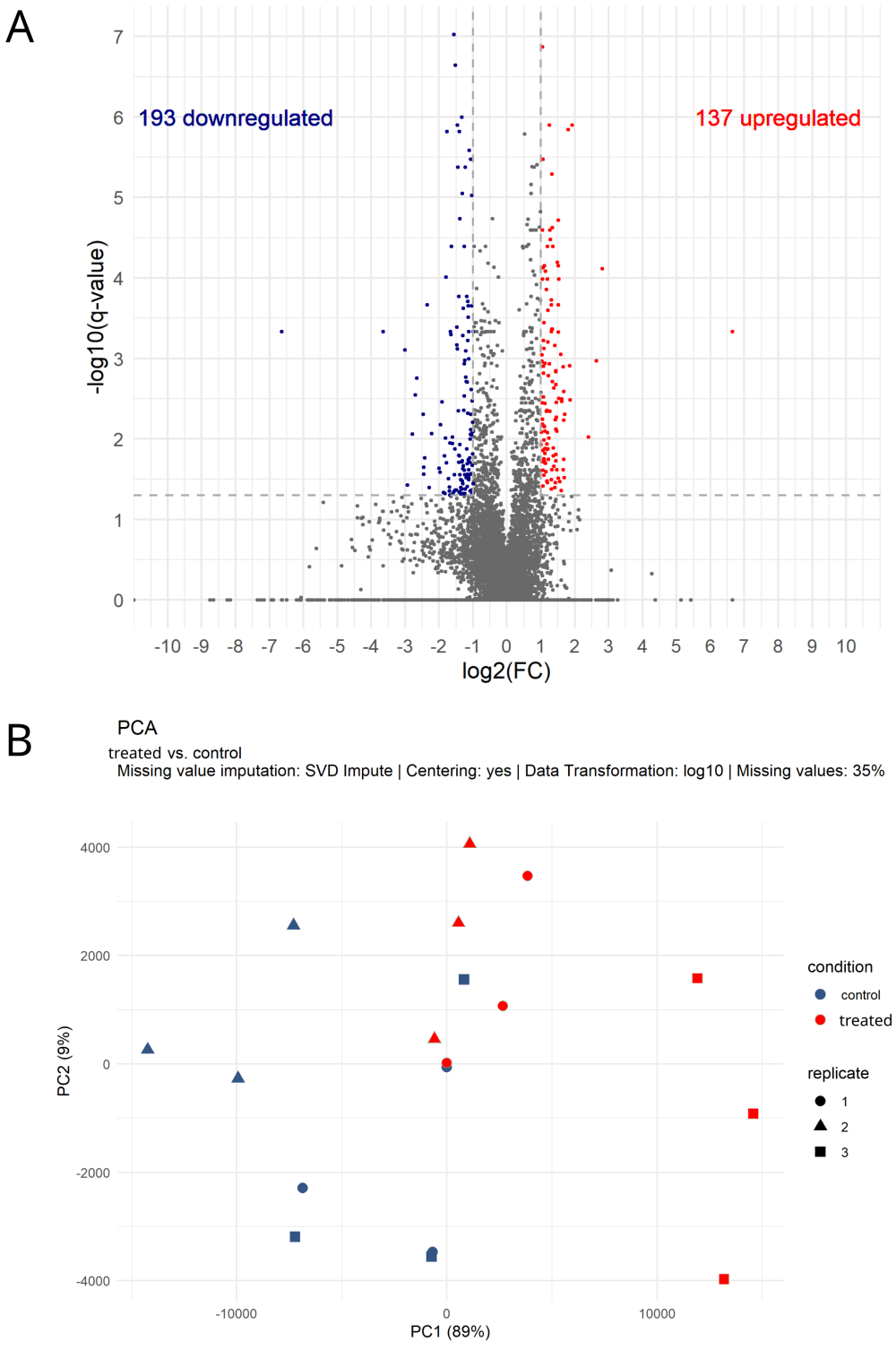

Figure 3.35: A) Results from statistical testing (t-test) as volcano plot from treated versus control murine Th17 cells of the sample set presented in figure 3.34. B) Multivariate analysis by principal components. Control samples and treated samples cluster together, separated over a of PC1, which accounts for 89 % of the observed variances

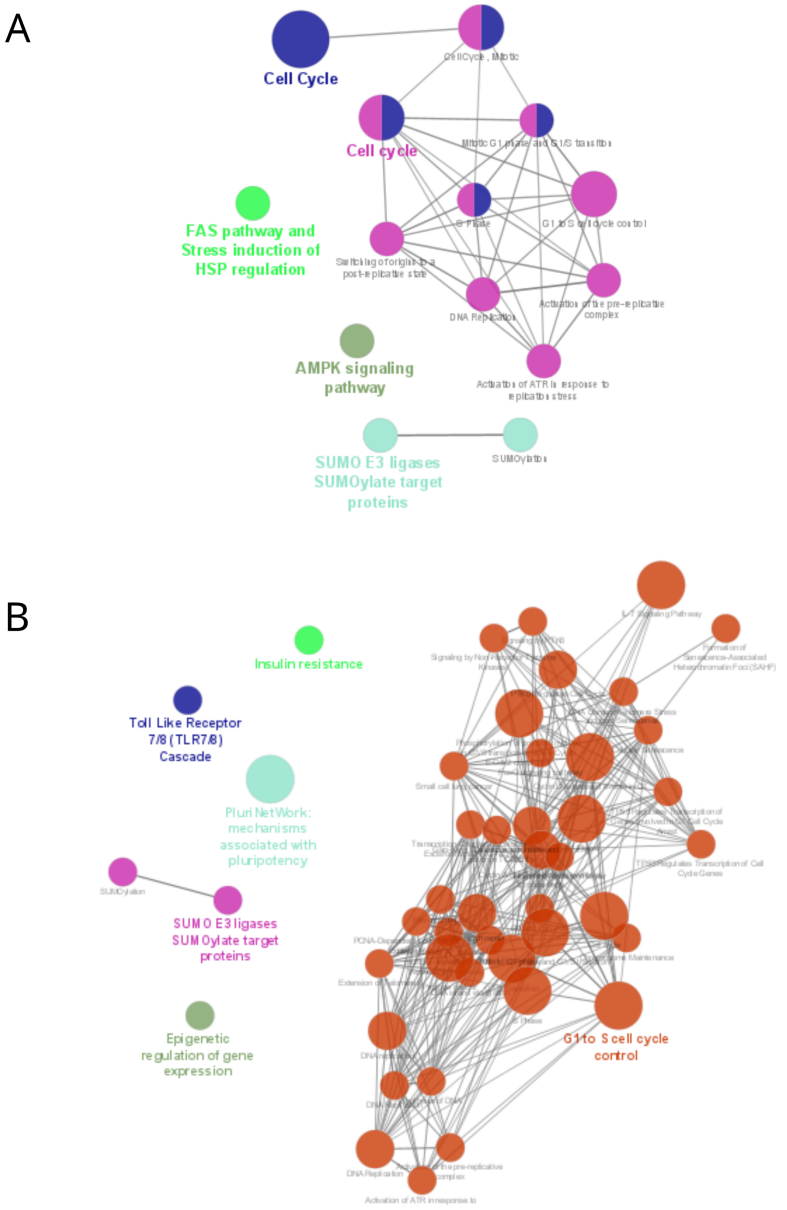

Figure 3.36: ClueGO network from significantly enriched GO terms of phosphopeptides significantly changing in murine Th17 cells after treatment of the sample set presented in figure 3.34 A) from 25 µg. B) from 1000 µg.

enrichment strategies, although 40-fold less protein material was required for the result.

In summary, with this experiment we could not only demonstrate a exceptional phosphopeptide data quality by reproducible number of identifications, excellent correlation in abundance and meaningful identification of differences in principal component analysis, but also validate the accuracy of the presented low amount phosphoproteomics workflow on a biological level.

4 Summary

In this work, novel strategies for phosphoproteomic analysis for sample preparation, data acquisition and statistical analysis were presented. The advantages achieved on three different biological questions were demonstrated:

1. Rapid evolutionary adaptation of osmoregulatory pathways in *Magnaporthe oryzae*
2. Phosphoproteomic response of a novel therapeutic strategy against human osteosarcoma
3. Phosphoproteomic response of treated murine T-Cells

For elucidating rapid evolutionary adaptation in *M.oryzae*, a highly robust phosphoproteomitcs workflow suitable for high-throughput is required. Using conventional lysis and sample preparation strategies, low protein yield due to comparably resistant cell wall and high protease activity leading to an increased number of unspecifically cleaved peptides, is observed. Common strategies, such as the addition of protease inhibitors, could not be applied as this would either decrease tryptic digest, necessary for proteome analysis, or even make it impossible. Therefore, a sample preparation strategy has been developed and successfully published as book chapter [26], that involves heat inactivation of degrading enzymes in combination with harsh cell wall lysis conditions by high concentrations of denaturating agents such as SDS and DTT. In addition to that, phosphoproteomics has been suffering from a high number of missing values due to either a random phosphopeptide enrichment selectivity or borderline signal intensities, which causes the exclusion for fragmentation using the commonly applied data dependent acquisition mode. Consequently, an incomplete dataset reduces confidence in the subsequent statistical testing. We successfully applied data independent acquisition, to include such cases and could prove that while maintaining data quality (such as phosphosite and peptide sequence confidence),

the data completeness increases. To further increase the confidence in the results, we applied a already widely applied statistical analysis from transcriptome analysis, *i.e.* linear models instead of t-test. This approach has proven in literature to be more sensitive on proteome and peptide level, but has no evidence for superior performance demonstrated on phosphopeptide level. Here, we could prove the suitability of the analytical stragety with comparing osmoregulatory response of wild type samples with previously acquired knowledge from literature.

The developed strategy was applied to a time-course experiment of osmosregulatory deficient sample set comprised from MoHog1 deleted mutants (osmosensitive) that show a) no adaptation and are osmosensitive b) reversible adaptation that lose osmoregulatory phenotype and c) irreversibly and stable adapted, that regains osmsoregulatory capabilities. Empirically determined settings for k-means clustering were applied to cluster the temporal profiles of proteome and phosphoproteome response to understand the biological relevance of each functional cluster in more detail. We present this dataset as resource for further investigation of underlying biological processes to understand the adaptation process in more detail.

In addition to that, a phosphopeptide enrichment method has been developed to meet the needs of clinical proteomics and *in-vivo* animal experiments, that in most cases only yield very minute protein amounts which is not sufficient for TiO_2 phosphopeptide enrichment in spin-tips. We could demonstrate the accuracy and reproducibility of the developed workflow using Zr-IMAC magnetic beads in combination with the previously described data independent acqusition by analysis of treated versus control of as low as 25 µg of human osteosarcoma and murine T-Cells. For the human osteosarcoma dataset we also investigated the impact of ion mobility separation for the elucidation of co-eluting and isobaric phosphopeptide isomers, that differ in phosphosite position. The Bruker timsTOF Pro 2 offers an integrated trapped ion mobility separation (tims) before actual mass spectrometry measurement and the the benefit of this procedure on such phosphopeptide isomers has never been demonstrated. We could show that the accuracy of identification increases and the *tims* based ion mobility separation adds confidence and comprehensiveness to phosphoproteomic studies.

The presented strategies for phosphoproteomic analysis for sample preparation, data ac-

quisition and statistical analysis proved as valid and useful alternative to conventional phosphoproteomics studies and open the door for further investigations. Especially current developments in the field aiming at single cell analysis and high throughput pinpoint the demand for further miniaturized methods. For example, one-pot or *in-situ* on tissue sample preparation including phosphopeptide enrichment could resolve spatial differences in proteomics and phosphoproteomics in tissues. This would aid to identify and deconvolute tissue specific responses or drug resistance mechanisms, *e.g.* from cancer FFPE tissues or fresh biopsies. In general, the use of DIA and linear models for data acquisition and statistical analysis will improve the data completeness and support comprehensive knowledge extraction and reduce the impact of intriguing missing value imputation. The added value of this work is double: We pushed phosphoproteomics techniques beyond it's current limits and could provide a comprehensive resource for investigation of phosphoproteomics in *Magnaporthe oryzae*.

5 References

[1] Ralph Dean et al. "The Top 10 fungal pathogens in molecular plant pathology". In: *Molecular plant pathology* 13 (4 May 2012), pp. 414–430. ISSN: 1364-3703. DOI: 10.1111/J.1364-3703.2011.00783.X.

[2] Naomi K. Fukagawa and Lewis H. Ziska. "Rice: Importance for Global Nutrition". In: *Journal of nutritional science and vitaminology* 65 (Supplement 2019), S2–S3. ISSN: 1881-7742. DOI: 10.3177/JNSV.65.S2.

[3] Ralph A. Dean et al. "The genome sequence of the rice blast fungus Magnaporthe grisea". In: *Nature 2005 434:7036* 434 (7036 Apr. 2005), pp. 980–986. ISSN: 1476-4687. DOI: 10.1038/nature03449.

[4] Stefan Bohnert et al. "Rapid adaptation of signaling networks in the fungal pathogen Magnaporthe oryzae". In: *BMC Genomics* 20 (1 Oct. 2019), pp. 1–16. ISSN: 14712164. DOI: 10.1186/S12864-019-6113-3/FIGURES/9.

[5] Hajar Yaakoub et al. "The high osmolarity glycerol (HOG) pathway in fungi-". In: *https://doi.org/10.1080/1040841X.2021.2011834* (2021). ISSN: 15497828. DOI: 10.1080/1040841X.2021.2011834.

[6] Natalie L. Catlett, Olen C. Yoder, and B. Gillian Turgeon. "Whole-Genome Analysis of Two-Component Signal Transduction Genes in Fungal Pathogens". In: *Eukaryotic Cell* 2 (6 Dec. 2003), p. 1151. ISSN: 15359778. DOI: 10.1128/EC.2.6.1151-1161.2003.

[7] Stefan Jacob et al. "Histidine kinases mediate differentiation, stress response, and pathogenicity in Magnaporthe oryzae". In: *MicrobiologyOpen* 3 (5 Oct. 2014), pp. 668–687. ISSN: 2045-8827. DOI: 10.1002/MBO3.197.

[8] Stefan S. Bielack et al. "Second and subsequent recurrences of osteosarcoma: Presentation, treatment, and outcomes of 249 consecutive cooperative osteosarcoma study group patients". In: *Journal of Clinical Oncology* 27 (4 Feb. 2009), pp. 557–565. ISSN: 0732183X. DOI: 10.1200/JCO.2008.16.2305.

[9] Cristina Meazza and Paolo Scanagatta. "Metastatic osteosarcoma: a challenging multidisciplinary treatment". In: *Expert review of anticancer therapy* 16 (5 May 2016), pp. 543–556. ISSN: 1744-8328. DOI: 10.1586/14737140.2016.1168697.

[10] Yu Hsuan Lin et al. "Osteosarcoma: Molecular Pathogenesis and iPSC Modeling". In: *Trends in molecular medicine* 23 (8 Aug. 2017), p. 737. ISSN: 1471499X. DOI: 10.1016/J.MOLMED.2017.06.004.

[11] Susanne Lorenz et al. "Unscrambling the genomic chaos of osteosarcoma reveals extensive transcript fusion, recurrent rearrangements and frequent novel TP53 aberrations". In: *Oncotarget* 7 (5 2016), pp. 5273–5288. ISSN: 1949-2553. DOI: 10.18632/ONCOTARGET.6567.

[12] Jilong Yang et al. "Recurrent LRP1-SNRNP25 and KCNMB4-CCND3 fusion genes promote tumor cell motility in human osteosarcoma". In: *Journal of hematology & oncology* 7 (1 Oct. 2014). ISSN: 1756-8722. DOI: 10.1186/S13045-014-0076-2.

[13] Su Young Kim et al. "The role of IGF-1R in pediatric malignancies". In: *The oncologist* 14 (1 Jan. 2009), pp. 83–91. ISSN: 1549-490X. DOI: 10.1634/THEONCOLOGIST.2008-0189.

[14] Junko Takita. "The role of anaplastic lymphoma kinase in pediatric cancers". In: *Cancer Science* 108 (10 Oct. 2017), p. 1913. ISSN: 13497006. DOI: 10.1111/CAS.13333.

[15] Anton Wellstein. "ALK receptor activation, ligands and therapeutic targeting in glioblastoma and in other cancers". In: *Frontiers in Oncology* 0 (2012), p. 192. ISSN: 2234-943X. DOI: 10.3389/FONC.2012.00192.

[16] Fang Wang et al. "Inhibition of insulin-like growth factor 1 receptor enhances the efficacy of sorafenib in inhibiting hepatocellular carcinoma cell growth and survival". In: *Hepatology Communications* 2 (6 June 2018), pp. 732–746. ISSN: 2471254X. DOI: 10.1002/HEP4.1181/FULL.

[17] Nadine Vewinger et al. "IGF1R Is a Potential New Therapeutic Target for HGNET-BCOR Brain Tumor Patients". In: *International journal of molecular sciences* 20 (12 June 2019). ISSN: 1422-0067. DOI: 10.3390/IJMS20123027.

[18] Emma D. Deeks. "Ceritinib: a Review in ALK-Positive Advanced NSCLC". In: *Targeted oncology* 11 (5 Oct. 2016), pp. 693–700. ISSN: 1776-260X. DOI: 10.1007/S11523-016-0460-7.

[19] Thomas H. Marsilje et al. "Synthesis, structure-activity relationships, and in vivo efficacy of the novel potent and selective anaplastic lymphoma kinase (ALK) inhibitor 5-chloro-N2-(2-isopropoxy-5-methyl-4-(piperidin-4-yl)phenyl)-N4-(2-(isopropylsulfonyl)pl 2,4-diamine (LDK378) currently in phase 1 and phase 2 clinical trials". In: *Journal of medicinal chemistry* 56 (14 July 2013), pp. 5675–5690. ISSN: 1520-4804. DOI: 10.1021/JM400402Q.

[20] Anke E.M. van Erp et al. "Targeting Anaplastic Lymphoma Kinase (ALK) in Rhabdomyosarcoma (RMS) with the Second-Generation ALK Inhibitor Ceritinib". In: *Targeted oncology* 12 (6 Dec. 2017), pp. 815–826. ISSN: 1776-260X. DOI: 10.1007/S11523-017-0528-Z.

[21] Deric L. Wheeler, Mari Iida, and Emily F. Dunn. "The Role of Src in Solid Tumors". In: *The Oncologist* 14 (7 July 2009), pp. 667–678. ISSN: 1083-7159. DOI: 10.1634/THEONCOLOGIST.2009-0009.

[22] Olaf Beck et al. "Safety and Activity of the Combination of Ceritinib and Dasatinib in Osteosarcoma". In: *Cancers* 12 (4 Apr. 2020). ISSN: 20726694. DOI: 10.3390/CANCERS12040793.

[23] Fiorella A. Solari et al. "Why phosphoproteomics is still a challenge". In: *Molecular BioSystems* 11 (6 May 2015), pp. 1487–1493. ISSN: 1742-206X. DOI: 10.1039/C5MB00024F.

[24] Sean J. Humphrey et al. "High-throughput and high-sensitivity phosphoproteomics with the EasyPhos platform". In: *Nature protocols* 13 (9 Sept. 2018), pp. 1897–1916. ISSN: 1750-2799. DOI: 10.1038/S41596-018-0014-9.

[25] Harm Post et al. "Robust, Sensitive, and Automated Phosphopeptide Enrichment Optimized for Low Sample Amounts Applied to Primary Hippocampal Neurons". In: *Journal of proteome research* 16 (2 Feb. 2017), pp. 728–737. ISSN: 1535-3907. DOI: 10.1021/ACS.JPROTEOME.6B00753.

[26] Thomas Michna and Stefan Tenzer. "Quantitative Proteome and Phosphoproteome Profiling in Magnaporthe oryzae". In: *Methods in molecular biology (Clifton, N.J.)* 2356 (2021), pp. 109–119. ISSN: 1940-6029. DOI: 10.1007/978-1-0716-1613-0_9.

[27] Ignacio Arribas Diez et al. "Zirconium(IV)-IMAC Revisited: Improved Performance and Phosphoproteome Coverage by Magnetic Microparticles for Phosphopeptide Affinity Enrichment". In: *Journal of proteome research* 20 (1 Jan. 2021), pp. 453–462. ISSN: 1535-3907. DOI: 10.1021/ACS.JPROTEOME.0C00508.

[28] Matthew S. Glover et al. "Examining the Influence of Phosphorylation on Peptide Ion Structure by Ion Mobility Spectrometry-Mass Spectrometry". In: *Journal of the American Society for Mass Spectrometry* 27 (5 May 2016), pp. 786–794. ISSN: 1879-1123. DOI: 10.1007/S13361-016-1343-Y.

[29] Lars Fugger, Lise Torp Jensen, and Jamie Rossjohn. "Challenges, Progress, and Prospects of Developing Therapies to Treat Autoimmune Diseases". In: *Cell* 181 (1 Apr. 2020), pp. 63–80. ISSN: 0092-8674. DOI: 10.1016/J.CELL.2020.03.007.

[30] A. F. Maarten Altelaar, Javier Munoz, and Albert J. R. Heck. "Next-generation proteomics: towards an integrative view of proteome dynamics". In: *Nature Reviews Genetics* 14 (1 Jan. 2013), pp. 35–48. ISSN: 1471-0056. DOI: 10.1038/nrg3356.

[31] Philip Cohen. "The regulation of protein function by multisite phosphorylation–a 25 year update". In: *Trends in biochemical sciences* 25 (12 Dec. 2000), pp. 596–601. ISSN: 0968-0004. DOI: 10.1016/S0968-0004(00)01712-6.

[32] James I. Garrels. "Quantitative two-dimensional gel electrophoresis of proteins". In: *Methods in Enzymology* 100 (Jan. 1983), pp. 411–423. ISSN: 0076-6879. DOI: 10.1016/0076-6879(83)00070-1.

[33] Cosette Abdallah et al. "Gel-based and gel-free quantitative proteomics approaches at a glance." In: *International journal of plant genomics* 2012 (Nov. 2012), p. 494572. ISSN: 1687-5389. DOI: 10.1155/2012/494572.

[34] Ute Distler et al. "Label-free quantification in ion mobility-enhanced data-independent acquisition proteomics". In: *Nature Protocols* 11 (4 Apr. 2016), pp. 795–812. ISSN: 1754-2189. DOI: 10.1038/nprot.2016.042.

[35] Liisa Arike and Lauri Peil. *Spectral Counting Label-Free Proteomics*. 2014. DOI: 10.1007/978-1-4939-0685-7_14.

[36] Fatima Ardito et al. "The crucial role of protein phosphorylation in cell signaling and its use as targeted therapy (Review)". In: *International Journal of Molecular Medicine* 40 (2 Aug. 2017), pp. 271–280. ISSN: 1791244X. DOI: 10.3892/IJMM.2017.3036/HTML.

[37] Cheryl L. Mathis and Amy M. Barrios. "Histidine phosphorylation in metalloprotein binding sites". In: *Journal of inorganic biochemistry* 225 (Dec. 2021). ISSN: 1873-3344. DOI: 10.1016/J.JINORGBIO.2021.111606.

[38] Ao Zhang, Fr-d-rique Pompeo, and Anne Galinier. "Overview of protein phosphorylation in bacteria with a main focus on unusual protein kinases in Bacillus subtilis". In: *Research in microbiology* 172 (7-8 Nov. 2021). ISSN: 1769-7123. DOI: 10.1016/J.RESMIC.2021.103871.

[39] Matthias Mann et al. "Analysis of protein phosphorylation using mass spectrometry: deciphering the phosphoproteome". In: *Trends in Biotechnology* 20 (6 June 2002), pp. 261–268. ISSN: 0167-7799. DOI: 10.1016/S0167-7799(02)01944-3.

[40] Mostafa Zarei et al. "Comparison of ERLIC-TiO2, HILIC-TiO2, and SCX-TiO2 for global phosphoproteomics approaches." In: *Journal of proteome research* 10 (8 Aug. 2011), pp. 3474–83. ISSN: 1535-3907. DOI: 10.1021/pr200092z.

[41] Maria Stella Ritorto et al. "Hydrophilic Strong Anion Exchange (hSAX) Chromatography for Highly Orthogonal Peptide Separation of Complex Proteomes". In: *Journal of Proteome Research* 12 (6 June 2013), pp. 2449–2457. ISSN: 1535-3893. DOI: 10.1021/pr301011r.

[42] Wan Mohd Aizat and Maizom Hassan. "Proteomics in systems biology". In: *Advances in Experimental Medicine and Biology* 1102 (2018), pp. 31–49. ISSN: 22148019. DOI: 10.1007/978-3-319-98758-3_3/COVER/.

[43] Daniel P. Donnelly et al. "Best practices and benchmarks for intact protein analysis for top-down mass spectrometry". In: *Nature Methods 2019 16:7* 16 (7 June 2019), pp. 587–594. ISSN: 1548-7105. DOI: 10.1038/s41592-019-0457-0.

[44] Neil L. Kelleher. "Peer Reviewed: Top-Down Proteomics". In: *Analytical Chemistry* 76 (11 June 2004), 196 A–203 A. ISSN: 0003-2700. DOI: 10.1021/AC0415657.

[45] Philip R. Gafken. "An overview of the qualitative analysis of phosphoproteins by mass spectrometry." In: *Methods in molecular biology (Clifton, N.J.)* 527 (2009), pp. 159–172. ISSN: 10643745. DOI: 10.1007/978-1-60327-834-8_12/FIGURES/4_12.

[46] Martin C. Stumpe and Helmut Grubm-ller. "Interaction of urea with amino acids: Implications for urea-induced protein denaturation". In: *Journal of the American Chemical Society* 129 (51 Dec. 2007), pp. 16126–16131. ISSN: 00027863. DOI: 10.1021/JA076216J/SUPPL_FILE/JA076216JSI20071004_081248.PDF.

[47] A. Wallqvist, D. G. Covell, and D. Thirumalai. "Hydrophobic interactions in aqueous urea solutions with implications for the mechanism of protein denaturation". In: *Journal of the American Chemical Society* 120 (2 Jan. 1998), pp. 427–428. ISSN: 00027863. DOI: 10.1021/JA972053V/ASSET/IMAGES/LARGE/JA972053VF00003.JPEG.

[48] Giovanni Salvi, Paolo De Los Rios, and Michele Vendruscolo. "Effective interactions between chaotropic agents and proteins". In: *Proteins: Structure, Function, and Bioinformatics* 61 (3 Nov. 2005), pp. 492–499. ISSN: 1097-0134. DOI: 10.1002/PROT.20626.

[49] Mark P. Kamps and Bartholomew M. Sefton. "Acid and base hydrolysis of phosphoproteins bound to Immobilon facilitates analysis of phosphoamino acids in gel-fractionated proteins". In: *Analytical Biochemistry* 176.1 (1989), pp. 22–27. ISSN: 0003-2697.

[50] Benjamin Schlager, Anna Straessle, and Ernst Hafen. "Use of anionic denaturing detergents to purify insoluble proteins after overexpression". In: *BMC Biotechnology* 12 (1 Dec. 2012), pp. 1–7. ISSN: 14726750. DOI: 10.1186/1472-6750-12-95/TABLES/2.

[51] Giedre Ratkeviciute, Benjamin F. Cooper, and Timothy J. Knowles. "Methods for the solubilisation of membrane proteins: the micelle-aneous world of membrane protein solubilisation". In: *Biochemical Society Transactions* 49 (4 Aug. 2021), p. 1763. ISSN: 14708752. DOI: 10.1042/BST20210181.

[52] M. J. Ruiz-Angel et al. “Performance of different C18 columns in reversed-phase liquid chromatography with hydro-organic and micellar-organic mobile phases”. In: *Journal of chromatography. A* 1344 (May 2014), pp. 76–82. ISSN: 1873-3778. DOI: 10.1016/J.CHROMA.2014.04.011.

[53] Joselito P. Quirino. “Sodium dodecyl sulfate removal during electrospray ionization using cyclodextrins as simple sample solution additive for improved mass spectrometric detection of peptides”. In: *Analytica chimica acta* 1005 (Apr. 2018), pp. 54–60. ISSN: 1873-4324. DOI: 10.1016/J.ACA.2017.12.012.

[54] Malte Sielaff et al. “Evaluation of FASP, SP3, and iST Protocols for Proteomic Sample Preparation in the Low Microgram Range”. In: *Journal of proteome research* 16 (11 Nov. 2017), pp. 4060–4072. ISSN: 1535-3907. DOI: 10.1021/ACS.JPROTEOME.7B00433.

[55] Jure Zevnik and Matev- Dular. “Cavitation bubble interaction with compliant structures on a microscale: A contribution to the understanding of bacterial cell lysis by cavitation treatment”. In: *Ultrasonics sonochemistry* 87 (June 2022), p. 106053. ISSN: 1873-2828. DOI: 10.1016/J.ULTSONCH.2022.106053.

[56] Torsten M-ller and Dominic Winter. “Systematic Evaluation of Protein Reduction and Alkylation Reveals Massive Unspecific Side Effects by Iodine-containing Reagents”. In: *Molecular & Cellular Proteomics : MCP* 16 (7 July 2017), p. 1173. ISSN: 15359484. DOI: 10.1074/MCP.M116.064048.

[57] Yvonne Markert et al. “Proline versus charge concept for protein stabilization against proteolytic attack”. In: *Protein Engineering, Design and Selection* 16 (12 Dec. 2003), pp. 1041–1046. ISSN: 1741-0126. DOI: 10.1093/PROTEIN/GZG136.

[58] Mark J. Wall et al. “Implications of partial tryptic digestion in organic-aqueous solvent systems for bottom-up proteome analysis”. In: *Analytica chimica acta* 703 (2 Oct. 2011), pp. 194–203. ISSN: 1873-4324. DOI: 10.1016/J.ACA.2011.07.025.

[59] Marek -ebela et al. “Thermostable trypsin conjugates for high-throughput proteomics: synthesis and performance evaluation”. In: *Proteomics* 6 (10 May 2006), pp. 2959–2963. ISSN: 1615-9853. DOI: 10.1002/PMIC.200500576.

[60] Harsha P. Gunawardena, Joshua F. Emory, and Scott A. McLuckey. "Phosphopeptide Anion Characterization via Sequential Charge Inversion and Electron Transfer Dissociation". In: *Analytical chemistry* 78 (11 June 2006), p. 3788. ISSN: 00032700. DOI: 10.1021/AC060164J.

[61] Martin R. Larsen et al. "Highly Selective Enrichment of Phosphorylated Peptides from Peptide Mixtures Using Titanium Dioxide Microcolumns". In: *Molecular & Cellular Proteomics* 4 (7 July 2005), pp. 873–886. ISSN: 1535-9476. DOI: 10.1074/MCP.T500007-MCP200.

[62] Uma K. Aryal and Andrew R.S. Ross. "Enrichment and analysis of phosphopeptides under different experimental conditions using titanium dioxide affinity chromatography and mass spectrometry". In: *Rapid communications in mass spectrometry : RCM* 24 (2 Jan. 2010), pp. 219–231. ISSN: 1097-0231. DOI: 10.1002/RCM.4377.

[63] Tine E Thingholm et al. "SIMAC (Sequential Elution from IMAC), a Phosphoproteomics Strategy for the Rapid Separation of Monophosphorylated from Multiply Phosphorylated Peptides* - S". In: *Molecular and Cellular Proteomics* 7 (2008), pp. 661–671. DOI: 10.1074/mcp.M700362-MCP200.

[64] Yeonyee Oh, William L. Franck, and Ralph A. Dean. *Sequential Phosphopeptide Enrichment for Phosphoproteome Analysis of Filamentous Fungi: A Test Case Using Magnaporthe oryzae*. 2018. DOI: 10.1007/978-1-4939-8724-5_7.

[65] P. A. Connor and A. J. McQuillan. "Phosphate adsorption onto TiO2 from aqueous solutions: an in situ internal reflection infrared spectroscopic study". In: *Langmuir* 15 (8 Apr. 1999), pp. 2916–2921. ISSN: 07437463. DOI: 10.1021/LA980894P/ASSET/IMAGES/LARGE/LA980894PF00008.JPEG.

[66] Lucrece Matheron et al. "Characterization of biases in phosphopeptide enrichment by Ti 4+-immobilized metal affinity chromatography and TiO2 using a massive synthetic library and human cell digests". In: *Analytical Chemistry* 86 (16 Aug. 2014), pp. 8312–8320. ISSN: 15206882. DOI: 10.1021/AC501803Z/SUPPL_FILE/AC501803Z_SI_003.XLSX.

[67] Arthur R. Salomon et al. "Profiling of tyrosine phosphorylation pathways in human cells using mass spectrometry". In: *Proceedings of the National Academy of Sciences of the United States of America* 100 (2 Jan. 2003), pp. 443–448. ISSN: 00278424. DOI: 10.1073/PNAS.2436191100/ASSET/B241601E-13E6-46CC-9A60-4EADB69EEF6D/ASSETS/GRAPHIC/PQ2426191003.JPEG.

[68] Clement M. Potel et al. "Widespread bacterial protein histidine phosphorylation revealed by mass spectrometry-based proteomics". In: *Nature methods* 15 (3 Mar. 2018), pp. 187–190. ISSN: 1548-7105. DOI: 10.1038/NMETH.4580.

[69] Haixing Wang et al. "Development and evaluation of a micro- and nanoscale proteomic sample preparation method". In: *Journal of Proteome Research* 4 (6 Nov. 2005), pp. 2397–2403. ISSN: 15353893. DOI: 10.1021/PR050160F/ASSET/IMAGES/LARGE/PR050160FF00002.JPEG.

[70] Wiebke Maria Nadler et al. "MALDI versus ESI: The Impact of the Ion Source on Peptide Identification". In: *Journal of Proteome Research* 16 (3 Mar. 2017), pp. 1207–1215. ISSN: 15353907. DOI: 10.1021/ACS.JPROTEOME.6B00805/ASSET/IMAGES/MEDIUM/PR-2016-008057_0006.GIF.

[71] D. J. Douglas. "Linear quadrupoles in mass spectrometry". In: *Mass spectrometry reviews* 28 (6 Nov. 2009), pp. 937–960. ISSN: 1098-2787. DOI: 10.1002/MAS.20249.

[72] Enrique Calvo et al. "Applying selected reaction monitoring to targeted proteomics". In: *Expert review of proteomics* 8 (2 Apr. 2011), pp. 165–173. ISSN: 1744-8387. DOI: 10.1586/EPR.11.11.

[73] a. D. Mc Naught and a Wilkinson. "Compendium of Chemical Terminology-Gold Book". In: *Iupac* (2012), p. 1670. ISSN: 0033-4545. DOI: 10.1351/goldbook.

[74] Igor V. Chernushevich, Alexander V. Loboda, and Bruce A. Thomson. "An introduction to quadrupole-time-of-flight mass spectrometry". In: *Journal of mass spectrometry : JMS* 36 (8 2001), pp. 849–865. ISSN: 1076-5174. DOI: 10.1002/JMS.207.

[75] Roman A. Zubarev and Alexander Makarov. "Orbitrap mass spectrometry". In: *Analytical Chemistry* 85 (11 June 2013), pp. 5288–5296. ISSN: 00032700. DOI: 10.1021/AC4001223/ASSET/IMAGES/LARGE/AC-2013-001223_0006.JPEG.

[76] Adrian Guthals and Nuno Bandeira. "Peptide Identification by Tandem Mass Spectrometry with Alternate Fragmentation Modes". In: *Molecular & Cellular Proteomics : MCP* 11 (9 Sept. 2012), p. 550. ISSN: 15359476. DOI: 10.1074/MCP.R112.018556.

[77] Alexander Makarov et al. "Dynamic range of mass accuracy in LTQ Orbitrap hybrid mass spectrometer". In: *Journal of the American Society for Mass Spectrometry* 17 (7 2006), pp. 977–982. ISSN: 1044-0305. DOI: 10.1016/J.JASMS.2006.03.006.

[78] Helmut E. Meyer et al. "Massenspektrometrie". In: *Bioanalytik* (2022), pp. 359–414. DOI: 10.1007/978-3-662-61707-6_16.

[79] Karsten Michelmann et al. "Fundamentals of trapped ion mobility spectrometry". In: *Journal of the American Society for Mass Spectrometry* 26 (1 2014), pp. 14–24. ISSN: 18791123. DOI: 10.1007/S13361-014-0999-4/SUPPL_FILE/JS8B04886_SI_001.DOCX.

[80] James N. Dodds and Erin S. Baker. "Ion Mobility Spectrometry: Fundamental Concepts, Instrumentation, Applications, and the Road Ahead". In: *Journal of the American Society for Mass Spectrometry* 30 (11 Nov. 2019), p. 2185. ISSN: 18791123. DOI: 10.1007/S13361-019-02288-2.

[81] Catherine G. Vasilopoulou et al. "Trapped ion mobility spectrometry and PASEF enable in-depth lipidomics from minimal sample amounts". In: *Nature Communications 2020 11:1* 11 (1 Jan. 2020), pp. 1–11. ISSN: 2041-1723. DOI: 10.1038/s41467-019-14044-x.

[82] Joshua A. Silveira et al. "Parallel accumulation for 100% duty cycle trapped ion mobility-mass spectrometry". In: *International Journal of Mass Spectrometry* 413 (Feb. 2017), pp. 168–175. ISSN: 1387-3806. DOI: 10.1016/J.IJMS.2016.03.004.

[83] Lukas Krasny and Paul H. Huang. "Data-independent acquisition mass spectrometry (DIA-MS) for proteomic applications in oncology". In: *Molecular Omics* 17 (1 Feb. 2021), pp. 29–42. ISSN: 2515-4184. DOI: 10.1039/D0MO00072H.

[84] Vadim Demichev et al. "DIA-NN: neural networks and interference correction enable deep proteome coverage in high throughput". In: *Nature methods* 17 (1 Jan. 2020), pp. 41–44. ISSN: 1548-7105. DOI: 10.1038/S41592-019-0638-X.

[85] Alex Bateman et al. "UniProt: the universal protein knowledgebase in 2021". In: *Nucleic Acids Research* 49 (D1 Jan. 2021), pp. D480–D489. ISSN: 0305-1048. DOI: 10.1093/NAR/GKAA1100.

[86] Jürgen Cox and Matthias Mann. "MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification". In: *Nature biotechnology* 26 (12 Dec. 2008), pp. 1367–1372. ISSN: 1546-1696. DOI: 10.1038/NBT.1511.

[87] J-rgen Cox et al. "Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ". In: *Molecular & cellular proteomics : MCP* 13 (9 Sept. 2014), pp. 2513–2526. ISSN: 1535-9484. DOI: 10.1074/MCP.M113.031591.

[88] Jeremy D. O'Connell et al. "Proteome-Wide Evaluation of Two Common Protein Quantification Methods". In: *Journal of proteome research* 17 (5 May 2018), p. 1934. ISSN: 15353907. DOI: 10.1021/ACS.JPROTEOME.8B00016.

[89] Andy T. Kong et al. "MSFragger: ultrafast and comprehensive peptide identification in shotgun proteomics". In: *Nature methods* 14 (5 Apr. 2017), p. 513. ISSN: 15487105. DOI: 10.1038/NMETH.4256.

[90] Vadim Demichev et al. "High sensitivity dia-PASEF proteomics with DIA-NN and FragPipe". In: *bioRxiv* (Mar. 2021), p. 2021.03.08.434385. DOI: 10.1101/2021.03.08.434385.

[91] Jing Zhang et al. "PEAKS DB: De Novo Sequencing Assisted Database Search for Sensitive and Accurate Peptide Identification". In: *Molecular & Cellular Proteomics : MCP* 11 (4 Apr. 2012). ISSN: 15359476. DOI: 10.1074/MCP.M111.010587.

[92] Ana Marcu et al. "Original research: HLA Ligand Atlas: a benign reference of HLA-presented peptides to improve T-cell-based cancer immunotherapy". In: *Journal for Immunotherapy of Cancer* 9 (4 Apr. 2021), p. 2071. ISSN: 20511426. DOI: 10.1136/JITC-2020-002071.

[93] Aivett Bilbao et al. "Processing strategies and software solutions for data-independent acquisition in mass spectrometry". In: *Proteomics* 15 (5-6 Mar. 2015), pp. 964–980. ISSN: 1615-9861. DOI: 10.1002/PMIC.201400323.

[94] George Rosenberger et al. "Statistical control of peptide and protein error rates in large-scale targeted data-independent acquisition analyses". In: *Nature Methods 2017 14:9* 14 (9 Aug. 2017), pp. 921–927. ISSN: 1548-7105. DOI: 10.1038/nmeth.4398.

[95] Ying S. Ting et al. "Peptide-Centric Proteome Analysis: An Alternative Strategy for the Analysis of Tandem Mass Spectrometry Data". In: *Molecular & Cellular Proteomics : MCP* 14 (9 Sept. 2015), p. 2301. ISSN: 15359484. DOI: 10.1074/MCP.O114.047035.

[96] Siegfried Gessulat et al. "Prosit: proteome-wide prediction of peptide tandem mass spectra by deep learning". In: *Nature methods* 16 (6 June 2019), pp. 509–518. ISSN: 1548-7105. DOI: 10.1038/S41592-019-0426-7.

[97] Dorte B. Bekker-Jensen et al. "Rapid and site-specific deep phosphoproteome profiling by data-independent acquisition without the need for spectral libraries". In: *Nature Communications 2020 11:1* 11 (1 Feb. 2020), pp. 1–12. ISSN: 2041-1723. DOI: 10.1038/s41467-020-14609-1.

[98] Ana Martinez-Val et al. "Data Processing and Analysis for DIA-Based Phosphoproteomics Using Spectronaut". In: *Methods in molecular biology (Clifton, N.J.)* 2361 (2021), pp. 95–107. ISSN: 1940-6029. DOI: 10.1007/978-1-0716-1641-3_6.

[99] Brendan MacLean et al. "Skyline: an open source document editor for creating and analyzing targeted proteomics experiments". In: *Bioinformatics (Oxford, England)* 26 (7 Feb. 2010), pp. 966–968. ISSN: 1367-4811. DOI: 10.1093/BIOINFORMATICS/BTQ054.

[100] Tommi V-likangas, Tomi Suomi, and Laura L. Elo. "A systematic evaluation of normalization methods in quantitative label-free proteomics". In: *Briefings in Bioinformatics* 19 (1 Jan. 2018), p. 1. ISSN: 14774054. DOI: 10.1093/BIB/BBW095.

[101] Sushma Anand et al. "Label-based and label-free strategies for protein quantitation". In: *Methods in Molecular Biology* 1549 (2017), pp. 31–43. ISSN: 10643745. DOI: 10.1007/978-1-4939-6740-7_4/FIGURES/1.

[102] Jiaming Li et al. "TMTpro-18plex: The Expanded and Complete Set of TMTpro Reagents for Sample Multiplexing". In: *Journal of Proteome Research* 20 (5 May

2021), pp. 2964–2972. ISSN: 15353907. DOI: 10.1021/ACS.JPROTEOME.1C00168/SUPPL_FILE/PR1C00168_SI_005.XLSX.

[103] Bj-rn Schwanhausser et al. "Global quantification of mammalian gene expression control". In: *Nature* 473 (7347 May 2011), pp. 337–342. ISSN: 1476-4687. DOI: 10.1038/NATURE10098.

[104] Wolfgang Huber et al. "Variance stabilization applied to microarray data calibration and to the quantification of differential expression". In: *Bioinformatics (Oxford, England)* 18 Suppl 1 (SUPPL. 1 2002). ISSN: 1367-4803. DOI: 10.1093/BIOINFORMATICS/18.SUPPL_1.S96.

[105] Stephen J. Callister et al. "Normalization approaches for removing systematic biases associated with mass spectrometry and label-free proteomics". In: *Journal of proteome research* 5 (2 Feb. 2006), pp. 277–286. ISSN: 1535-3893. DOI: 10.1021/PR050300L.

[106] Olga Troyanskaya et al. "Missing value estimation methods for DNA microarrays". In: *Bioinformatics (Oxford, England)* 17 (6 2001), pp. 520–525. ISSN: 1367-4803. DOI: 10.1093/BIOINFORMATICS/17.6.520.

[107] P. Griffiths and Jack Needleman. "Statistical significance testing and p-values: Defending the indefensible? A discussion paper and position statement". In: *International journal of nursing studies* 99 (Nov. 2019). ISSN: 1873-491X. DOI: 10.1016/J.IJNURSTU.2019.07.001.

[108] Dieter Rasch, Klaus D. Kubinger, and Karl Moder. "The two-sample t test: pre-testing its assumptions does not pay off". In: *Statistical Papers 2009 52:1* 52 (1 Apr. 2009), pp. 219–231. ISSN: 1613-9798. DOI: 10.1007/S00362-009-0224-X.

[109] Prabhaker Mishra et al. "Application of student's t-test, analysis of variance, and covariance". In: *Annals of cardiac anaesthesia* 22 (4 2019), p. 407. ISSN: 0974-5181. DOI: 10.4103/ACA.ACA_94_19.

[110] Manfei Xu et al. "The Differences and Similarities Between Two-Sample T-Test and Paired T-Test". In: *Shanghai archives of psychiatry* 29 (3 June 2017), pp. 184–188. ISSN: 1002-0829. DOI: 10.11919/J.ISSN.1002-0829.217070.

[111] Wolfgang Huber et al. "Orchestrating high-throughput genomic analysis with Bioconductor". In: *Nature Methods 2015 12:2* 12 (2 Jan. 2015), pp. 115–121. ISSN: 1548-7105. DOI: 10.1038/nmeth.3252.

[112] Michiel P. van Ooijen et al. "Identification of differentially expressed peptides in high-throughput proteomics data". In: *Briefings in Bioinformatics* 19 (5 Sept. 2018), pp. 971–981. ISSN: 1467-5463. DOI: 10.1093/BIB/BBX031.

[113] Shi Yi Chen, Zhe Feng, and Xiaolian Yi. "A general introduction to adjustment for multiple comparisons". In: *Journal of Thoracic Disease* 9 (6 June 2017), p. 1725. ISSN: 20776624. DOI: 10.21037/JTD.2017.05.34.

[114] Yoav Benjamini and Yosef Hochberg. "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing". In: *Journal of the Royal Statistical Society: Series B (Methodological)* 57 (1 Jan. 1995), pp. 289–300. ISSN: 2517-6161. DOI: 10.1111/J.2517-6161.1995.TB02031.X.

[115] Detlef Groth et al. "Principal components analysis". In: *Methods in molecular biology (Clifton, N.J.)* 930 (2013), pp. 527–547. ISSN: 1940-6029. DOI: 10.1007/978-1-62703-059-5_22.

[116] Leland McInnes et al. "UMAP: Uniform Manifold Approximation and Projection". In: *Journal of Open Source Software* 3 (29 Sept. 2018), p. 861. ISSN: 2475-9066. DOI: 10.21105/JOSS.00861.

[117] Laurens Van Der Maaten and Geoffrey Hinton. "Visualizing Data using t-SNE". In: *Journal of Machine Learning Research* 9 (2008), pp. 2579–2605.

[118] Bill Andreopoulos et al. "A roadmap of clustering algorithms: finding a match for a biomedical application". In: *Briefings in bioinformatics* 10 (3 2009), pp. 297–314. ISSN: 1477-4054. DOI: 10.1093/BIB/BBN058.

[119] Damian Szklarczyk et al. "The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets". In: *Nucleic acids research* 49 (D1 Jan. 2021), pp. D605–D612. ISSN: 1362-4962. DOI: 10.1093/NAR/GKAA1074.

[120] Huaiyu Mi, Anushya Muruganujan, and Paul D. Thomas. "PANTHER in 2013: modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees". In: *Nucleic acids research* 41 (Database issue Jan. 2013). ISSN: 1362-4962. DOI: 10.1093/NAR/GKS1118.

[121] Gabriela Bindea et al. "ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks". In: *Bioinformatics (Oxford, England)* 25 (8 2009), pp. 1091–1093. ISSN: 1367-4811. DOI: 10.1093/BIOINFORMATICS/BTP101.

[122] Peter V. Hornbeck et al. "PhosphoSitePlus, 2014: mutations, PTMs and recalibrations". In: *Nucleic acids research* 43 (Database issue Jan. 2015), pp. D512–D520. ISSN: 1362-4962. DOI: 10.1093/NAR/GKU1267.

[123] Danica D. Wiredja, Mehmet Koyut-rk, and Mark R. Chance. "The KSEA App: a web-based tool for kinase activity inference from quantitative phosphoproteomics". In: *Bioinformatics* 33 (21 Nov. 2017), pp. 3489–3491. ISSN: 1367-4803. DOI: 10.1093/BIOINFORMATICS/BTX415.

[124] Robin Beekhof et al. "INKA, an integrative data analysis pipeline for phosphoproteomic inference of active kinases". In: *Molecular Systems Biology* 15 (4 Apr. 2019), e8250. ISSN: 1744-4292. DOI: 10.15252/MSB.20188250.

[125] -zg-n Babur et al. "Causal interactions from proteomic profiles: Molecular data meet pathway knowledge". In: *Patterns* 2 (6 June 2021), p. 100257. ISSN: 26663899. DOI: 10.1016/J.PATTER.2021.100257/ATTACHMENT/BA851778-043C-4584-A133-AA7E8DB52474/MMC4.XLSX.

[126] Andrew J. Alpert. "Electrostatic repulsion hydrophilic interaction chromatography for isocratic separation of charged solutes and selective isolation of phosphopeptides". In: *Analytical Chemistry* 80 (1 Jan. 2008), pp. 62–76. ISSN: 00032700. DOI: 10.1021/AC070997P/ASSET/IMAGES/LARGE/AC070997PF00021.JPEG.

[127] Shujiro Okuda et al. "jPOSTrepo: an international standard data repository for proteomes". In: *Nucleic acids research* 45 (D1 Jan. 2017), pp. D1107–D1111. ISSN: 1362-4962. DOI: 10.1093/NAR/GKW1080.

[128] Juan A. Vizca-no et al. "ProteomeXchange provides globally coordinated proteomics data submission and dissemination". In: *Nature Biotechnology 2014 32:3* 32 (3 Mar. 2014), pp. 223–226. ISSN: 1546-1696. DOI: 10.1038/nbt.2839.

[129] Carol A. Munro. "Chitin and Glucan, the Yin and Yang of the Fungal Cell Wall, Implications for Antifungal Drug Discovery and Therapy". In: *Advances in Applied Microbiology* 83 (Jan. 2013), pp. 145–172. ISSN: 0065-2164. DOI: 10.1016/B978-0-12-407678-5.00004-0.

[130] Jennifer L. Proc et al. "A quantitative study of the effects of chaotropic agents, surfactants, and solvents on the digestion efficiency of human plasma proteins by trypsin". In: *Journal of proteome research* 9 (10 Oct. 2010), pp. 5422–5437. ISSN: 1535-3907. DOI: 10.1021/PR100656U.

[131] Paula Monteiro de Souza et al. "A biotechnology perspective of fungal proteases". In: *Brazilian Journal of Microbiology* 46 (2 June 2015), p. 337. ISSN: 16784405. DOI: 10.1590/S1517-838246220140359.

[132] Juri Rappsilber, Matthias Mann, and Yasushi Ishihama. "Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips". In: *Nature protocols* 2 (8 Aug. 2007), pp. 1896–1906. ISSN: 1750-2799. DOI: 10.1038/NPROT.2007.261.

[133] Dennis A Dougherty. "The Cation– Interaction". In: *Acc Chem Res* 46 (4 2013), pp. 885–893. DOI: 10.1021/ar300265y.

[134] Eric D. Glendening and Frank Weinhold. "Pauling-s Conceptions of Hybridization and Resonance in Modern Quantum Chemistry". In: *Molecules 2021, Vol. 26, Page 4110* 26 (14 July 2021), p. 4110. ISSN: 1420-3049. DOI: 10.3390/MOLECULES26144110.

[135] Ewelina Eckert, Fatemeh Bamdad, and Lingyun Chen. "Metal solubility enhancing peptides derived from barley protein". In: *Food Chemistry* 159 (Sept. 2014), pp. 498–506. ISSN: 0308-8146. DOI: 10.1016/J.FOODCHEM.2014.03.061.

[136] Jordy J. Hsiao et al. "Improved LC/MS Methods for the Analysis of Metal-Sensitive Analytes Using Medronic Acid as a Mobile Phase Additive". In: *Analytical chemistry* 90 (15 Aug. 2018), pp. 9457–9464. ISSN: 1520-6882. DOI: 10.1021/ACS.ANALCHEM.8B02100.

[137] Robert E. Birdsall et al. "Application of mobile phase additives to reduce metal-ion mediated adsorption of non-phosphorylated peptides in RPLC/MS-based assays". In: *Journal of chromatography. B, Analytical technologies in the biomedical and life sciences* 1126-1127 (Sept. 2019). ISSN: 1873-376X. DOI: 10.1016/J.JCHROMB.2019.121773.

[138] Martin R. Larsen et al. "Improved detection of hydrophilic phosphopeptides using graphite powder microcolumns and mass spectrometry: evidence for in vivo doubly phosphorylated dynamin I and dynamin III". In: *Molecular & cellular proteomics : MCP* 3 (5 May 2004), pp. 456–465. ISSN: 1535-9476. DOI: 10.1074/MCP.M300105-MCP200.

[139] Stoyan Stoychev. *Magnetic HILIC: An enabling & versatile tool for robust automated MS sample preparation work*. 2018.

[140] Andrew J. Alpert, Otto Hudecz, and Karl Mechtler. "Anion-exchange chromatography of phosphopeptides: Weak anion exchange versus strong anion exchange and anion-exchange chromatography versus electrostatic repulsion-hydrophilic interaction chromatography". In: *Analytical Chemistry* 87 (9 May 2015), pp. 4704–4711. ISSN: 15206882. DOI: 10.1021/AC504420C/ASSET/IMAGES/LARGE/AC-2014-04420C_0008.JPEG.

[141] Mark A. Strege. "Hydrophilic interaction chromatography-electrospray mass spectrometry analysis of polar compounds for natural product drug discovery". In: *Analytical chemistry* 70 (13 July 1998), pp. 2439–2445. ISSN: 0003-2700. DOI: 10.1021/AC9802271.

[142] Yusi Cui et al. "Counterion Optimization Dramatically Improves Selectivity for Phosphopeptides and Glycopeptides in Electrostatic Repulsion-Hydrophilic Interaction Chromatography". In: *Analytical Chemistry* 93 (22 June 2021), pp. 7908–7916. ISSN: 15206882. DOI: 10.1021/ACS.ANALCHEM.1C00615/ASSET/IMAGES/LARGE/AC1C00615_0007.JPEG.

[143] Kyung Cho Cho et al. "Developing Workflow for Simultaneous Analyses of Phosphopeptides and Glycopeptides". In: *ACS Chemical Biology* 14 (1 Jan. 2019), pp. 58–66. ISSN: 15548937. DOI: 10.1021/ACSCHEMBIO.8B00902/SUPPL_FILE/CB8B00902_SI_004.XLSX.

[144] Qi Lu et al. "High-Efficiency Phosphopeptide and Glycopeptide Simultaneous Enrichment by Hydrogen Bond-based Bifunctional Smart Polymer". In: *Analytical chemistry* 92 (9 May 2020), pp. 6269–6277. ISSN: 1520-6882. DOI: 10.1021/ACS.ANALCHEM.9B02643.

[145] Yusi Cui et al. "Finding the Sweet Spot in ERLIC Mobile Phase for Simultaneous Enrichment of N-Glyco and Phosphopeptides". In: *Journal of the American Society for Mass Spectrometry* 30 (12 Dec. 2019), pp. 2491–2501. ISSN: 18791123. DOI: 10.1007/S13361-019-02230-6/FIGURES/4.

[146] Reta Birhanu Kitata et al. "A data-independent acquisition-based global phosphoproteomics system enables deep profiling". In: *Nature Communications 2021 12:1* 12 (1 May 2021), pp. 1–14. ISSN: 2041-1723. DOI: 10.1038/s41467-021-22759-z.

[147] Roland Bruderer et al. "Optimization of Experimental Parameters in Data-Independent Mass Spectrometry Significantly Increases Depth and Reproducibility of Results". In: *Molecular & cellular proteomics : MCP* 16 (12 Dec. 2017), pp. 2296–2309. ISSN: 1535-9484. DOI: 10.1074/MCP.RA117.000314.

[148] Jesper V. Olsen et al. "Global, In Vivo, and Site-Specific Phosphorylation Dynamics in Signaling Networks". In: *Cell* 127 (3 Nov. 2006), pp. 635–648. ISSN: 00928674. DOI: 10.1016/J.CELL.2006.09.026/ATTACHMENT/5BA60E23-CFB8-4412-A1A4-891107A3D788/MMC2.XLS.

[149] Qijun Zhao et al. "Proteomic analysis reveals that naturally produced citral can significantly disturb physiological and metabolic processes in the rice blast fungus Magnaporthe oryzae". In: *Pesticide Biochemistry and Physiology* 175 (June 2021), p. 104835. ISSN: 0048-3575. DOI: 10.1016/J.PESTBP.2021.104835.

[150] Jaclyn Gowen Kalmar et al. "Comparative Proteomic Analysis of Wild Type and Mutant Lacking an SCF E3 Ligase F-Box Protein in Magnaporthe oryzae". In: *Journal of proteome research* 19 (9 Sept. 2020), pp. 3761–3768. ISSN: 1535-3907. DOI: 10.1021/ACS.JPROTEOME.0C00294.

[151] Veit Schw-mmle, Ileana Rodr-guez Le-n, and Ole N-rregaard Jensen. "Assessment and improvement of statistical tools for comparative proteomics analysis of sparse data sets with few experimental replicates". In: *Journal of proteome research* 12 (9 Sept. 2013), pp. 3874–3883. ISSN: 1535-3907. DOI: 10.1021/PR400045U.

[152] Gordon K. Smyth, Jo-lle Michaud, and Hamish S. Scott. "Use of within-array replicate spots for assessing differential expression in microarray experiments". In: *Bioinformatics (Oxford, England)* 21 (9 May 2005), pp. 2067–2075. ISSN: 1367-4803. DOI: 10.1093/BIOINFORMATICS/BTI270.

[153] Giulia Mantini et al. "Computational Analysis of Phosphoproteomics Data in Multi-Omics Cancer Studies". In: *PROTEOMICS* 21 (3-4 Feb. 2021), p. 1900312. ISSN: 1615-9861. DOI: 10.1002/PMIC.201900312.

[154] Scott P. Lyons et al. "Proteomics and phosphoproteomics datasets of a muscle-specific STIM1 loss-of-function mouse model". In: *Data in Brief* 42 (June 2022), p. 108051. ISSN: 2352-3409. DOI: 10.1016/J.DIB.2022.108051.

[155] Aaron J. Storey et al. "ProteoViz: a tool for the analysis and interactive visualization of phosphoproteomics data". In: *Molecular Omics* 16 (4 Aug. 2020), pp. 316–326. ISSN: 25154184. DOI: 10.1039/C9MO00149B.

[156] Mingyi Liu and Ashok Dongre. "Proper imputation of missing values in proteomics datasets for differential expression analysis". In: *Briefings in bioinformatics* 22 (3 May 2021). ISSN: 1477-4054. DOI: 10.1093/BIB/BBAA112.

[157] Hassan Dihazi, Renate Kessler, and Klaus Eschrich. "High osmolarity glycerol (HOG) pathway-induced phosphorylation and activation of 6-phosphofructo-2-kinase are essential for glycerol accumulation and yeast cell proliferation under hyperosmotic stress". In: *The Journal of biological chemistry* 279 (23 June 2004), pp. 23961–23968. ISSN: 0021-9258. DOI: 10.1074/JBC.M312974200.

[158] Douglas Steinley. "K-means clustering: a half-century synthesis". In: *The British journal of mathematical and statistical psychology* 59 (Pt 1 May 2006), pp. 1–34. ISSN: 0007-1102. DOI: 10.1348/000711005X48266.

[159] Martin Ester et al. "A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise". In: (1996).

[160] Natalia A. Petushkova et al. "Applying of hierarchical clustering to analysis of protein patterns in the human cancer-associated liver". In: *PloS one* 9 (8 Aug. 2014). ISSN: 1932-6203. DOI: 10.1371/JOURNAL.PONE.0103950.

[161] Alexander Muir et al. "Down-regulation of TORC2-Ypk1 signaling promotes MAPK-independent survival under hyperosmotic stress". In: *eLife* 4 (AUGUST2015 Aug. 2015). ISSN: 2050084X. DOI: 10.7554/ELIFE.09336.

[162] Susumu Morigasaki et al. "Modulation of TOR complex 2 signaling by the stress-activated MAPK pathway in fission yeast". In: *Journal of cell science* 132 (19 Oct. 2019). ISSN: 1477-9137. DOI: 10.1242/JCS.236133.

[163] Alexander Muir et al. "TORC2-dependent protein kinase Ypk1 phosphorylates ceramide synthase to stimulate synthesis of complex sphingolipids". In: *eLife* 3 (2014). ISSN: 2050-084X. DOI: 10.7554/ELIFE.03779.

[164] Mervi H. Toivari et al. "Metabolic engineering of Saccharomyces cerevisiae for conversion of D-glucose to xylitol and other five-carbon sugars and sugar alcohols". In: *Applied and environmental microbiology* 73 (17 Sept. 2007), pp. 5471–5476. ISSN: 0099-2240. DOI: 10.1128/AEM.02707-06.

[165] William L. Franck et al. "Phosphoproteome Analysis Links Protein Phosphorylation to Cellular Remodeling and Metabolic Adaptation during Magnaporthe oryzae Appressorium Development". In: *Journal of proteome research* 14 (6 June 2015), pp. 2408–2424. ISSN: 1535-3907. DOI: 10.1021/PR501064Q.

[166] Mark O. Collins, Lu Yu, and Jyoti S. Choudhary. "Analysis of protein phosphorylation on a proteome-scale". In: *PROTEOMICS* 7 (16 Aug. 2007), pp. 2751–2768. ISSN: 1615-9861. DOI: 10.1002/PMIC.200700145.

[167] Matthias Blazek et al. "Analysis of fast protein phosphorylation kinetics in single cells on a microfluidic chip". In: *Lab on a chip* 15 (3 Feb. 2015), pp. 726–734. ISSN: 1473-0189. DOI: 10.1039/C4LC00797B.

[168] Daniel R. Pentland et al. "Ras signalling in pathogenic yeasts". In: *Microbial Cell* 5 (2 Feb. 2017), p. 63. ISSN: 23112638. DOI: 10.15698/MIC2018.02.612.

[169] Xia Yan and Nicholas J. Talbot. "Investigating the cell biology of plant infection by the rice blast fungus Magnaporthe oryzae". In: *Current opinion in microbiology* 34 (Dec. 2016), pp. 147–153. ISSN: 1879-0364. DOI: 10.1016/J.MIB.2016.10.001.

[170] Tengsheng Zhou et al. "The glycogen synthase kinase MoGsk1, regulated by Mps1 MAP kinase, is required for fungal development and pathogenicity in Magnaporthe oryzae". In: *Scientific reports* 7 (1 Dec. 2017). ISSN: 2045-2322. DOI: 10.1038/S41598-017-01006-W.

[171] Diane O. Inglis and Gavin Sherlock. "Ras Signaling Gets Fine-Tuned: Regulation of Multiple Pathogenic Traits of Candida albicans". In: *Eukaryotic Cell* 12 (10 Oct. 2013), p. 1316. ISSN: 15359778. DOI: 10.1128/EC.00094-13.

[172] Sara Manzanares-Estreder et al. "Multilayered control of peroxisomal activity upon salt stress in Saccharomyces cerevisiae". In: *Molecular microbiology* 104 (5 June 2017), pp. 851–868. ISSN: 1365-2958. DOI: 10.1111/MMI.13669.

[173] Amparo Pascual-Ahuir et al. "Ask yeast how to burn your fats: lessons learned from the metabolic adaptation to salt stress". In: *Current genetics* 64 (1 Feb. 2018), pp. 63–69. ISSN: 1432-0983. DOI: 10.1007/S00294-017-0724-5.

[174] Carmen Herrero-de-dios et al. "Hog1 Controls Lipids Homeostasis Upon Osmotic Stress in Candida albicans". In: *Journal of fungi (Basel, Switzerland)* 6 (4 Dec. 2020), pp. 1–13. ISSN: 2309-608X. DOI: 10.3390/JOF6040355.

[175] E. Thines, R. W.S. Weber, and N. J. Talbot. "MAP kinase and protein kinase A-dependent mobilization of triacylglycerol and glycogen during appressorium turgor generation by Magnaporthe grisea". In: *The Plant cell* 12 (9 2000), pp. 1703–1718. ISSN: 1040-4651. DOI: 10.1105/TPC.12.9.1703.

[176] Yeonyee Oh et al. "Transcriptome analysis reveals new insight into appressorium formation and function in the rice blast fungus Magnaporthe oryzae". In: *Genome biology* 9 (5 May 2008). ISSN: 1474-760X. DOI: 10.1186/GB-2008-9-5-R85.

[177] Michael J. Kershaw et al. "Conidial Morphogenesis and Septin-Mediated Plant Infection Require Smo1, a Ras GTPase-Activating Protein in Magnaporthe oryzae". In: *Genetics* 211 (1 Jan. 2019), pp. 151–167. ISSN: 1943-2631. DOI: 10.1534/GENETICS.118.301490.

[178] Xuan Cai et al. "Deubiquitinase Ubp3 regulates ribophagy and deubiquitinates Smo1 for appressorium-mediated infection by Magnaporthe oryzae". In: *Molecular plant pathology* 23 (6 June 2022), pp. 832–844. ISSN: 1364-3703. DOI: 10.1111/MPP.13196.

[179] Roy Auty et al. “Purification of active TFIID from Saccharomyces cerevisiae. Extensive promoter contacts and co-activator function”. In: *The Journal of biological chemistry* 279 (48 Nov. 2004), pp. 49973–49981. ISSN: 0021-9258. DOI: 10.1074/JBC.M409849200.

[180] Carme Sol- et al. “Control of Ubp3 ubiquitin protease activity by the Hog1 SAPK modulates transcription upon osmostress”. In: *The EMBO journal* 30 (16 Aug. 2011), pp. 3274–3284. ISSN: 1460-2075. DOI: 10.1038/EMBOJ.2011.227.

[181] Shengpei Zhang et al. “System-Wide Characterization of MoArf GTPase Family Proteins and Adaptor Protein MoGga1 Involved in the Development and Pathogenicity of Magnaporthe oryzae”. In: *mBio* 10 (5 2019). ISSN: 2150-7511. DOI: 10.1128/MBIO.02398-19.

[182] Ruth Kabeche et al. “Eisosomes Regulate Phosphatidylinositol 4,5-Bisphosphate (PI(4,5)P2) Cortical Clusters and Mitogen-activated Protein (MAP) Kinase Signaling upon Osmotic Stress”. In: *The Journal of biological chemistry* 290 (43 Oct. 2015), pp. 25960–25973. ISSN: 1083-351X. DOI: 10.1074/JBC.M115.674192.

[183] Runmin Wei et al. “Missing Value Imputation Approach for Mass Spectrometry-based Metabolomics Data”. In: *Scientific reports* 8 (1 Dec. 2018). ISSN: 2045-2322. DOI: 10.1038/S41598-017-19120-0.

[184] Gouji Toyokawa and Takashi Seto. “Updated Evidence on the Mechanisms of Resistance to ALK Inhibitors and Strategies to Overcome Such Resistance: Clinical and Preclinical Data”. In: *Oncology research and treatment* 38 (6 June 2015), pp. 291–298. ISSN: 2296-5262. DOI: 10.1159/000430852.

[185] Leonidas C. Platanias et al. “The type I interferon receptor mediates tyrosine phosphorylation of insulin receptor substrate 2”. In: *The Journal of biological chemistry* 271 (1 Jan. 1996), pp. 278–282. ISSN: 0021-9258. DOI: 10.1074/JBC.271.1.278.

[186] Ruey Hwa Chen et al. “Interleukin-6 inhibits transforming growth factor-beta-induced apoptosis through the phosphatidylinositol 3-kinase/Akt and signal transducers and activators of transcription 3 pathways”. In: *The Journal of biological chemistry* 274 (33 Aug. 1999), pp. 23013–23019. ISSN: 0021-9258. DOI: 10.1074/JBC.274.33.23013.

[187] G. Joshi-Tope et al. "Reactome: a knowledgebase of biological pathways". In: *Nucleic acids research* 33 (Database issue Jan. 2005). ISSN: 1362-4962. DOI: 10.1093/NAR/GKI072.

[188] Thomas Kelder et al. "Mining biological pathways using WikiPathways web services". In: *PloS one* 4 (7 July 2009). ISSN: 1932-6203. DOI: 10.1371/JOURNAL.PONE.0006447.

[189] Hiroyuki Ogata et al. "Computation with the KEGG pathway database". In: *Bio Systems* 47 (1-2 June 1998), pp. 119–128. ISSN: 0303-2647. DOI: 10.1016/S0303-2647(98)00017-3.

[190] Matteo Parri and Paola Chiarugi. "Rac and Rho GTPases in cancer cell motility control". In: *Cell Communication and Signaling 2010 8:1* 8 (1 Sept. 2010), pp. 1–14. ISSN: 1478-811X. DOI: 10.1186/1478-811X-8-23.

[191] Yingyue Lou et al. "Role of RhoC in cancer cell migration". In: *Cancer Cell International 2021 21:1* 21 (1 Oct. 2021), pp. 1–16. ISSN: 1475-2867. DOI: 10.1186/S12935-021-02234-X.

[192] Moredreck Chibi et al. "RBBP6 Interacts with Multifunctional Protein YB-1 through Its RING Finger Domain, Leading to Ubiquitination and Proteosomal Degradation of YB-1". In: *Journal of Molecular Biology* 384 (4 Dec. 2008), pp. 908–916. ISSN: 0022-2836. DOI: 10.1016/J.JMB.2008.09.060.

[193] Ki Hyuk Shin et al. "Heterogeneous nuclear ribonucleoprotein G shows tumor suppressive effect against oral squamous cell carcinoma cells". In: *Clinical cancer research : an official journal of the American Association for Cancer Research* 12 (10 May 2006), pp. 3222–3228. ISSN: 1078-0432. DOI: 10.1158/1078-0432.CCR-05-2656.

[194] Dong Wan Kim et al. "Intracranial and whole-body response of ceritinib in ALK inhibitor-na-ve and previously ALK inhibitor-treated patients with ALK-rearranged non-small-cell lung cancer (NSCLC): updated results from the phase 1, multicentre, open-label ASCEND-1 trial". In: *The Lancet. Oncology* 17 (4 Apr. 2016), p. 452. ISSN: 14745488. DOI: 10.1016/S1470-2045(15)00614-2.

[195] Joost C.M. Uitdehaag et al. "Combined cellular and biochemical profiling to identify predictive drug response biomarkers for kinase inhibitors approved for clinical use between 2013 and 2017". In: *Molecular Cancer Therapeutics* 18 (2 Feb. 2019), pp. 470–481. ISSN: 15388514. DOI: 10.1158/1535-7163.MCT-18-0877/87367/AM/COMBINED-CELLULAR-AND-BIOCHEMICAL-PROFILING-TO.

[196] Yukiya Sako et al. "Development of an orally available inhibitor of CLK1 for skipping a mutated dystrophin exon in Duchenne muscular dystrophy". In: *Scientific Reports 2017 7:1* 7 (1 May 2017), pp. 1–9. ISSN: 2045-2322. DOI: 10.1038/srep46126.

[197] Fred M. Moeslein, Michael P. Myers, and Gary E. Landreth. "The CLK family kinases, CLK1 and CLK2, phosphorylate and activate the tyrosine phosphatase, PTP-1B". In: *The Journal of biological chemistry* 274 (38 Sept. 1999), pp. 26697–26704. ISSN: 0021-9258. DOI: 10.1074/JBC.274.38.26697.

[198] Luis E. Arias-Romero et al. "Activation of Src by Protein Tyrosine Phosphatase-1B is required for ErbB2 transformation of human breast epithelial cells". In: *Cancer research* 69 (11 June 2009), p. 4582. ISSN: 00085472. DOI: 10.1158/0008-5472.CAN-08-4001.

[199] Christopher D. Chouinard et al. "Improved Sensitivity and Separations for Phosphopeptides using Online Liquid Chromotography Coupled with Structures for Lossless Ion Manipulations Ion Mobility-Mass Spectrometry". In: *Analytical Chemistry* 90 (18 Sept. 2018), pp. 10889–10896. ISSN: 15206882. DOI: 10.1021/ACS.ANALCHEM.8B02397/ASSET/IMAGES/LARGE/AC-2018-023979_0005.JPEG.

[200] Kosuke Ogata, Chih Hsiang Chang, and Yasushi Ishihama. "Effect of Phosphorylation on the Collision Cross Sections of Peptide Ions in Ion Mobility Spectrometry". In: *Mass Spectrometry* 10 (1 2021), pp. 1–8. ISSN: 21865116. DOI: 10.5702/MASSSPECTROMETRY.A0093.

6 Supplementary data

Window	center (m/z)	Isolation Window (m/z)	m/z start	m/z end
1	392.58	95.2	344.98	440.18
2	459.56	39.8	439.66	479.46
3	494.25	30.6	478.95	509.55
4	522.78	27.5	509.03	536.53
5	548.53	25	536.03	561.03
6	572.89	24.7	560.54	585.24
7	596.9	24.3	584.75	609.05
8	620.81	24.5	608.56	633.06
9	645.09	25.1	632.54	657.64
10	670.1	26	657.1	683.1
11	696.1	27	682.6	709.6
12	723.35	28.5	709.1	737.6
13	752.37	30.6	737.07	767.67
14	783.12	31.9	767.17	799.07
15	816.29	35.4	798.59	833.99
16	852.84	38.7	833.49	872.19
17	893.43	43.5	871.68	915.18
18	939.95	50.6	914.65	965.25
19	997.57	65.7	964.72	1030.42
20	1070.86	81.9	1029.91	1111.81
21	1180.66	138.7	1111.31	1250.01

Table 6.1: Variable window sizes for DIA acquisition with Orbitrap Exploris 480. Applied for the DIA data acquisition of the *M.oryzae* osmostress resource

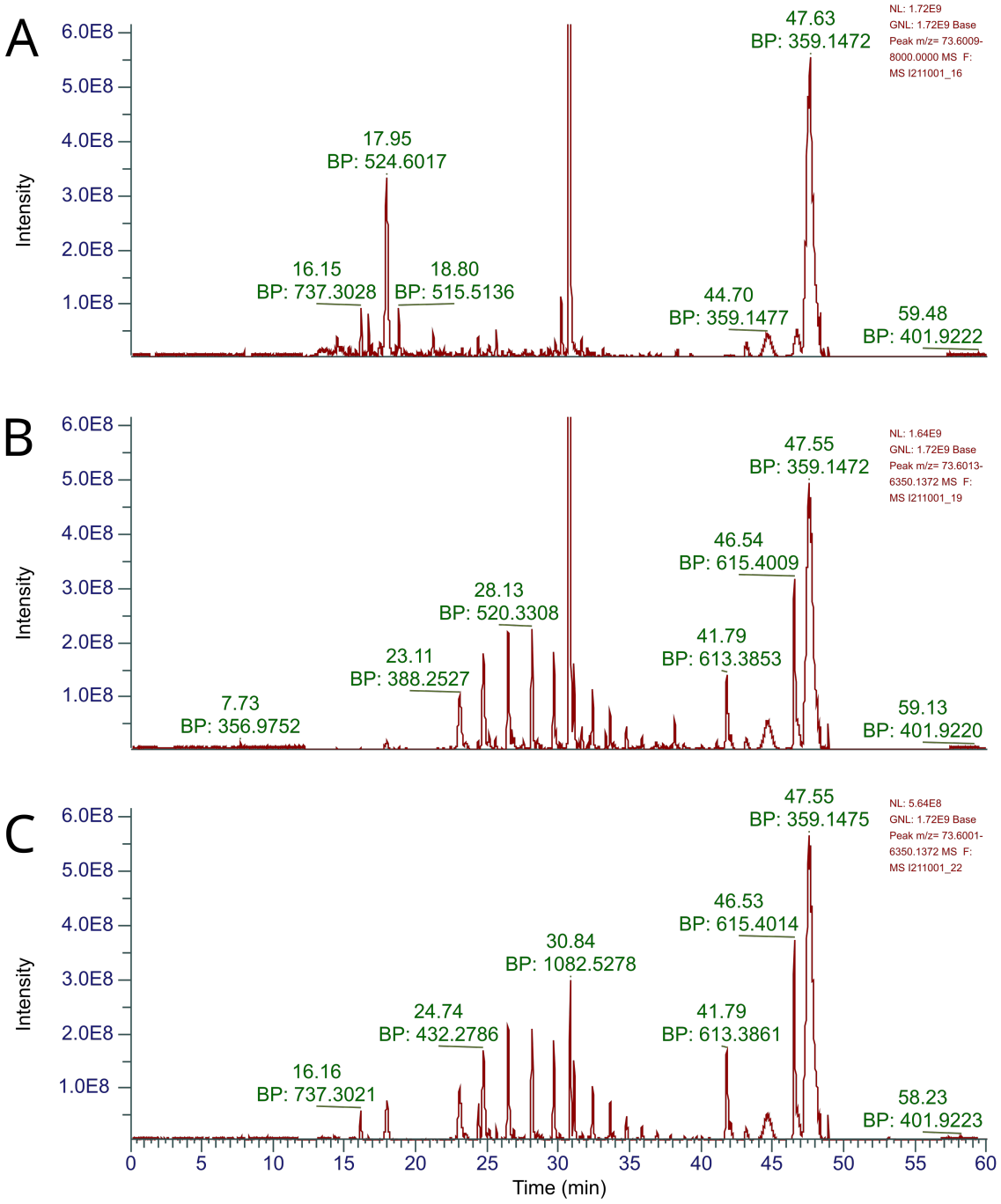

Figure 6.1: Example chromatograms of 25 µg mouse brain phosphopeptides after enrichment A) without desalting B) after SePak tC18 desalting C) after Oasis HLB desalting

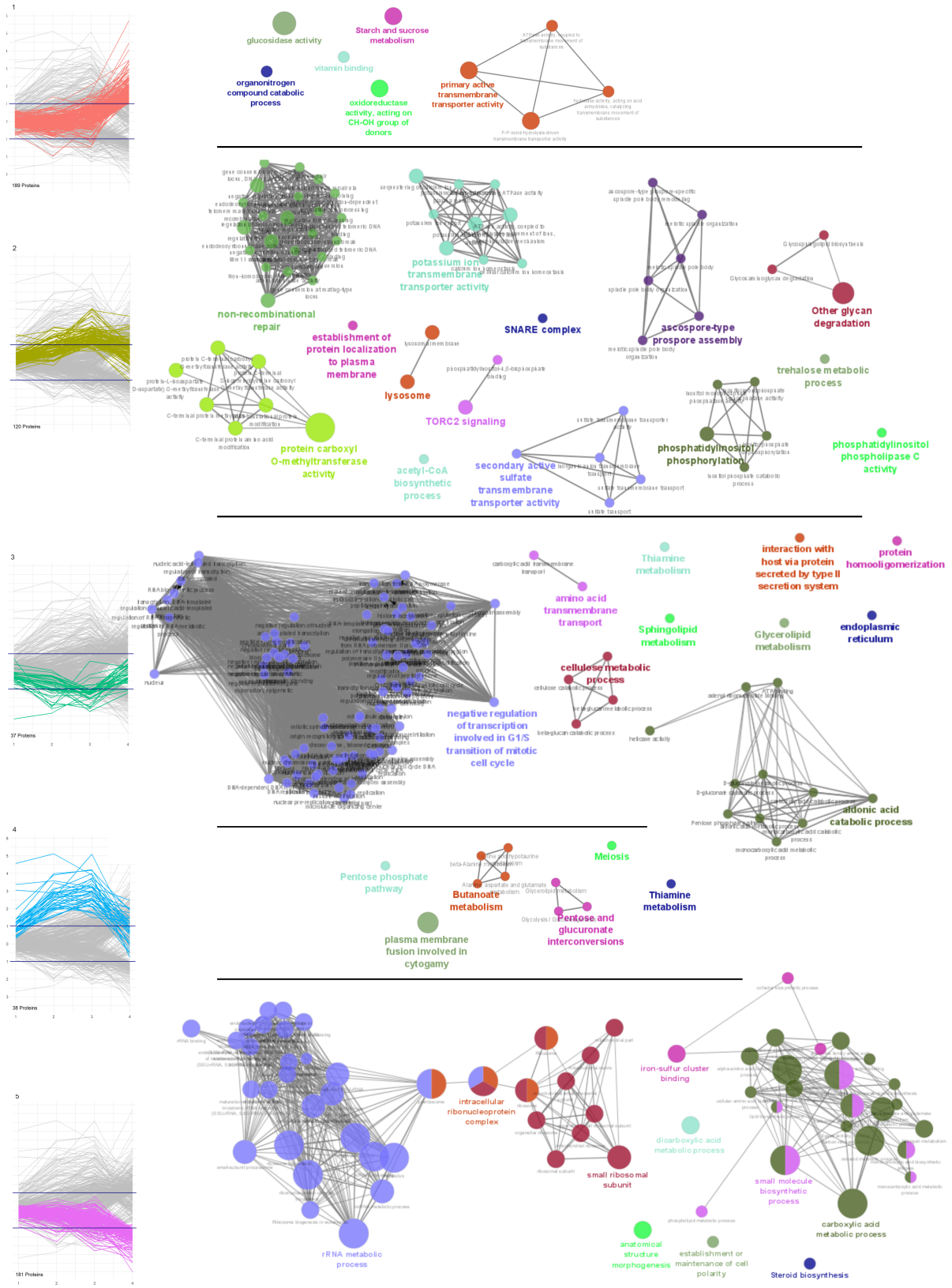

Figure 6.2: Gene ontology enrichment networks for each cluster of temporal proteome response of wild type *M.oryzae* during the time course of 24h after osmotic stress.

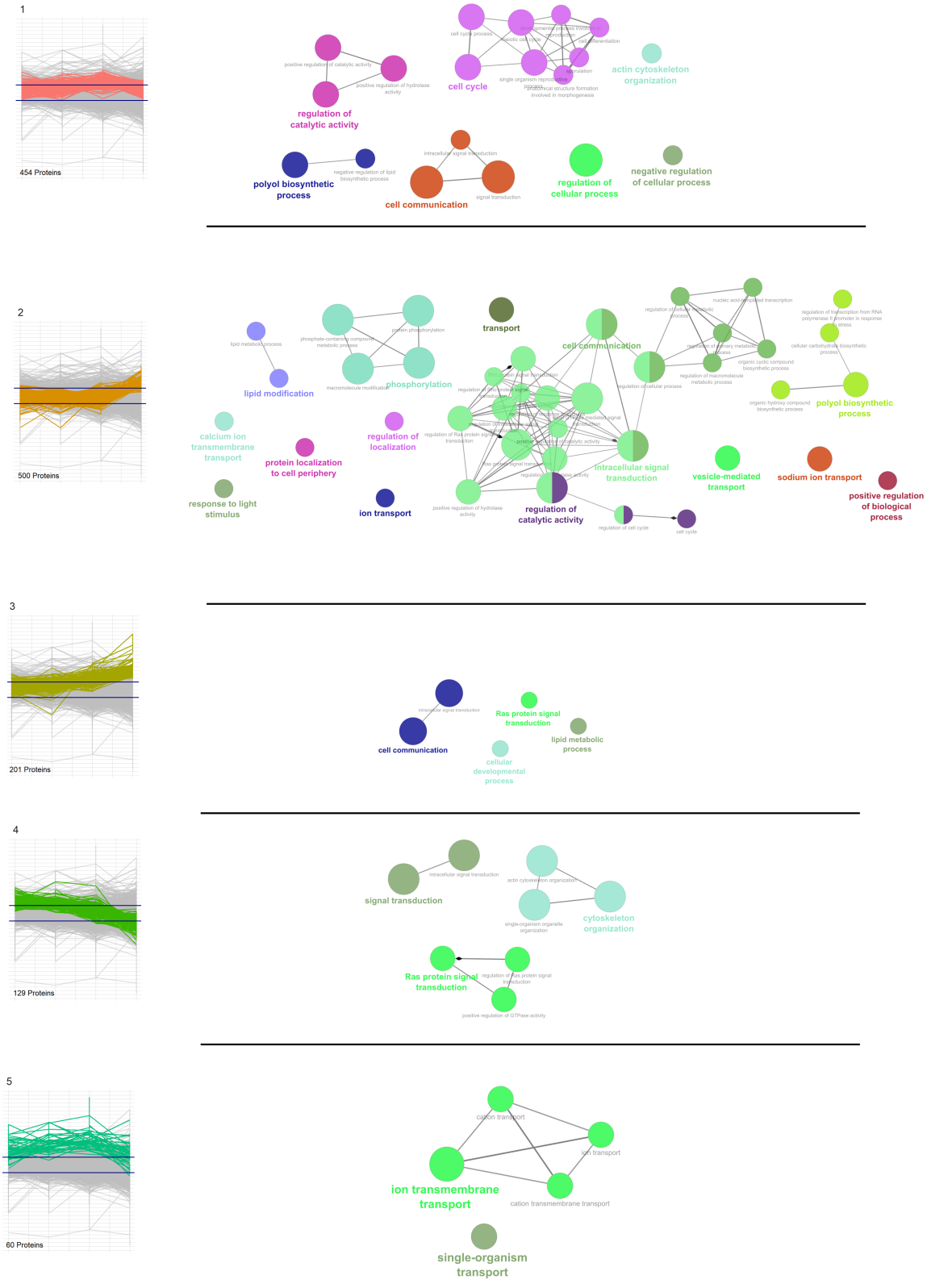

Figure 6.3: Gene ontology enrichment networks for each cluster of temporal phosphopeptide response (1/2) of wild type *M.oryzae* during the time course of 24h after osmotic stress.

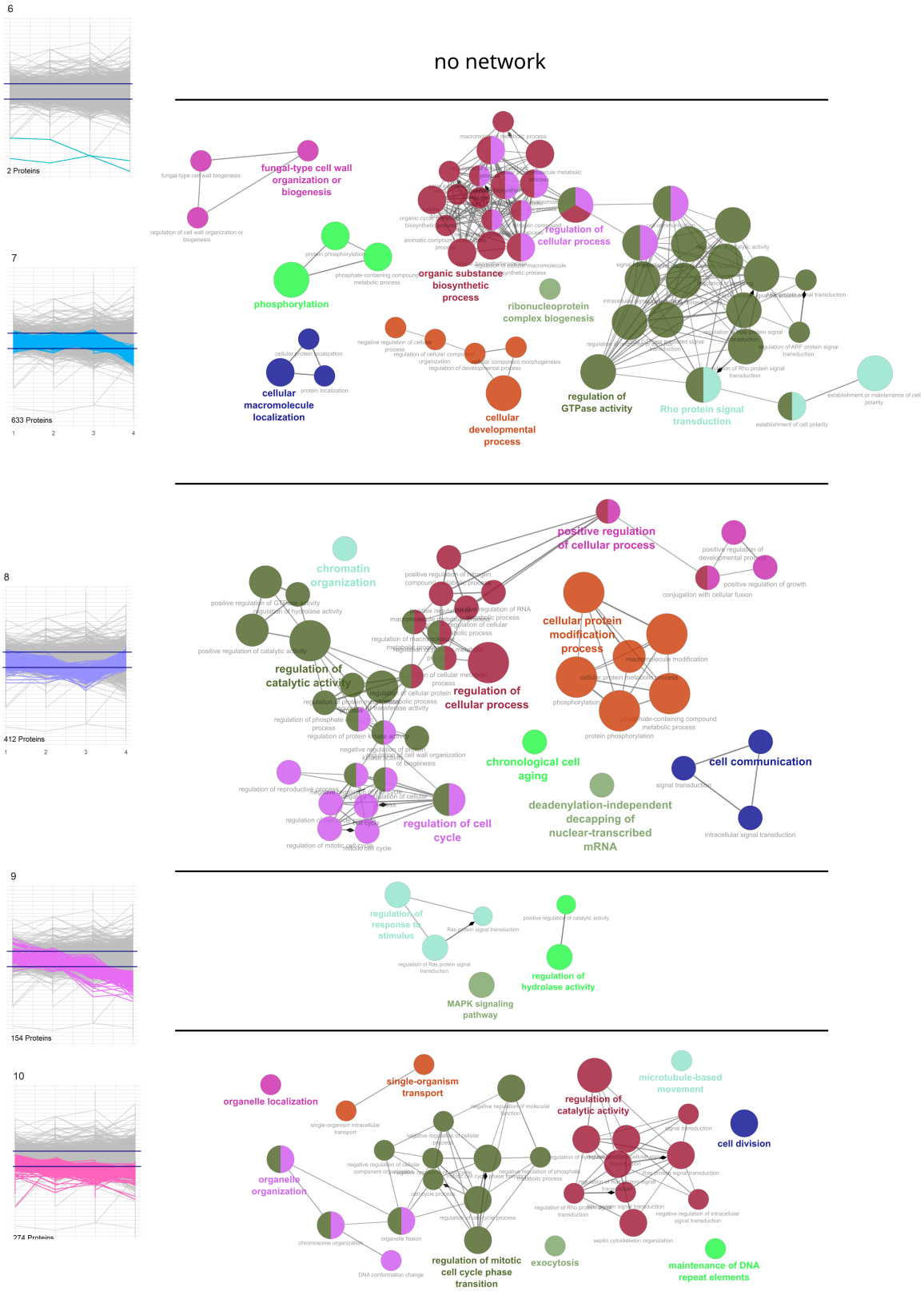

Figure 6.4: Gene ontology enrichment networks for each cluster of temporal phosphopeptide response (2/2) of wild type *M.oryzae* during the time course of 24h after osmotic stress.

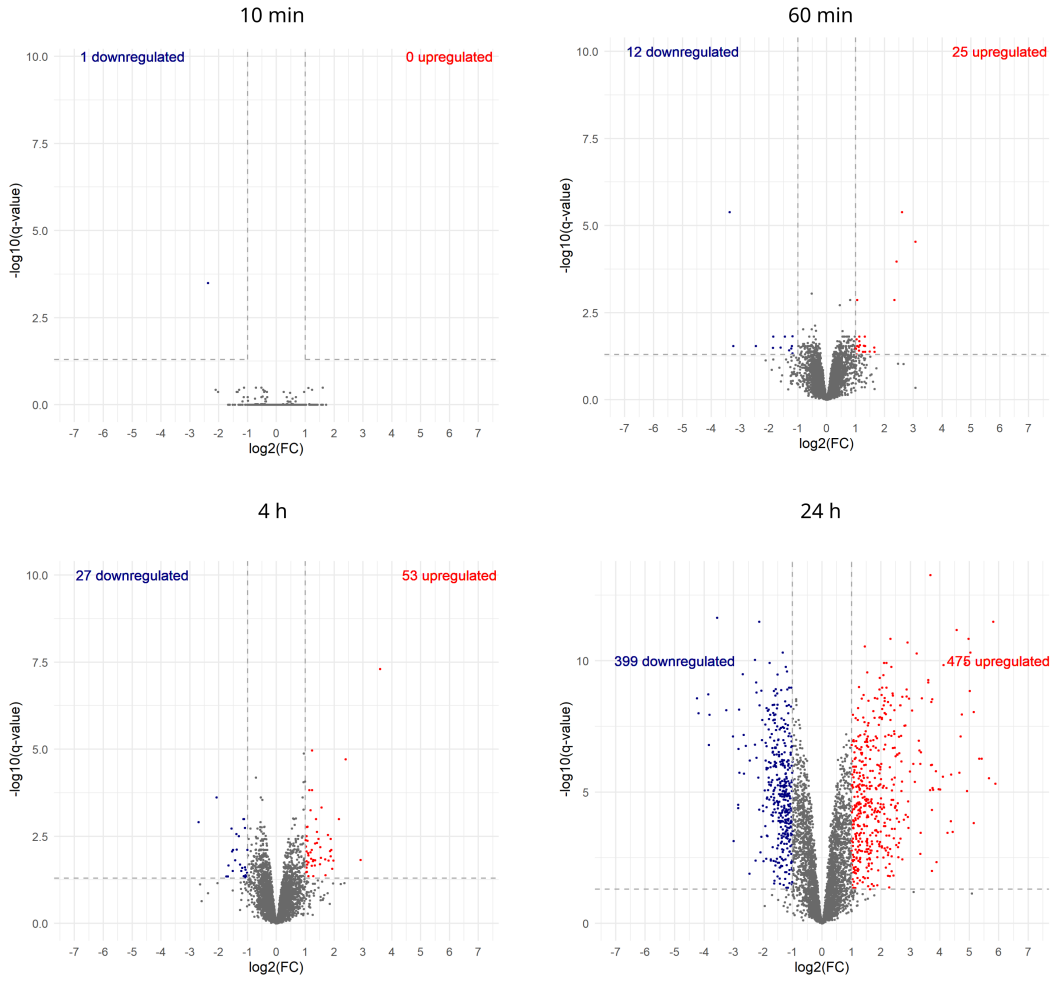

Figure 6.5: Volcano plots of temporal changes in proteome of the irreversibly adapted Hog1 deletion mutant during the time course of 24h after osmotic stress.

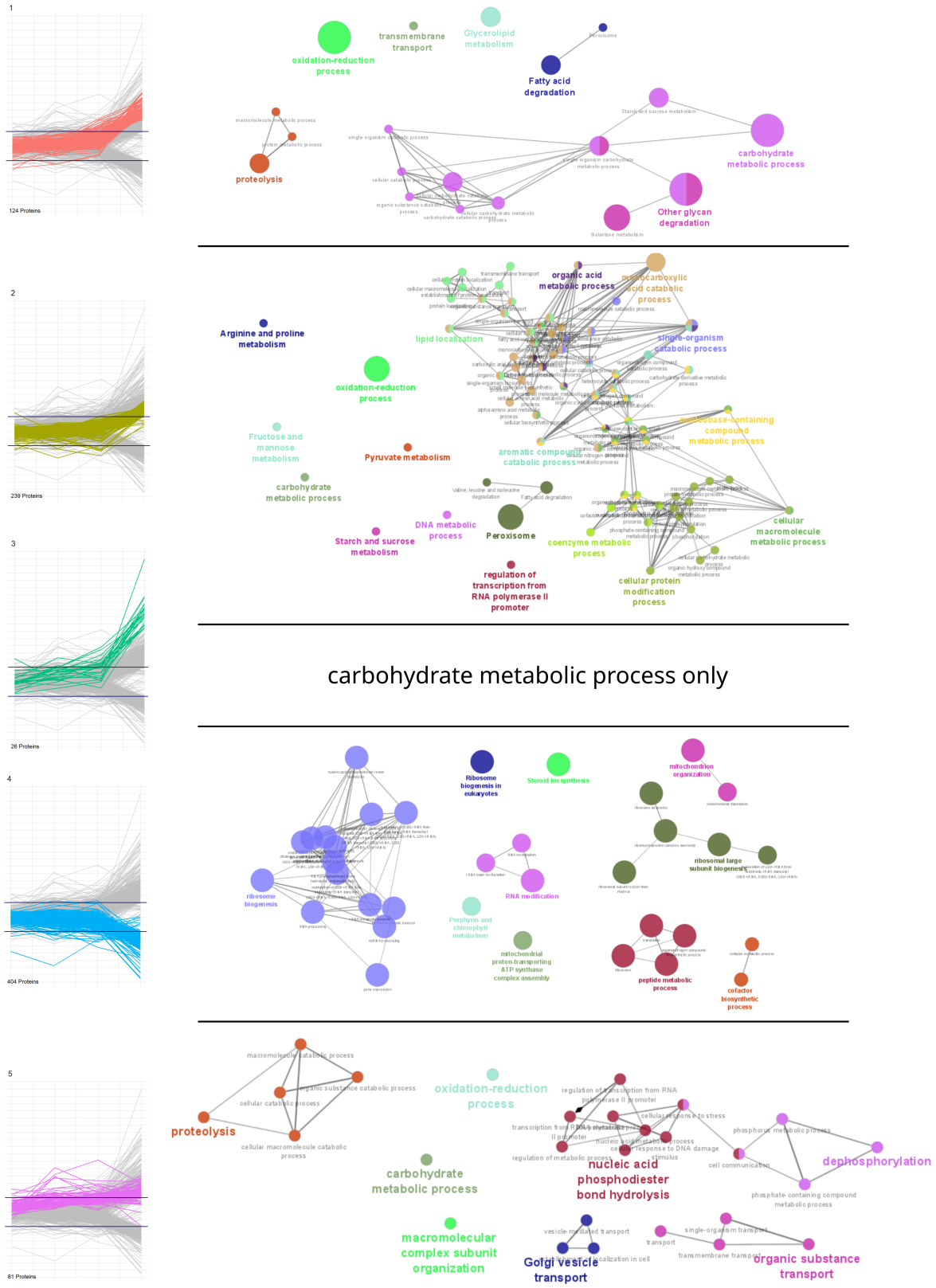

Figure 6.6: Gene ontology enrichment networks for each cluster of temporal proteome response in the adapted Hog1 deletion mutant during the time course of 24h after osmotic stress.

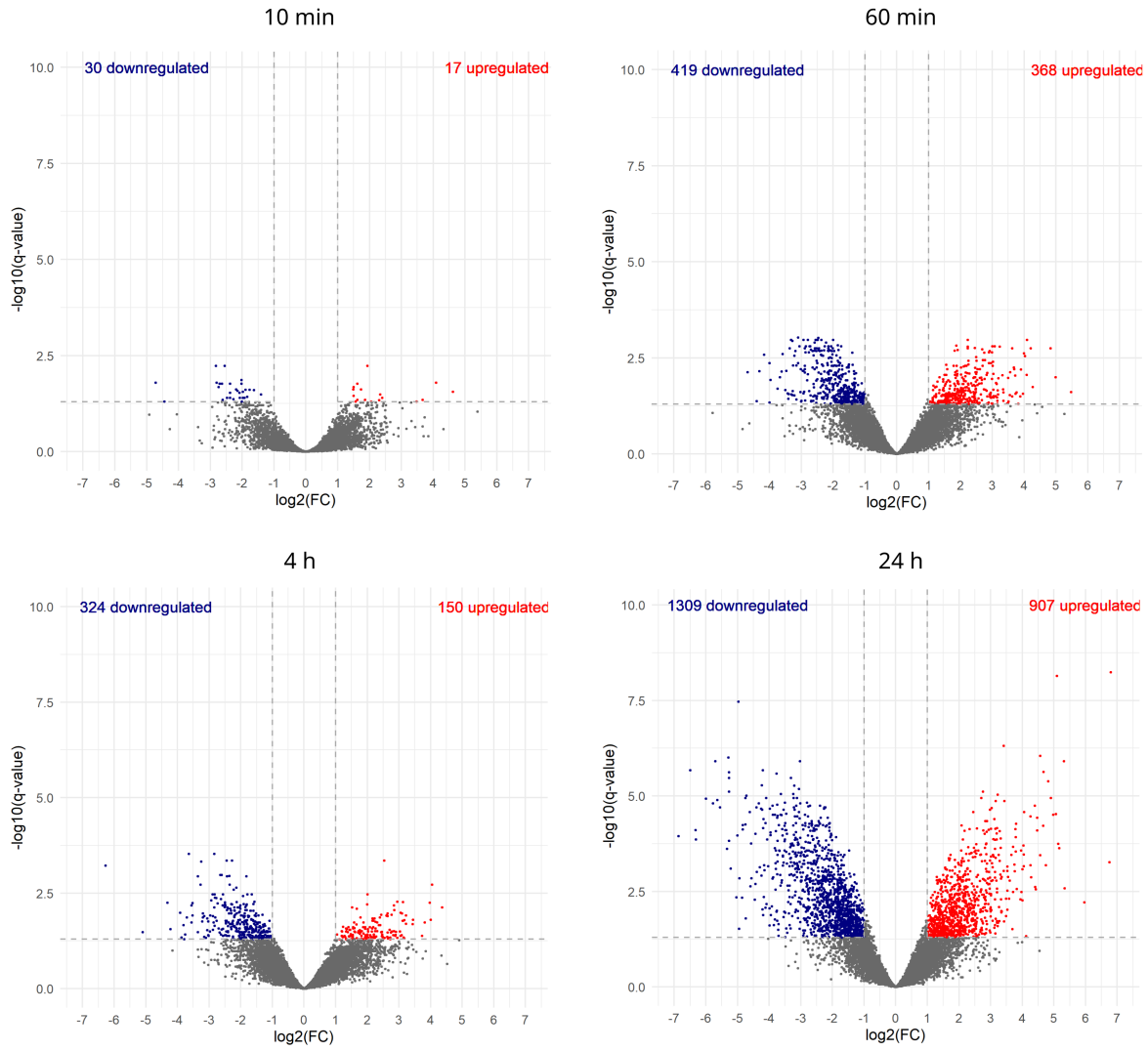

Figure 6.7: Volcano plots of temporal changes in phosphopeptides of the irreversibly adapted Hog1 deletion mutant during the time course of 24h after osmotic stress.

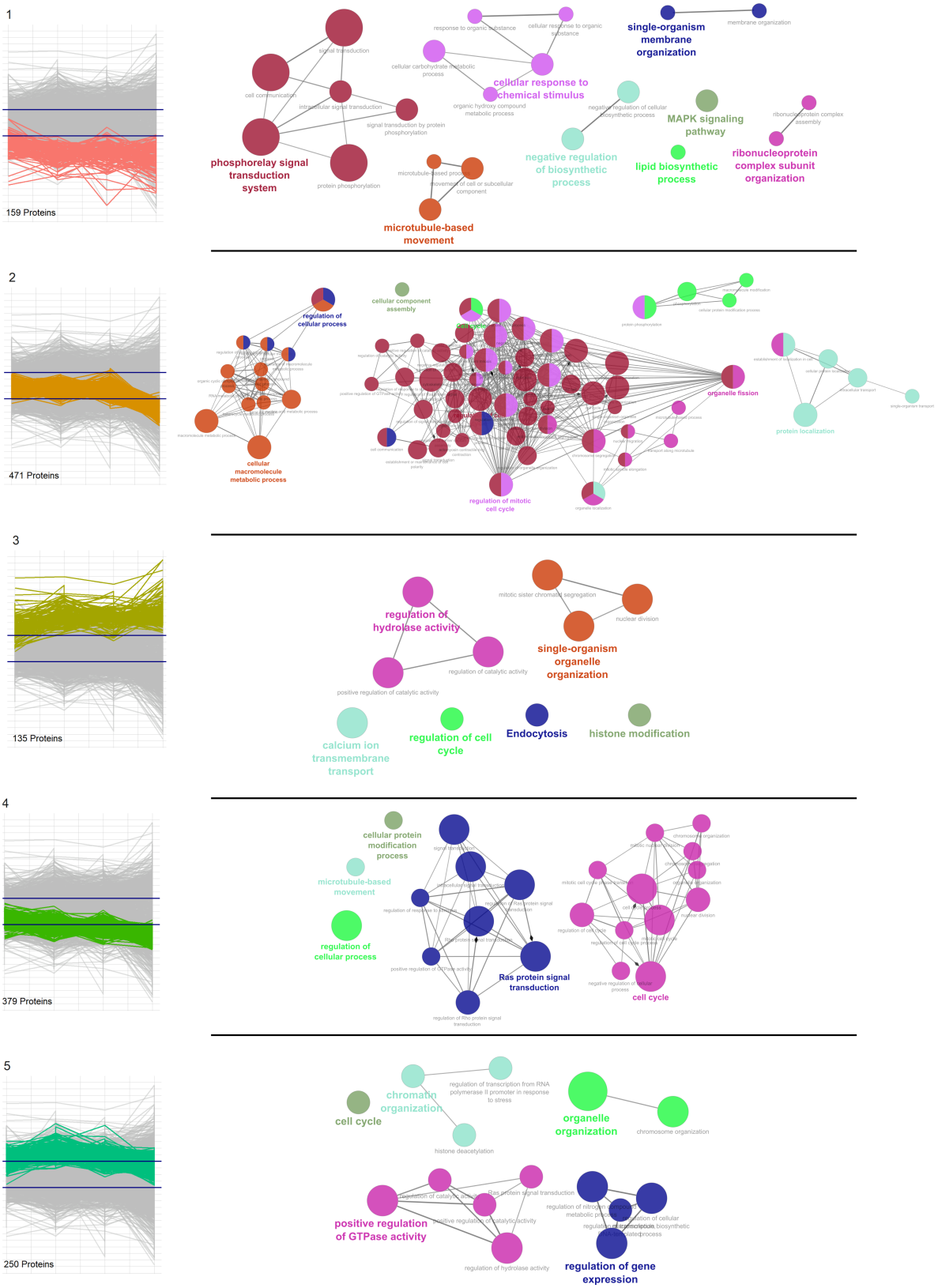

Figure 6.8: Gene ontology enrichment networks for each cluster of temporal phosphopeptide response in the adapted Hog1 deletion mutant (1/2) during the time course of 24h after osmotic stress.

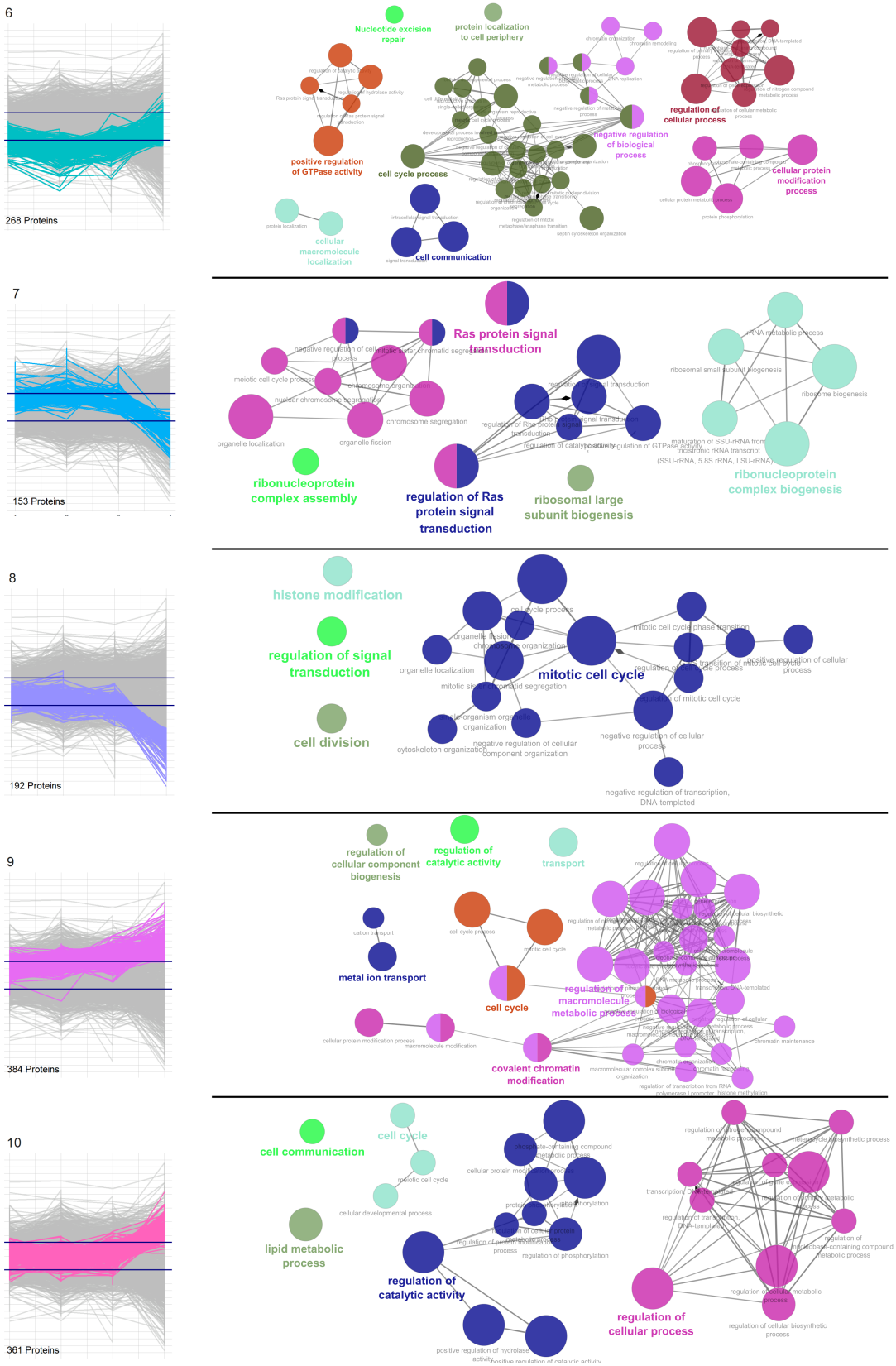

Figure 6.9: Gene ontology enrichment networks for each cluster of temporal phosphopeptide response in the adapted Hog1 deletion mutant (2/2) during the time course of 24h after osmotic stress.

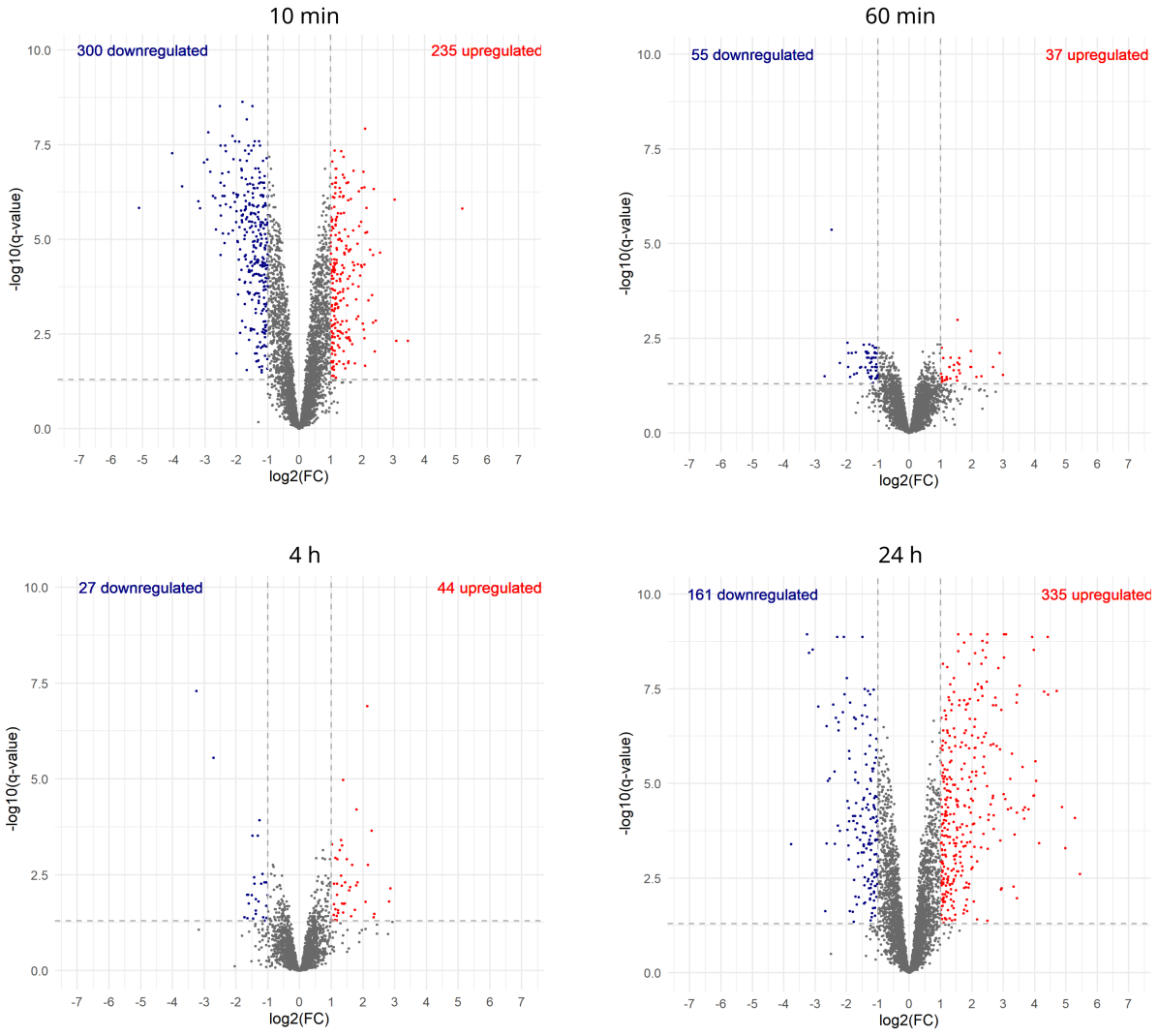

Figure 6.10: Volcano plots of temporal changes in proteome of the Hog1 deletion mutant (not adapted) during the time course of 24h after osmotic stress.

Figure 6.11: GO enrichment of proteome changes of the Hog1 deletion mutant (not adapted) during the time course of 24h after osmotic stress.

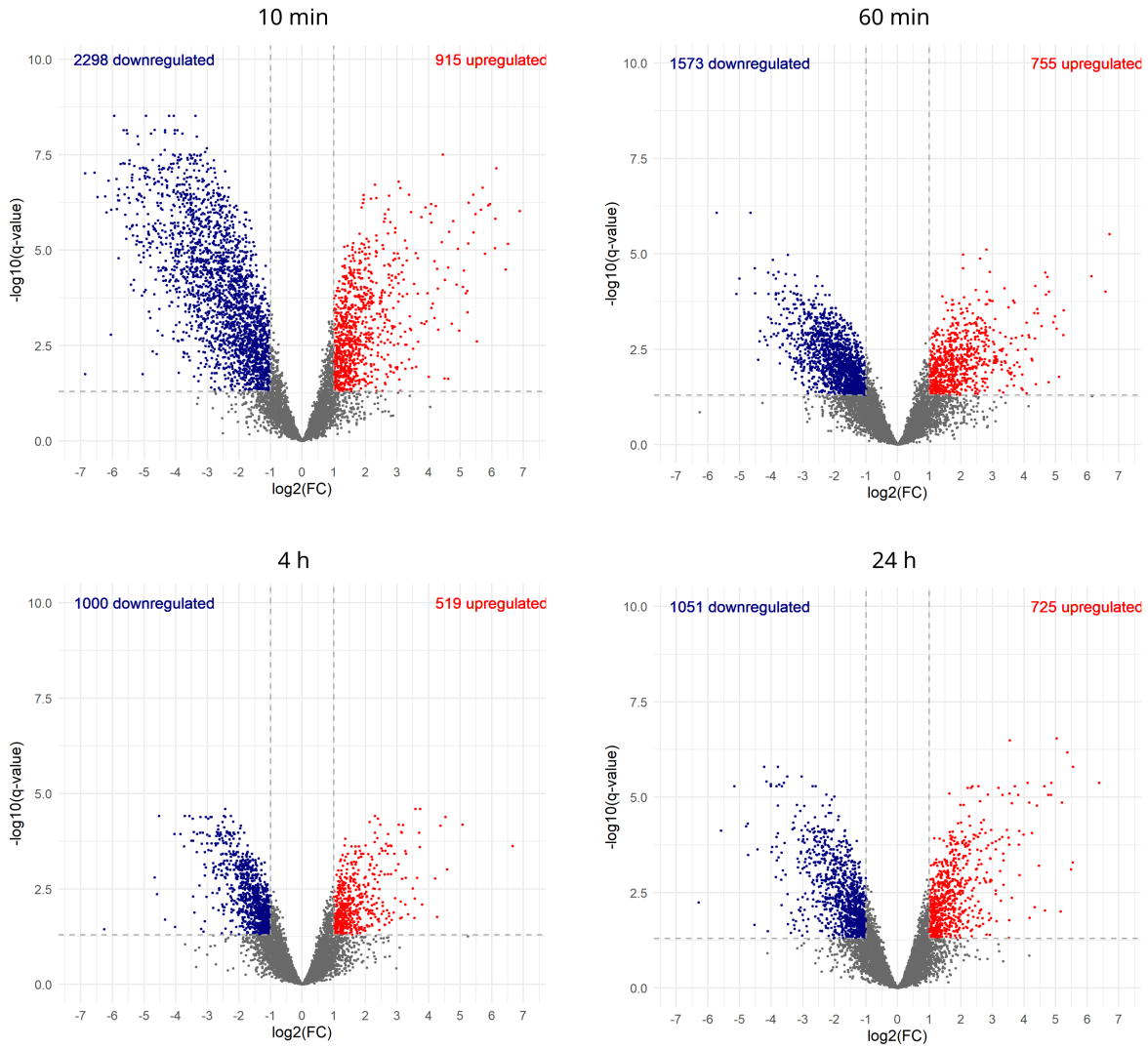

Figure 6.12: Volcano plots of temporal changes in phosphoprotein of the Hog1 deletion mutant (not adapted) during the time course of 24h after osmotic stress.

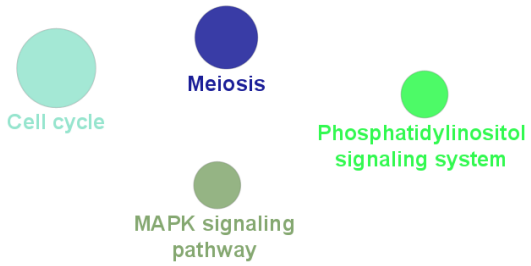

Figure 6.13: GO enrichment of phosphoprotein changes of the Hog1 deletion mutant (not adapted) during the time course of 24h after osmotic stress.

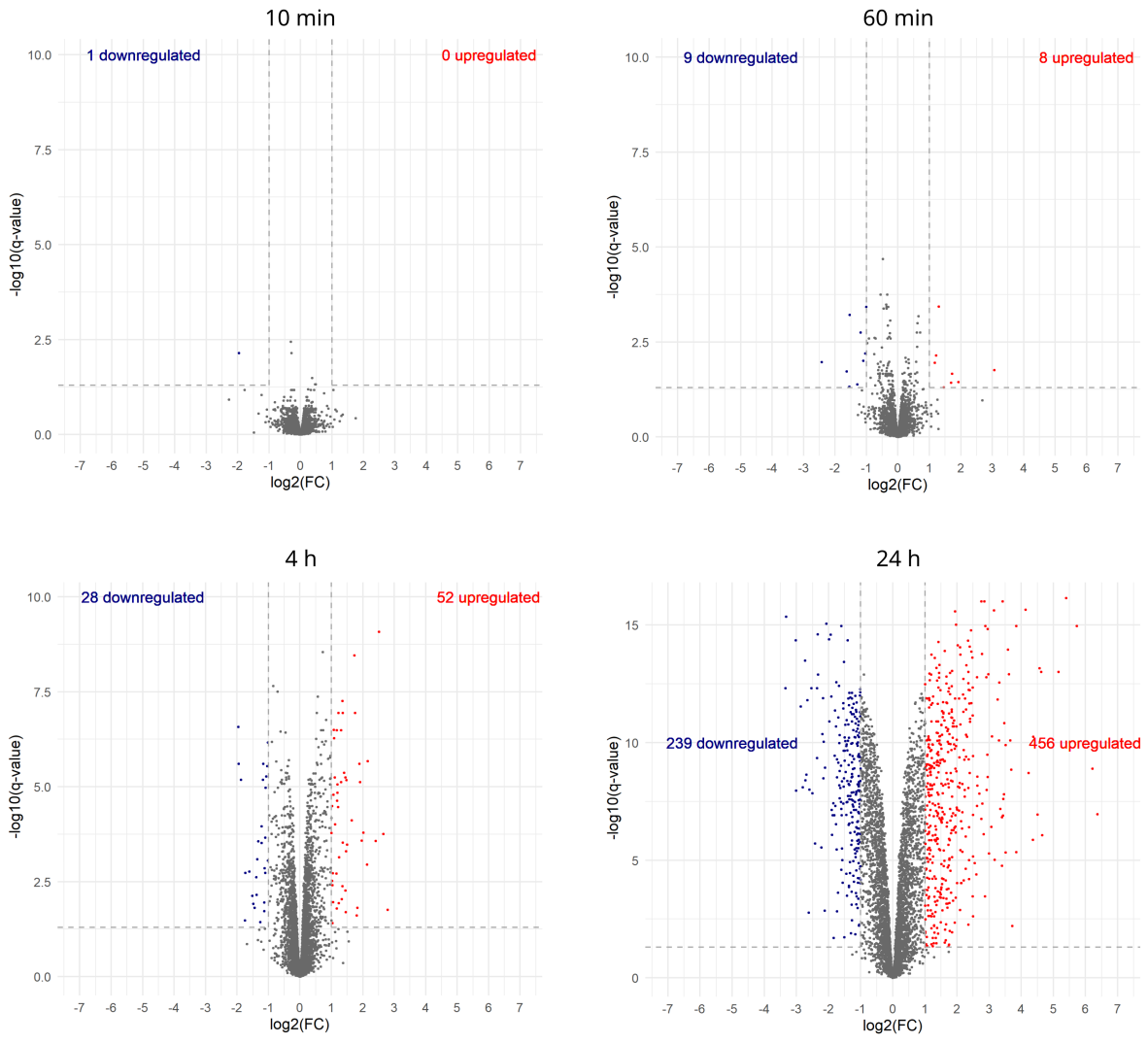

Figure 6.14: Volcano plots of temporal changes in proteome of the Hog1 deletion mutant (reversibly adapted) during the time course of 24h after osmotic stress.

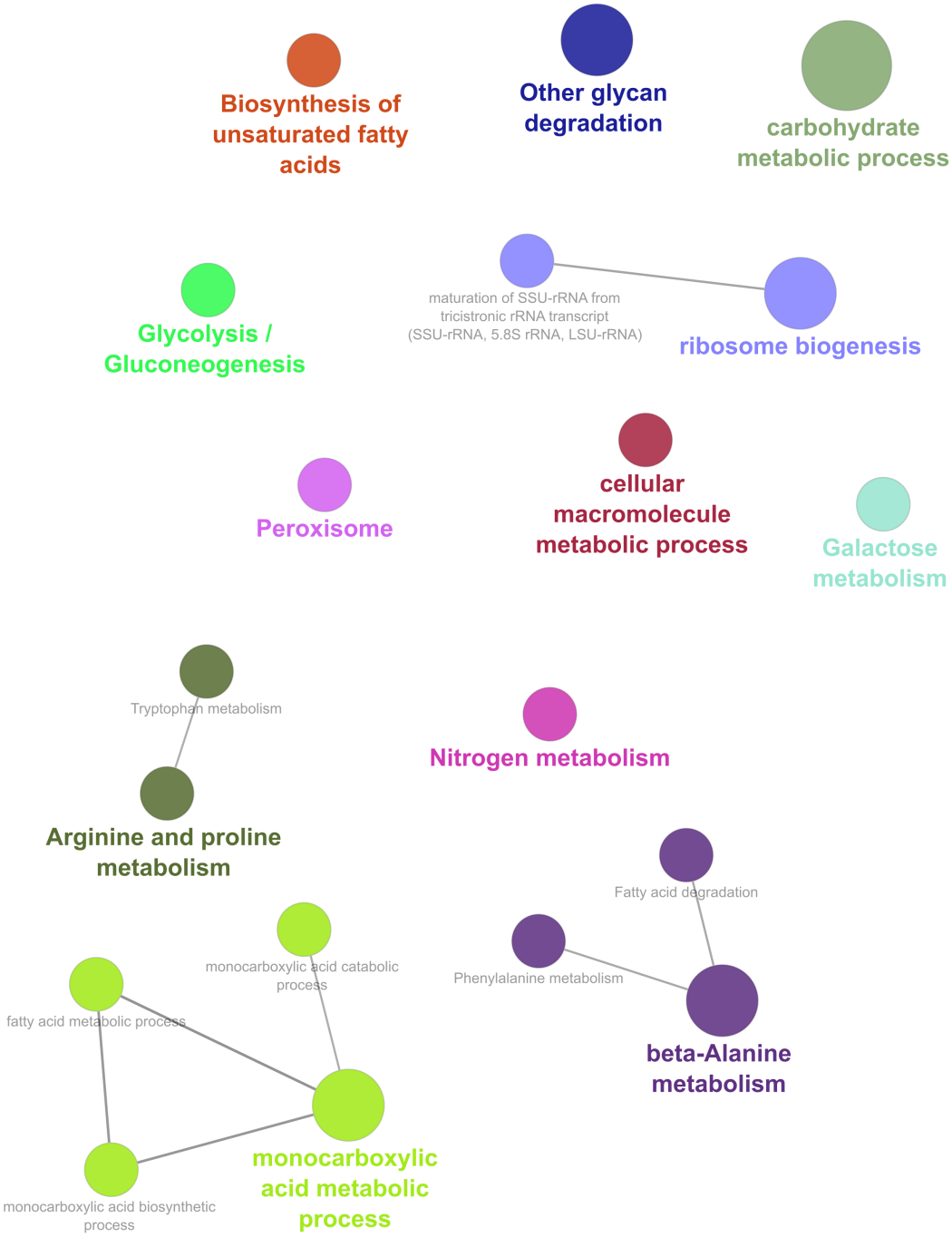

Figure 6.15: GO enrichment of proteome changes of the Hog1 deletion mutant (reversibly adapted) during the time course of 24h after osmotic stress.

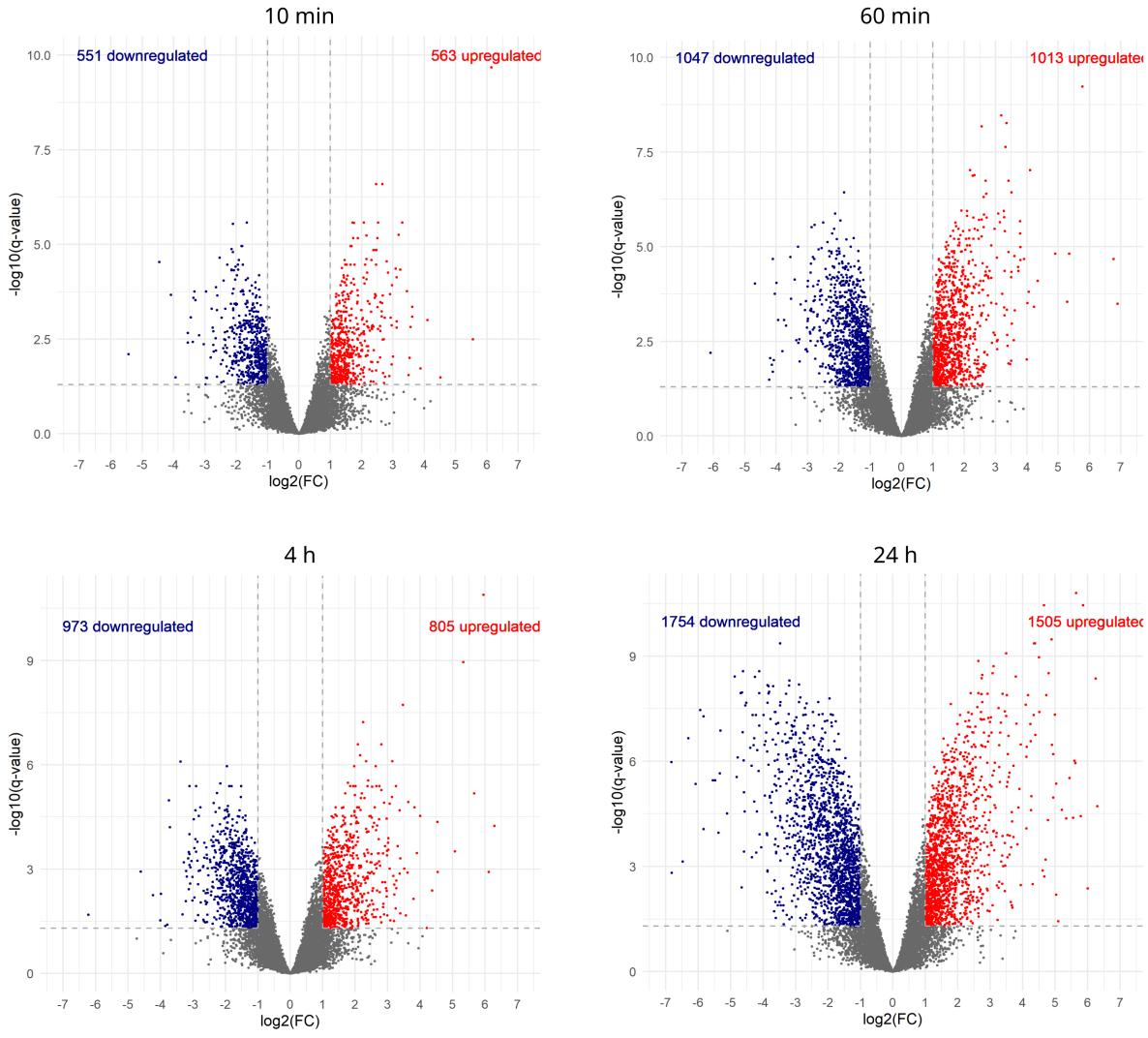

Figure 6.16: Volcano plots of temporal changes in phosphoprotein of the Hog1 deletion mutant (reversibly adapted) during the time course of 24h after osmotic stress.

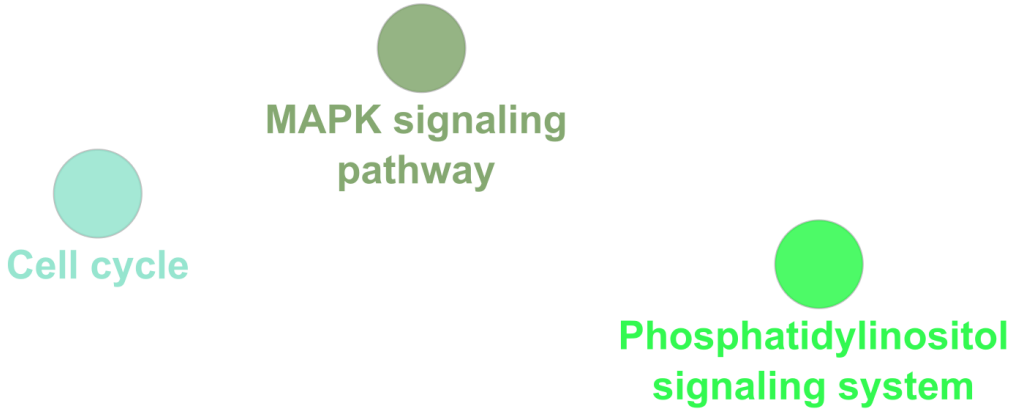

Figure 6.17: GO enrichment of phosphoprotein changes of the Hog1 deletion mutant (reversibly adapted) during the time course of 24h after osmotic stress.

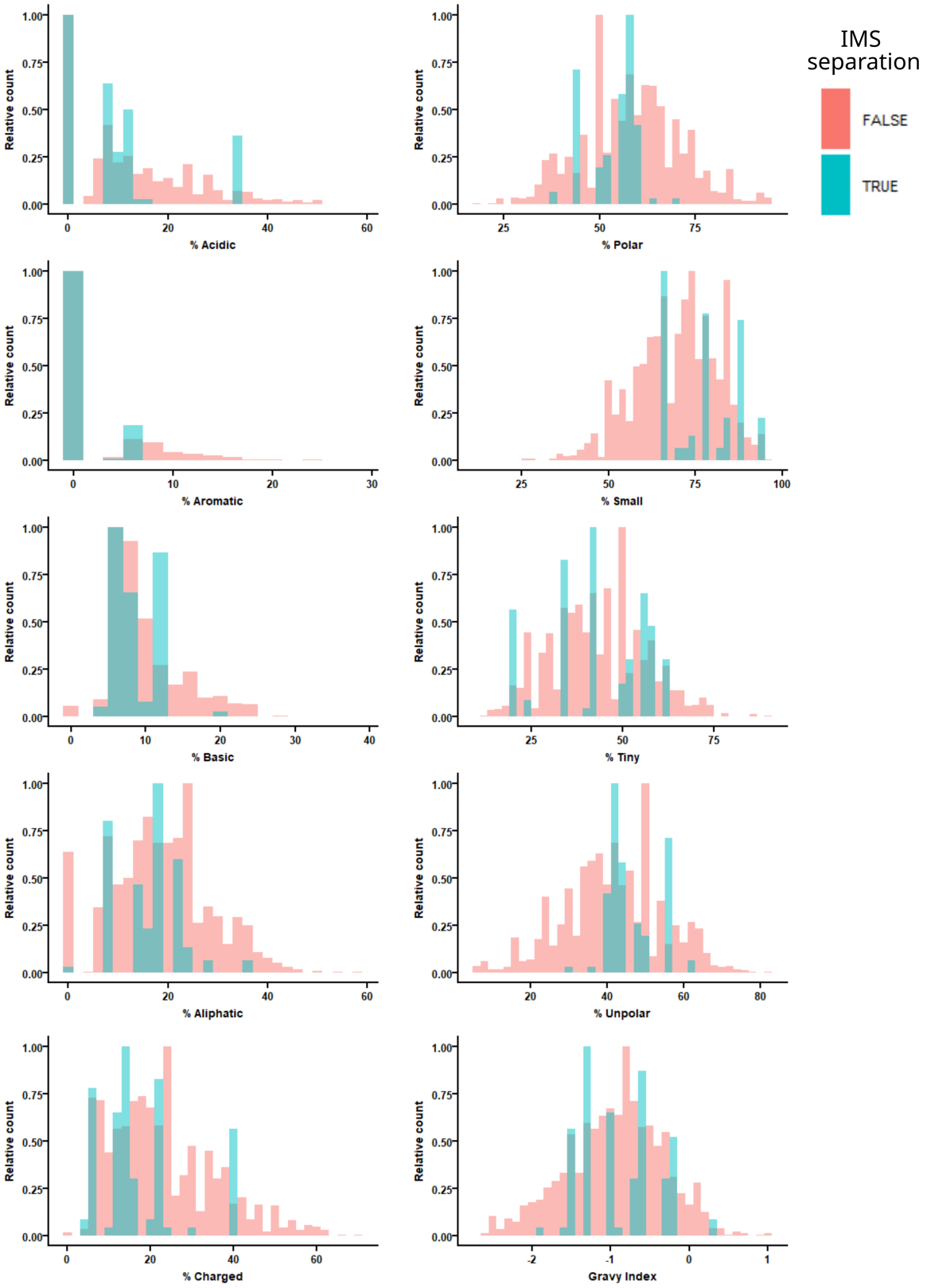

Figure 6.18: Normalized histograms of co-eluting isobaric phosphopeptide isomers enriched from 24 µg human osteosarcoma cells with co-elution (red) and separated (blue) in ion mobility.

7 Curriculum vitae

Education

08/2015 - 03/2017	**Master in Bio- and Pharmaceutical Analysis**; University of Applied Sciences Fresenius, Idstein (Germany); Thesis "Implementation of an hepatocyte based assay for the accurate prediction of metabolic elimination of small molecules in Human beings"
03/2011 - 08/2014	**Bachelor in Bioengineering** (formerly german engineering diploma); University of Applied Sciences Bingen, Bingen (Germany); Thesis "Robustness-Testing of a HPLC purity assay of gentamicin sulphate using Design of Experiments" ; Awarded as Merck Talent
08/2002 - 02/2010	**Max-Slevogt High Shool**; Landau in der Pfalz (Germany); Key subjects: Chemistry / Physics / English

Professional Experience and Trainings

11/2018 - 06/2022	**PhD Candidate in Proteomics**; Core Facility Mass spectrometry - Institute for Immunology - University Medical Center of the Johannes-Gutenberg University Mainz; Tasks: Proteome and Phosphoproteome analysis using LC-MS/MS, SDS-PAGE, Managing Core Facility Projects and Communication with external collaborators, Targeted- and bottom up proteomics with synthetic peptides, Instrument responsibility (operation and troubleshooting) for Orbitrap Exploris 480 incl. FAIMS unit, Statistical analysis with R and Python, Wet lab automation with Biomek i7

12/2017 - 10/2018	**Biopharm and Preformulation Automation Development Scientist** (External Consultant for Aerotek); Janssen Pharmaceutica -Beerse (Belgium); Tasks: In-vitro biopharm assay design (e.g. solubility in biorelevant media); Lab automation with Hamilton Star Plus
10/2014 - 11/2017	**Merck KGaA - DMPK eADME Department of Merck Serono**; Darmstadt (Germany); Tasks: Routine analysis of CYP-Inhibition (luminescence), Caco2-Permeability, Serum-Proteinbinding, Intrinsic Hepatic Clearance Prediction (MS/MS), Assay development and automation
08/2010 - 09/2014	**Merck KGaA - different Departments (paralell to studies)**; Darmstadt (Germany); Tasks: Apprenticeship, excerpt of the included work, each 3-6 months: GMP analysis of raw materials and pharmaceutical products with HPLC, TLC, IC, GC and wet chemistry. Method development and validation with DoE, Organic synthesis of small molecules in drug discovery, Suzuki-Coupling underdifferent conditions, various state-of-the-art organic reactions, Round robin tests of various UV/Vis spectrometers for the application as standard spectrometers for Merck, especially for automated solvent QC (Projectwork), Development of wet chemistry test kits (Spectroquant), Development of UV-stable reactive mesogenes in Chilworth, Southampton, UK (Leonardo-da-Vinci-Exchange), Training in Immunoassays and PCR techniques
04/2010 - 05/2010	**Institute of Grapevine Breeding Geilweilerhof**; Siebeldingen (Germany); Tasks: Internship. Sample preparation of grapevine samples and GC analysis

Method and Software Skills

Chromatography data systems	EZChrom / Empower / Analyst / MagIC Net / XCalibur / MassLynx / timsControl
Software	GraphPad Prism / Design Expert / R / Python / MS Office
MS/MS Analysis	with AB Sciex QTRAP 5500, API 4000 / AB Sciex 6500 + / waters Synapt G2-S / Bruker timsTOF Pro / Thermo Exploris 480
Laboratory Automation	with Hamilton Microlab Starline / Starline Plus / Tecan Evo / Biomek i7
Automation Software	Hamilton Venus 3 / Biomek i7 Method Editor
nUPLC/UPLC/HPLC	with Waters Aquitiy and I-Class, Agilent 1200 and 1290, Merck Hitachi LaChrom / Elite / Äkta Pure 20
IC Analysis	with Metrohm 761 Systems
Spectroscopic Analysis	with various equipment (Varian, Perkin Elmer,Bruker, Shimadzu, Thermo Fisher, Analytik Jena, Jasco, Tecan, BioTek)
(Phospho)proteomics	with SP3 / FASP/ TiO_2 / Zr- IMAC / ERLIC / DDA and DIA

Languages

German	native
English	fluent
Polish	good command
Dutch	good working knowlegde
Russian	very basic knowlegde
Latin	-

8 List of publications

1. Quantitative Proteome and Phosphoproteome Profiling in Magnaporthe oryzae; Michna T, Tenzer S.;Methods Mol Biol. 2021;2356:109-119. doi: 10.1007/978-1-0716-1613-09.; Contribution: Wrote the manuscript

2. Data-Independent Acquisition (DIA) Is Superior for High Precision Phospho-Peptide Quantification in Magnaporthe oryzae; Bersching K, Michna T, Tenzer S, Jacob S.; J Fungi (Basel). 2022 Dec 31;9(1):63. doi: 10.3390/jof9010063.; Contribution: Experiment planning, data acquisition, data analysis and wrote the manuscript

3. Evidence of a New MoYpd1p Phosphotransferase Isoform in the Multistep Phosphorelay System of Magnaporthe oryzae; Bühring S, Yemelin A, Michna T, Tenzer S, Jacob S.; J Fungi (Basel). 2021 May 15;7(5):389. doi: 10.3390/jof7050389.; Contribution: Experiment planning, data acquisition, data analysis and wrote the manuscript

4. Targeting myeloid cell coagulation signaling blocks MAP kinase/TGF-b1-driven fibrotic remodeling in ischemic heart failure; Garlapati V, Molitor M, Michna T, Harms GS, Finger S, Jung R, Lagrange J, Efentakis P, Wild J, Knorr M, Karbach S, Wild S, Vujacic-Mirski K, Münzel T, Daiber A, Brandt M, Gori T, Milting H, Tenzer S, Ruf W, Wenzel P.; J Clin Invest. 2023 Feb 15;133(4):e156436. doi: 10.1172/JCI156436.; Contribution: data acquisition, data analysis

5. Limited proteolysis by acrosin affects sperm-binding and mechanical resilience of the mouse zona pellucida; Kuske M, Floehr J, Yiallouros I, Michna T, Jahnen-Dechent W, Tenzer S, Stöcker W, Körschgen H.; Mol Hum Reprod. 2021 Mar 24;27(4):gaab022. doi: 10.1093/molehr/gaab022.; Contribution: data analysis

6. Determination of low intrinsic clearance in vitro: the benefit of a novel internal standard in human hepatocyte incubations; Zanelli U, Michna T, Petersson C.; Xenobiotica. 2019 Apr;49(4):381-387. doi: 10.1080/00498254.2018.1451010. Epub 2018 Mar 26.; Con-

tribution: Experiment planning, data acquisition, data analysis and wrote the manuscript

7. AKT activity orchestrates marginal zone B cell development in mice and humans; Cox EM, El-Behi M, Ries S, Vogt JF, Kohlhaas V, Michna T, Manfroi B, Al-Maarri M, Wanke F, Tirosh B, Pondarre C, Lezeau H, Yogev N, Mittenzwei R, Descatoire M, Weller S, Weill JC, Reynaud CA, Boudinot P, Jouneau L, Tenzer S, Distler U, Rensing-Ehl A, König C, Staniek J, Rizzi M, Magérus A, Rieux-Laucat F, Wunderlich FT, Hövelmeyer N, Fillatreau S.; Cell Rep. 2023 Apr 14;42(4):112378. doi: 10.1016/j.celrep.2023.112378. Online ahead of print.; Contribution: data acquisition, data analysis

9 Statutory declaration

I hereby certify in accordance with § 12, (2) of the doctoral regulations dated 01.04.2018:

- I have prepared the work now submitted as a dissertation myself and have used only the sources and aids indicated.

- I have not or have not yet submitted the work now presented as a dissertation to any other German or foreign university or comparable institution for the purpose of obtaining an academic degree.

- I have not yet unsuccessfully completed a PhD or comparable graduation procedure in the doctoral subject.

- I have not yet successfully completed a doctoral PhD,- or a comparable graduation procedure in the doctoral subject.

- I have not received any paid assistance from third parties, in particular from a doctoral advisor or a doctoral mediator, for the preparation of the thesis submitted.

Place, date and signature