Identification and Validation of Regulatory Genetic Variation in Yeasts

Inaugural-Dissertation

zur

Erlangung des Doktorgrades

der Mathematisch-Naturwissenschaftlichen Fakultät

der Universität zu Köln

vorgelegt von

Jan Großbach

geboren in Köln

Köln, 2020

Berichterstatter:

Prof. Dr. Andreas Beyer
Prof. Dr. Achim Tresch

Tag der letzten mündlichen Prüfung: 12.6.2019

Zusammenfassung

Genomische Unterschiede beeinflussen das molekulare Netzwerk, das alle zellulären Aufgaben ausführt. Da diese Aufgaben in erster Linie von Proteinen erledigt werden, ist es von zentraler Bedeutung die Effekte von genomischen Unterschieden auf die Häufigkeiten und Aktivitäten von Proteinen zu verstehen.

Wir haben in zwei unabhängigen Studien in der Spalthefe *Schizosaccharomyces pombe* und der Bäckerhefe *Saccharomyces cerevisiae* die Effekte von genetischen Veränderungen auf das molekulare Netzwerk erforscht, indem wir Genloci mit Einflüssen auf quantitative Merkmale (QTL) identifiziert haben. Um das molekulare Netzwerk in mehr Zuständen analysieren zu können, haben wir *S. pombe* Kulturen oxidativem Stress ausgesetzt. Die Effekte der meisten QTL haben sich zwischen beiden experimentellen Bedingungen unterschieden und wir haben genetische Polymorphismen identifiziert (QTL-Hotspots), die die Expression von Genen beeinflussen, welche in der Antwort auf zellulären Stress involviert sind. Für einen dieser Hotspots haben wir die ursächliche Mutation im *pka1*-Gen identifiziert und validiert. *pka1* ist ein Bestandteil des RAS-Signalwegs, der an der Antwort auf Umweltstress in *S. pombe* beteiligt ist.

In der zweiten Studie haben wir Daten über die Phosphorylierung von hunderten von Proteinen mit Daten über die Häufigkeiten der Proteine und ihrer Transkripte kombiniert. Das Phosphoproteom wurde in starkem Maße von genetischen Unterschieden beeinflusst und es konnte eine hohen Zahl an Polymorphismen mit einem Effekt auf Phosphorylierung identifiziert werden. QTL-Hotspots mit Auswirkungen auf anderen molekular Ebenen, sowie Mutationen in der Nähe der betroffenen Proteine haben auch das Phosphoproteom beeinflusst. Der zusätzliche Informationswert der Phospho-Daten wurde zudem bei der Analyse von betroffenen Signalwegen deutlich. Die umfassende Charakterisierung des Proteoms in beiden Studien hat es uns ermöglicht, die Beziehung von Effekten auf der Transkript- und Proteinebene zu erforschen. Diese Beziehung war komplex und unterschied sich deutlich zwischen verschiedenen Genen mit unterschiedlichen Funktionen.

Diese Arbeit trägt zum tieferen Verständnis von molekularen Netzwerken bei, indem

die Effekte von genetischen Polymorphismen in den untersuchten Modellorganismen auf die Stressantwort, das Proteom und das Phosphoproteom bestimmt werden. Es wird zudem gezeigt, wie Veränderungen im molekularen Netzwerk durch verschiedene Ebenen hinweg analysiert werden können.

Abstract

Genomic variation impacts on the molecular network that performs all cellular functions. As these functions are largely carried out by proteins, it is crucial to understand how genomic variation contributes to differences in protein abundances and activities.

We investigated the effects of genetic variation on the molecular network in two independent studies in the fission yeast *Schizosaccharomyces pombe* and the budding yeast *Saccharomyces cerevisiae* through the mapping of quantitative trait loci (QTL). To be able to observe the molecular network in more states, we exposed fission yeast samples to oxidative stress. The effects of most QTL differed between these two experimental conditions. We identified QTL-hotspots that affected the expression of large numbers of stress response genes. For one of these hotspots we identified and validated *pka1* as the causal gene: a missense mutation in *pka1* caused a reduction in RAS signaling, which mediates part of the stress response in fission yeast.

In the second study, we integrated quantitative phosphorylation traits for hundreds of proteins with matched transcriptomic and proteomic data. We found the phosphoproteome to be controlled by genetic variation and identified numerous associations between genetic variants and phosphorylation traits. Hotspots for other molecular layers and local variation both impacted on the phosphorylation of proteins. The additional information of phsophorylation was further demonstrated by the analysis of affected signaling pathways. The comprehensive proteomic data in both projects allowed us to investigate the effects of changes in transcript levels on protein levels. We found the relationship between transcript and protein levels to be complex and variable across functional groups of genes.

This work contributes to the understanding of molecular networks by identifying the effects of genetic variation on the stress response, proteome, and phosphoproteome in the yeast models, studied here. We demonstrate how perturbations of this network can be tracked through multiple molecular layers.

Acknowledgements

This thesis represents several years of my life and a large portion of what has formed me during this time. The fact that these years have been filled with positive emotions and experiences is due to the people I share this great environment with.

I want to thank my supervisor Andreas Beyer not only for building this environment but for also supporting me in tremendous ways in my work and development. I have truly enjoyed my projects so far and I'm grateful that I got to be a part of them and the Beyer-Lab. I'm also especially thankful to Mathieu Clément-Ziza who let me be a part of his projects and taught me much of what I know today. I want to thank Corinna Schmalohr for our many discussions and her support that has helped me greatly. My thanks also go to Achim Tresch for giving great advice on a variety of projects and for his continuous support during my time as a student.

As a computational biologists, I'm truly dependent on my collaborators and I'm grateful that they shared the products of their work with me. My special thanks go to Stephan Kamrad and María Rodríguez-López.

I'm thankful to all current and past members of the lab for making this a great place.

Contents

Chapter 1

General introduction

Understanding how genetic variation expresses itself as changes in measurable phenotypes is one of the central questions of modern biology.

Many of the traits that are medically or economically relevant are quantitative, i.e. they follow a continuous distribution. The majority of clinically relevant traits is estimated to be quantitative (van der Sijde et al., 2014). Yield of crops and growth rates of animals are quantitative traits that are of economical relevance (King et al., 2015; Plastow et al., 2005).

Most phenotypes are affected by a combination of environmental and genetic influences. It has long been debated to which degree each of these factors contribute to these traits. The overall dependence of a quantitative trait on genetic variation within a population can be quantified and is referred to as the broad-sense heritability (reviewed in Visscher et al. (2008)). This heritability can be determined empirically by comparing the variance in the trait values of closely related individuals to the variance in the trait values of distantly related individuals, while controlling for environmental differences. Accurate estimates of heritability in human populations are challenging since environmental influences tend to be more complicated and harder to control than in populations of model organisms (reviewed in Visscher et al. (2008)). Aside from monozygotic twins, there are no humans with identical genomes available to be studied. In both humans and model organisms, the broad-sense heritability of many traits was found to be sizeable in the vast majority of instances (Göring et al.,

2007; Foss et al., 2011; Bloom et al., 2013; Polderman et al., 2015). The heritability for the same trait can differ between populations as it can be influenced by environmental and genetic factors that differ between the populations (Dandine-Roulland et al., 2016).

The inheritance of many quantitative traits was studied in humans through genome-wide association studies (GWASs, Visscher et al. (2017)). GWASs have identified thousands of cases in which genetic variation contributes to human disease (van der Sijde et al., 2014). Accordingly GWAS can be used to explain a portion of the disease risk between individuals (Purcell et al., 2009). A gap between estimates of broad-sense heritability and the variance that can be explained by such models still exists and is commonly referred to as 'missing heritability' (Manolio et al., 2009).

While GWASs often detect numerous associations between genomic regions and a trait of interest, the causal variant and the mechanism through which it affects the trait often remain hidden. Reasons for this are (i) the density of polymorphisms in the human genome, (ii) the linkage disequilibrium (LD) between polymorphisms in a region of interest, (iii) the incomplete annotation of regulatory elements and gene functions, and (iv) the incomplete knowledge about the etiology of many common diseases (van der Sijde et al., 2014). Different approaches to identify causal variants from a set of polymorphisms, by predicting their effects on protein function, have been proposed (Rentzsch et al., 2019; Kumar et al., 2009). An alternative approach, promising to elucidate the mechanisms that are underlying causal genetic variants, is systems genetics (Civelek and Lusis, 2014). Systems genetics uses knowledge of the molecular network to link known molecular changes in disease states with effects of genetic variants on these molecular traits (van der Sijde et al., 2014). If, for instance, a disease state is associated with a change in the expression of a specific group of genes and a genetic locus with an effect on disease risk contains such a gene, that gene would be a candidate causal gene. This molecular network is made up of multiple layers, of which some of the most important are the genome, transcriptome and proteome. While the genome and transcriptome are essential to this network as well, the proteome is largely responsible to perform any functions necessary for the survival of the cell, such as most structural or enzymatic functions. Consequently

protein levels are under stricter conservation (Khan et al., 2013).

Genetic polymorphisms can affect protein levels and activities in a large number of ways. Variants within regulatory regions such as enhancer, repressor, or promoter regions have the potential to modulate the expression of a transcript. For instance, a change in the binding behavior of transcription factors or the RNA-polymerase can lead to a change in transcription frequency or strength (reviewed in Albert and Kruglyak (2015)). Genetic polymorphisms within 3' and 5' untranslated regions were shown to affect both mRNA maturation and stability. In addition, they can affect the translation rates for the transcripts they are part of (reviewed in Pesole et al. (2001)). All of these changes to transcripts can result in changes in the levels of the respective proteins. Mutations within the coding region of a gene can lead to both silent polymorphisms that do not affect the amino acid sequence of the protein product, and non-silent polymorphisms. Non-silent mutations include missense mutations, nonsense mutations, shifts of the reading frame, and insertions and deletions. All of these kinds of polymorphisms have the potential to directly affect the function of a protein. Polymorphisms within a protein binding site or an active site of an enzyme, for instance, have been reported to drastically affect the activity of proteins (Okerberg et al., 2016). Mutations at phosphorylation sites can have a drastic effect on protein functions (Gupta et al., 2011). Changes in protein binding or folding can result in increased degradation or aggregation of the protein, (MacDonald et al., 1993). Even silent mutations can affect the folding of proteins and translation rates through the availability of the respective tRNAs (Spencer and Barral, 2012).

While data for many macroscopic phenotypes is often readily available, for instance through medical records, the costs associated with collecting comprehensive transcriptomic and proteomic data for a GWAS-cohort is often prohibitive. In order to reduce the necessary sample size for significant findings on the molecular network, studies investigating the genetic influences on molecular networks are often performed in special populations of model organisms (Brem et al., 2002; Holdt et al., 2013; Clement-Ziza et al., 2014; Snoek et al., 2017). These studies aim to detect quantitative trait loci (QTL), i.e. genomic regions with a significant effect on the quantitative trait of interest. QTL-studies can be considered as a subset of GWAS

and are characterized by several features. Populations in QTL-studies are often derived from a small number of founder-genotypes and have therefore relatively few genetic polymorphisms. Further, the minor allele frequencies (MAFs) in such populations are, on average, higher than that in natural populations (Clement-Ziza et al., 2014; Almasy and Blangero, 2009). Environmental influences can be controlled to a large degree. All of these factors improve the statistical power of QTL-studies compared to GWAS in humans. As fewer samples are required, QTL-studies can be used to systematically investigate molecular traits such as gene expression (eQTL), protein levels (pQTL) or the levels of metabolites (mQTL) (Brem et al., 2002; Foss et al., 2007; Albert et al., 2014a; Nagana Gowda and Djukovic, 2014).

As it is technically easier to quantify the transcriptome than the proteome, more eQTL studies have been performed than pQTL studies and the former have generally found a larger number of significant associations (Brem et al., 2002; Foss et al., 2011; Clement-Ziza et al., 2014; Albert et al., 2014a, 2018). While it is established that the absolute levels of transcripts and proteins are highly correlated across different genes within an organism, the dependence of protein levels on their transcripts across samples is less complete (Vogel and Marcotte, 2012). Recent studies suggest that the control of protein level variation between samples through transcript levels is limited and variable between genes (reviewed in Liu et al. (2016)). The biological relevance of protein abundances and their partial independence from transcript levels necessitate the direct quantification of the proteome to gain a full understanding of the molecular network.

Chapter 2

Genetic control of the response to oxidative stress in *Schizosaccharomyces pombe*

2.1 Introduction

Systems genetics aims to understand the overall genetic network which expresses all cellular traits (Civelek and Lusis, 2014). Environmental influences can affect which parts of this network are active. Most QTL studies consider the effects of genetic variation on quantitative traits under standard lab conditions (Brem et al., 2002; Clement-Ziza et al., 2014; Albert et al., 2018). While a large proportion of traits show considerable heritable variation in these conditions, many polymorphisms were shown to have different effects in different growth conditions (Smith and Kruglyak, 2008; Clement-Ziza et al., 2014; Ackermann et al., 2013). As their effects depend on the experimental condition, these QTL are referred to as 'conditional QTL' (Ackermann et al., 2013). Conditional QTL can elucidate aspects of the genetic network that would remain hidden from a study restricted to standard laboratory conditions. Environmental perturbations along with genetic perturbations increase the states in which the genetic network can be observed and therefore can be helpful to improve

14

our understanding of the network.

We chose to investigate the effects of oxidative stress on the expression of molecular traits. Aside from its general value for perturbing the genetic network, understanding how the cell responds to oxidative stress is of importance for medical research. Oxidative stress was shown to have some role in aging and pathophysiology (Bhattacharyya et al., 2014; Liguori et al., 2018). Several aspects of human immune function are in part dependent on the generation of oxidative stress. For instance the blood forms an oxidative environment which reduces the spread of metastases and phagocytes generate reactive oxygen species to attack pathogens (Piskounova et al., 2015; Babior, 2000). Further, oxidative stress has important signaling functions and low levels of oxidative stress are beneficial to many organisms (reviewed in Yan (2014)). We investigated the response to oxidative stress in the fission yeast *Schizosaccharomyces pombe*. Fission yeast differs from the budding yeast *Saccharomyces cerevisiae*, the most widely used yeast model organism for QTL mapping, in several characteristics that make it interesting as a model organism for QTL studies. While only a small proportion of genes in budding yeast contain introns and almost no transcripts are subject to alternative splicing, fission yeast is more similar to higher eukaryotes in its splicing behavior (Wood et al., 2002; Käufer and Potashkin, 2000). It is often used to study processes related to growth and cell division and it resembles higher eukaryotes such as humans in its chromatin structure, telomeres and lower number of duplicated genes than budding yeast (Hayles and Nurse, 2018).

The response to oxidative stress in fission yeast has been studied in the reference strain (JB22) before (Chen, 2003; Chen et al., 2008; Marguerat et al., 2014). Several key players that mediate the stress response have been identified (Toone et al., 1998; Rodriguez-Gabriel et al., 2006; Rodríguez-Gabriel et al., 2003; Wilkinson et al., 1996) and the response to oxidative stress overlaps with the core environmental stress response (CESR, Chen (2003)).

To understand the effects of natural genetic variation in fission yeast we measured the transcriptome, proteome and growth traits of a large number of strains from a fission yeast cross. The conditional effects of genetic variants on molecular traits can only be adequately analyzed if they were correctly identified. A common pitfall

in the comparison of results from different experiments is to compute the overlap in significant results and consider all other hits as being specific to a condition (reviewed in Ackermann et al. (2013)). In the case of QTL being mapped in different datasets, this approach would misclassify QTL that have effects in both conditions, but only reach significance in a subset, as being conditional. This is especially relevant if the statistical power in both conditions is not the same. Further, differences in the effects of QTL between the conditions might be overlooked if only significance is considered. Other approaches consider the change in trait levels for the same strains as a quantitative trait for QTL-mapping (Smith and Kruglyak, 2008; Ackermann et al., 2013). While this would lead to the detection of QTL that affect the change in a trait across conditions, further analysis would still be required to identify the condition in which the QTL affects the trait level and to identify QTL that have the same effect across conditions. We applied an approach that was previously proposed by Ackermann et al. and consists of QTL mapping which is followed by *post hoc* tests that classify a QTL based on its effects in the different conditions and determine the conditions in which it has an effect (Ackermann et al., 2013). We characterized a panel of fission yeast strains on the transcriptomic and proteomic layer in conditions with and without added oxidative stress. Further, we measured growth efficiency for these strains. We used these data to investigate the effects of genetic variation on the expression of these molecular and fitness traits. In particular we were interested in how genetic variation affects the response to oxidative stress in this cross.

2.2 Methods

2.2.1 Generation of a fission yeast cross for QTL studies

To generate a fission yeast cross in which the response to oxidative stress could be studied, we chose parental strains that exhibited differences in their resistance to oxidative stress in preliminary experiments. The three parental strains were JB50, a strain closely related to the reference strain JB22, JB759, a strain with increased sensitivity to oxidative stress compared to JB50, and JB760, a strain that is more

resistant to oxidative stress than JB50.

To obtain diploid hybrids between each pair of parental strains, we deleted the ade6-locus and replaced it with resistance genes for kanamycin (KAN, in JB50) nurseothricin (NAT, in JB759) or hygromycin B (HB, in JB760). We used plates with both fungicides for a pair of parental strains to select for diploid hybrids, carrying both resistance genes. F1 segregants were obtained from these diploid hybrids for each pairwise cross, and were then further crossed to obtain an F2 population. The cross included 150 strains in total.

2.2.2 RNA seq

Samples for the transcriptomic and proteomic analyses were prepared separately from different cultures grown from differing number of overlapping strains of the QTL-cross. The transcriptome of 286 samples, corresponding to 130 strains, was quantified. These strains included all of the strains whose proteome was quantified. Samples for the transcriptomic analysis were prepared by inocculating 50 ml YES medium at 32°. Samples were grown to an $OD_5 95$ of 0.4-0.5. Samples were either harvested before or after the exposure to 0.5 mM H2O2 for one hour. RNA was isolated by hot phenol extraction (Lyne et al., 2003). Poly(dT) enrichment was achieved through three rounds of Sera-Mag magnetic bead purification. The purity of RNA was determined with a Bioanalyzer. A total of 500 ng RNA per sample was used for a library preparation. cDNA was produced from the mRNA, ligated with SOLiD-adaptors and amplified in 16 PCR cycles. The libraries were sequenced on an ABI SOLiD V4.0 System to produce stranded and single-end reads with a length of 50 bp (Applied Biosystems, LifeTechnologies, United States).

Previous work showed that accounting for coding variants improves the quantification of transcript levels compared to using a reference genome that does not reflect all polymorphisms (Clement-Ziza et al., 2014). This has special relevance for the detection of local eQTL, as polymorphisms can reduce the mapping efficiency when the reference genome is used, thus resulting in false positive eQTL.

We mapped reads against the *Schizosaccharomyces pombe* reference genome using

Bowtie with the following parameters: `-C -n 3 -e 100 -best` (v.0.12.7, Langmead et al. (2009)). Read group information was added, and BAM files were sorted using Picard utilities (`http://broadinstitute.github.io/picard`). Further processing of the RNA-seq data was performed with the GATK pipeline (v3.4-46-gbc02625), according to best practice guidelines (Van der Auwera et al., 2013). Variants at the polymorphic sites were then called with UnifiedGenotyper. If the GATK score for a given site was below 20, the position was considered as unknown. We excluded polymorphic sites for which (i) the samples of the parental strains could not be called, (ii) or more than 50% of the segregants could not be called, or (iii) the minor allele frequency was less than 10% in our cross. We also excluded genetic markers if they were called differently than both closest neighbors (i.e. the nearest upstream marker and the nearest downstream marker) and those markers were within 50 kb. When possible, missing genotypes were inferred through the neighboring markers if they had a identical segregation patterns and were within 50 kb. In these cases the marker was assigned the same allele as the two flanking markers. VCFs were converted to strain-specific FASTA-files using the GenomeGenerator tool (Clement-Ziza et al., 2014).

2.2.3   RNA quantification

We mapped reads with Bowtie (v.0.12.7, Langmead et al. (2009)) against the strain-specific genomes we generated based on the RNA-seq data using the following options: `-C --best -m 1`. Genes to which no reads were mapped in at least 50% of samples were dismissed. To correct for differences in library size, we compute a correction factor for each sample. We determined the 20% and 80% quantiles for each sample i and computed the median M_i for all counts between these quantiles. The counts for gene j and sample i were then multiplied with the ratio of the median of the counts for this subset of genes and the mean of these medians across all samples to correct for different library sizes.

$$c'_{i,j} = c_{i,j} \cdot \frac{M_i}{\overline{M}}$$

Afterwards counts were divided by the length of the coding region of the corresponding gene to correct for gene length.

$$c''_{i,j} = \frac{c'_{i,j}}{l_j}$$

2.2.4 Proteomics

Culture preparation

Samples for the proteomics analysis were prepared by inocculating 50 ml YES medium at 32°. Samples were grown to an OD_{595} of 0.4-0.5. Samples were either harvested before or after the exposure to 0.5 mM H2O2 for one hour.

Protein extraction

Cells were harvested by centrifugation at 2,000xg, and resuspended in 200 μl of lysis buffer and 100 μl of glass beads. The cells were then lysed by vortexing three times for 1 minute and sonicated three times for 30 seconds. A small aliquot of the supernatant was taken to determine the protein concentration using a BCA assay (Thermo Fisher Scientific, United States) and the protein concentration was adjusted to 5 mg/ml using additional lysis buffer. Proteins obtained from the different samples were reduced with 5 mM TCEP for 60 min at 37 °C and alkylated with 10 mM iodoacetamide for 30 min in the dark at 25 °C. After quenching the reaction with 12 mM N-acetyl-cysteine the samples were diluted with 100 mM ammoniumbicarbonate buffer to a final urea concentration of 1.5 M. Proteins were digested by incubation with sequencing-grade modified trypsin (1/50, w/w; Promega, Madison, Wisconsin, United States) over night at 37°C. Then, the samples were acidified with 2 M HCl to a final concentration of 50 mM, incubated for 15 min at 37 °C and the cleaved detergent removed by centrifugation at 10,000xg for 5 min. Subsequently, all peptides were desalted on C18 reversed-phase spin columns according to the manufacturer's instructions (Macrospin, Harvard Apparatus, United States), dried under vacuum and stored at -80°C until further use. The hybrid Orbitrap-Velos mass spectrometer

was interfaced to a nanoelectrospray ion source coupled online to an Easy-nLC system (Thermo Fisher Scientific, United States). 1 μg of peptides were separated on a RP-LC column (75 μm x 20 cm) packed in-house with C18 resin (Magic C18 AQ 3 μm; Michrom BioResources, United States) using a linear gradient from 95% solvent A (98% water, 2% acetonitrile, 0.15% formic acid) and 5% solvent B (98% acetonitrile, 2% water, 0.15% formic acid) to 30 % solvent B over 120 minutes at a flow rate of 0.3 μl/min. Each survey scan acquired in the Orbitrap at 60,000 FWHM was followed by 10 MS/MS scans of the most intense precursor ions in the linear ion trap with enabled dynamic exclusion for 60 seconds. Charge state screening was employed to select for ions with at least two charges and rejecting ions with undetermined charge state. The normalized collision energy was set to 32% and one microscan was acquired for each spectrum. The acquired raw files were converted to centroid mzXML format using ReAdW (`http://tools.proteomecenter.org/wiki/index.php?title=Software:ReAdW`), MS/MS spectra were searched using the SORCERER-SEQUEST v4.0.3 algorithm against a decoy database (consisting of forward and reverse protein sequences) of the predicted proteome from *Schizosaccharomyces pombe* (`ftp://ftp.sanger.ac.uk/pub/yeast/pombe/Protein_data/pompep`). The database consists of 4974 S. pombe proteins as well as known contaminants such as porcine trypsin, human keratins and high abundant bovine serum proteins (Non-Redundant Protein Database, National Cancer Institute Advanced Biomedical Computing Center, 2004, `ftp://ftp.ncifcrf.gov/pub/nonredundant`) resulting in a total of 10 584 protein sequences. The search criteria were set as follows: full tryptic specificity was required (cleavage after lysine or arginine residues, unless followed by proline); 2 missed cleavages were allowed; carbamidomethylation (C) was set as fixed modification; oxidation (M) was applied as variable modifications; mass tolerance of 10 ppm (precursor) and 0.6 Da (fragments). The database search results were further processed using the PeptideProphet (Keller et al., 2002) and ProteinProphet (Nesvizhskii et al., 2003) program and the peptide false discovery rate (FDR) was set to 1% on the peptide and 1% on the protein level and validated using the number of reverse protein sequence hits in the datasets.

The raw-files were imported into the Progenesis LC-MS software (v2.5, Nonlinear Dy-

namics Limited, United States), which was used for label-free protein quantification applying the default parameters. Only unmodified peptides having a PeptideProphet score of 0.85, corresponding to an FDR of less than 1%, were considered for quantification. The quantitative data obtained were further normalized and statistically analyzed according to (Brusniak et al., 2008) using the Spotfire Decision Site program (version 9.1.1, TIBCO) and the guides provided for analyzing large transcriptomics datasets. In brief, we set a nominal lower bound value (noise level) as the minimum measured intensity and replaced with it missing values and values below it. We then calculated fold change ratios (in log-scale) between control and perturbed samples. On the protein level, the ProteinProphet probability were employed to set the FDR to 1% based on the number of reverse protein hits. Only proteins with a 1.5-fold change in abundance and a p-value below 0.05 were considered significant.

Preprocessing

The resulting peptide counts per sample were processed further to generate suitable traits for QTL-mapping. Abundances of zero were considered as missing values. To estimate differences in the total amount of peptide counts between samples, we identified those peptides that were identified in all samples and averaged their counts per sample. The average mean $\overline{m}$ was computed and each m_i was divided by it, resulting in m_i'. All abundances in sample i were then divided by m_i' to produce the normalized abundances. We removed peptides that were detected in only three or less samples of more than one batch, indicating that their measurement was strongly dependent on batch effects. This step was necessitated by the observation that certain peptides were only detected in some batches but not others. Peptides that were not detected in the majority of samples were also removed.

MS-proteomics quantifies distinct peptides. The number of peptides that is quantified for each protein differs based on the size of the protein, its abundance, the uniqueness of its sequence, and other technical considerations. The task of estimating the abundance of the whole protein based on the abundances of multiple peptides is not trivial and multiple methods have been proposed to solve it (MacCoss et al., 2003; Teo et al., 2015). In species without extensive alternative splicing, different

21

peptides from the same protein are expected to be strongly positively correlated. Possible reasons for a lack of positive correlation between the levels of different peptides for the same protein might be general technical noise or batch effects that affect different peptides to different degrees. Approaches that combine all available levels by summing them up, or averaging them, generate protein levels that are still affected by the aforementioned technical sources of variance. A common strategy to avoid this bias is to focus on the most abundant peptides for the respective protein with the potential downside of ignoring a large number of valuable data points that can be used to reduce the noise in the resulting protein levels (as possible in Choi et al. (2014)). We chose to employ an approach that compares peptide levels for the same protein and combines them to groups if they correlate strongly. This approach allows for the generation of multiple levels for each protein if the peptide levels do not agree.

Depending on the size and abundance of a protein, the number of peptides that were measured and kept after filtering can differ. For proteins for which only one peptide level was available, the abundance of this peptide was centered by its median and considered as a peptide group.

For proteins with multiple peptides we computed the pairwise Spearman's ρ for all pairs of peptide levels.

If two peptides were never measured for the same sample, we considered their correlation to be 0. We defined the distance between the levels of two peptides as $d = 1 - \rho$ and used the `hclust` and `cuttree` functions to separate the peptides into clusters in which all peptides had a correlation of $\rho > 0.3$ (R Development Core Team, 2016). If there were less than two samples in which all peptides for a group were measured, we used the mean value of the peptides of the group per sample and mean centered these values. For all other peptide groups we centered each peptide by median and computed the mean of these values for each sample.

Batch correction was performed separately for peptide groups with and without missing values. Peptide groups with available values for all samples were corrected with comBat (Johnson et al., 2007). Because comBat does not allow for missing data, we used a different approach to correct for batch effects in the peptide groups

with missing values. Each peptide group for which comBat could not be used, was corrected by mean centering the abundances for this peptide group per batch.

2.2.5 Growth profiling

Precultures were grown in 10 ml YES medium in Falcon tubes over 36 hours. Cultures were then inoculated in flower-shaped 48 well plates in YES medium using the bioLector system (m2p-labs, Germany). The cell density of each well was adjusted so an optical density of 0.2 at a wavelength of 595 nm was reached. The optical density of each sample was determined in 3 minute intervals for at least 25 hours. After five to twelve hours, the samples were exposed to 0, 0.5 or 1 mM hydrogen peroxide. Finally light scattering values were converted to optical densities with a linear model. ΔOD was computed as the difference between the initial and final OD. Growth was measured in 28 batches. We corrected for batch effects by subtracting the mean of all measurements for the batch from each of these measurements. Afterwards we employed a step-wise procedure to remove batches that showed much more variation than the rest of the batches. First, we computed the variance of all trait values per batch. Second, if any batch had a variance that was 2.5 times as large as the average variance of all batches, we removed the batch with the highest variance. Then we repeated this step until the variance for no batch exceeded 250% of the average variance of the remaining batches. We removed four batches with this procedure. 1152 trait values remained after batch exclusion, spread across conditions and strains. We had at least one measurement per condition for 114 strains after batch removal.

2.2.6 Mapping data

To generate suitable data for QTL-mapping we averaged each trait over all replicates for a given strain in each condition separately. Here we ignored missing values, meaning that we took the average of present data. All vectors of trait values were centered to a mean of 0 and scaled to a variance of 1. We removed two strains from the pQTL mapping data as they were measured only in one condition. As we

planned to apply *post hoc* tests to determine in which condition the QTL is active, we preferred to have balanced data to avoid differences in power between the conditions. The mapping data included 130 strains and 7160 traits for the eQTL data and 59 strains and 6853 traits for the pQTL data. We collapsed identical genetic markers to predictor groups. As some markers were identical between the smaller group of strains for which the proteome was quantified but not for the larger set of strains for which the transcriptome was quantified, the final number of markers for the pQTL mapping data (497) was smaller than for the eQTL mapping data (619). To capture the effects of population structure on the traits and avoid false positive associations as a result of population structure, we compute covariates for population structure as described before (Clement-Ziza et al., 2014). The first four principal components explained more than 90% of the genotypic variance and were included in the models we generated for QTL-mapping.

2.2.7 QTL-mapping

We employed a technique based on Random Forest for QTL-mapping. This approach was developed previously and applied to molecular QTL-mapping (Michaelson et al., 2010; Ackermann et al., 2013; Clement-Ziza et al., 2014). It is well suited for QTL-mapping for several reasons: (i) it considers multiple markers in the same model and (ii) it models epistatic interactions implicitly. RF is a machine learning algorithm that builds an ensemble of decision trees predicting a categorical or quantitative dependent variable or outcome based on a set of independent variables or predictors (Breiman, 2001). Because the predictors used to split a node are chosen independently for each node relationships between predictors and subpopulations in the dataset can be efficiently modelled. In effect this allows for the modelling of epistatic interactions without prior knowledge of their existence (Michaelson et al., 2010).

We generated RF-models with a modified version of the randomForest package (RandomForestExtended, available at `http://cellnet-sb.cecad.uni-koeln.de/resources/qtl-mapping/`, Liaw and Wiener (2002)).

As the alleles for some strains and markers were not known and RF does not accept missing data, we randomly imputed alleles for the missing genotypes. We did this by sampling alleles with replacement from the strains for which the genotype at this marker was not missing. To avoid the creation of artificial biases, we repeated this imputation step 100 times per trait and grew a separate forest for each imputed genotype matrix. A forest for each imputation was grown with the following parameters: `ntree=200, importance=TRUE, keep.forest=F`. In the case of the pQTL data, many trait data were incomplete, i.e. some strains had no measurements. We removed those data points and mapped the respective traits with a smaller sample size. We combined the forests generated with each imputation of missing alleles to one forest with the `combine` function. Each of the forests that was generated in this fashion had 20,000 trees. To assess the importance of each predictor in the model that was generated we extracted the 'residual sum of squares' and 'permutation importance' (RSS and PI respectively) for each marker i and combined them as follows:

$$S_i = max(0, PI_i) \cdot max(0, RSS_i)$$

To estimate the significance of each S_{ij}, we generated null distributions of S_i by permuting the trait vectors. Importantly, the predictors representing the population structure were permuted in the same pattern. This way a possible association between a covariate representing population structure and the original trait vector would not be removed in the permuted trait vector. Removing possible associations between the predictors representing the population structure and phenotypes could result in inflated scores for the rest of the predictors. Overall this might lead to false negative associations. We permuted only those traits without missing values. For each marker in each permutation we computed a score S as described above. We generated 1,432,000 permutations for the eQTL data in each condition and 829,200 permutations for the pQTL data in each condition. The growth traits were permuted 995,000 times. We permuted the trait vectors within each cross (i.e. all strains that derive from the same pair of parental strains), to remove the association of all variants to the trait while preserving differences between the crosses. Parental strains

were permuted along with the strains that derived from them. We computed a p-value p_{ij} for a marker i and trait j by comparing S_{ij} against the distribution of S_i of the same marker in the permuted traits k in $1..l$ of the same molecular layer and condition.

$$p_{i,j} = max(\frac{1}{l}, 1 - \frac{\sum S_{i,j} > S_{i,k}}{l})$$

We showed earlier that the pooling of importance scores across the permutations of different traits does not reduce the precision or power of QTL-mapping if the traits have the same variance (Michaelson et al., 2010). We performed FDR-correction across all markers separately for each molecular layer and condition, i.e. separately for eQTL in N, eQTL in H, pQTL in N, and pQTL in H).

As neighboring markers are often highly correlated, they will often be significant for the same traits because they reflect the same underlying biological relationship between QTL and target. To improve the interpretation of these results we combined significant markers to QTL based on the following criteria. If two markers were significant for the same trait and had an absolute correlation of 0.8 or higher they were combined to the same QTL. If they were on the same chromosome and not more than 9 markers were located between them, all markers located between were also added to the QTL.

2.2.8   Identification of conditional QTL

To identify QTL that have effects that differ between experimental conditions, we employed a previously published approach (Ackermann et al., 2013). First we combined all QTL at FDR<10% from both conditions for each dataset (eQTL and pQTL). If two QTL that were detected for the same trait in different conditions overlapped, we combined them to a single one. For each QTL from this combined set we performed an ANOVA to determine if there is a significant interaction between a marker representing the QTL and the condition represented by a dummy-variable with the trait levels of the QTL-target as the dependent variable. These p-values were corrected

with the Benjamini-Hochberg procedure across all QTL for the dataset (i.e. all eQTL or all pQTL, but not both at the same time). We used a significance threshold of FDR<10% for the interaction effect. We also performed ANOVAs to determine how significant the effect of the QTL was in each separate condition. We adjusted the p-values for condition-specific ANOVAs with the Bonferroni-procedure. We used a p-value threshold of $\alpha = 0.005$ for both conditions. These tests were performed using the marker with the smallest FDR for the target trait in either condition within the QTL. Based on the corrected p-values for the interaction effect and the effect strength in each condition, we classified QTL into four classes, differing in their activity pattern: static QTL, conditional QTL active in H, conditional QTL active in N, and conditional QTL active in H and N. If the interaction effect was not significant, the QTL was classified as static. If the interaction effect was significant and the condition-specific ANOVAs were significant in at least one condition, the QTL was classified as being conditional and being active in those conditions which had significant p-values. A conditional QTL can be active in all conditions if the effects in the different conditions differ sufficiently in strength or if they differ in direction. If the interaction effect was significant but no condition had a significant p-value, the QTL was classified as static.

2.2.9   Stress response

We computed the log2 fold changes for the RNA and protein levels of the JB50-strain as a reference for the stress response. We used non-centered and non-scaled mean values of all replicates in the respective condition for JB50. The stress response r_j for a gene j was defined as

$$r_j = eN_j - eH_j$$

where eN and eH are defined as the mean expression in the JB50-strain in the normal and stress-condition, respectively. We computed the log2 fold change between the conditions for each peptide group and transcript as described above. We averaged

the log2 fold changes for peptide groups belonging to the same protein. We compared the stress response on the transcript level to that published by Chen et al. (Chen, 2003). We accessed their data at `ftp://ftp.sanger.ac.uk/pub/postgenomics/s_pombe/wtaverage.txt`. The "H2O2_60" column was log2 transformed. We used this data as we also exposed the strains to hydrogen peroxide for an hour. We directly compared those values to the stress response of the JB50 strain in our data for all genes measured in both experiments. As the stress response in our data was computed as N-H and Chen et al. computed it as H-N, we multiplied the correlation with -1.

2.2.10   Identification of local and distant QTL

In the case of QTL directly affecting a gene product, a spatial relationship between the QTL and the target can be determined. A QTL affecting the trait of a gene that is encoded in the vicinity of the QTL is often referred to as 'local' (Smith and Kruglyak, 2008). QTL that are more distant or even on different chromosomes than the trait which they affect are termed 'distant'. The same QTL can be local for one target and distant for another. For each gene we determined the midpoint between the start and the end of the coding sequence. We determined the closest marker on each side of that position. A QTL was called local if any of the markers belonging to had an absolute Pearson's correlation of 0.8 or higher with any of the markers next to the midpoint of the coding region of the target gene. To determine if we saw more local QTL than expected, we permuted the targets of all eQTL 200 times and computed the amount of local QTL.

2.2.11   Hotspot detection

Regions harboring QTL for a large number of traits are often referred to as QTL hotspots. We used a previously published approach to detect QTL hotspots (Brem et al., 2002). We divided the genome into bins with a size of 50 kb. Bins at the end of chromosomes were slightly smaller. We assigned QTL to bins based on the location of marker with the smallest p-value within the QTL. Based on the total number of

QTL for the molecular layer and condition we computed the expected number of QTL per bin. A bin containing more QTL than expected at a threshold of $\alpha = 0.05$ was considered a hotspot, using a Poisson distribution. Consecutive bins on the same chromosome were combined to one hotspot. As the number of QTL detected differed between the conditions and molecular layers, the amount of QTL in a bin that is required for hotspot was not the same for both conditions and molecular layers.

2.2.12 Functional enrichment analysis

Gene ontology (GO) enrichment analysis was performed with the topGO package (v2.30.1, Alexa (2007)). The annotations were downloaded from `www.pombase.com` (uploaded on 1st Sept 2015, Wood et al. (2012)). All genes with an FDR<=10% were included in the testset while all other genes formed the background, if not indicated otherwise. Importantly only those genes with available measurements were included in the background to avoid false positive enrichments. `nodeSize` was set to 10. We performed Fisher's exact tests with the `weight01`-algorithm. Supplementary tables contain maximal ten terms per ontology with a p-value $p < 0.01$.

2.2.13 Generation of a *pka1* allele-replacement strain

To investigate the effects of the polymorphism in *pka1* (C358F between JB50 and JB760) we generated an allele replacement strain. This strain was identical to JB50 aside from position 1073 in the coding sequence of pka1, changing the respective codon from UGU, coding for cystein, to UUU, coding for phenylalanine. The strain was generated using a CRISPR/Cas9 method as described before (Rodríguez-López et al., 2017). It is referred to as PKA1Rep in the following. We used the following primers:
F: acataacctgtaccgaagaaAGCAACTGTTGTACTCTTTGgttttagagctagaaatagc
R: gctatttctagctctaaaacCAAAGAGTACAACAGTTGCTttcttcggtacaggttatgt
We used Sanger sequencing to validate that we successfully introduced the JB760-allele of *pka1* in the JB50-background.

To measure the effects of the *pka1*-allele replacement on the transcriptome and proteome, we grew three replicates each of JB50 and PKA1Rep under normal growth conditions and three replicates, for each strain, which were exposed to increased oxidative stress (0.5 mM H2O2) for one hour before the samples were harvested. Each sample was separated into a fraction for RNA extraction and one for protein quantification. The fraction for protein quantification was centrifuged and the pellets washed with cold PBS and snap frozen for transport.

2.2.14   Transcriptome quantification of the replacement strain

RNA was extracted with the hot phenol method described in (Lyne et al., 2003). RNA was further purified with Qiagen RNAeasy columns, and DNAse treatment was performed in the columns (as suggested by manufacturer) prior to library preparation. RNA quality was assessed with a Bioanalyzer instrument (Agilent, United States), and all samples had a RIN (RNA Integrity Number) > 9. cDNA libraries were prepared with the Illumina TruSeq stranded mRNA kit, according to the manufacturer's specifications, by the Cologne Centre for Genomics (CCG) facility. The samples were sequenced on a single lane of a Illumina Hiseq4000 to produce stranded 2x75 bp reads. Reads were trimmed with Trimmomatric (v0.36, Bolger et al. (2014)), with the following parameters differing from default settings: `LEADING:0 TRAILING:0 SLIDINGWINDOW:4:15 MINLEN:25`. The reference genome was indexed with bowtie2-build with default settings. Paired reads were aligned to the reference genome using bowtie2 with default settings (v2.3.4.1, Langmead and Salzberg (2013)). In the case of the allele replacement strain, the reference genome was edited to reflect the base substitution within pka1. Aligned reads were counted using intersect from the bedtools package (v2.27.1, Quinlan and Hall (2010)), with the parameters -wb -f 0.55 -s -bed. Identical reads were only counted once. Readcounts were tested for differential expression between strains using DESeq2 v1.18.1, with default settings (Love et al., 2013). We tested differential expression between strains separately in the two treatments. The stress response was also estimated separately per strain.

2.2.15 Proteome quantification of the replacement strain

$2 \cdot 10^7$ cells were lysed in 8M Urea (Sigma Aldrich, United States), 0.1M ammonium bicarbonate in presence of phosphatase inhibitors (Sigma Aldrich P5726&P0044) using strong ultra-sonication (Bioruptor, 10 cycles, 30 seconds on/off, Diagenode, Belgium). Protein concentration was determined by BCA assay (Thermo Fisher Scientific, United States) using a small sample aliquot. $200\mu g$ of proteins were digested as described previously (Ahrné et al., 2016), reduced with 5 mM TCEP for 60 min at 37 °C and alkylated with 10 mM chloroacetamide for 30 min at 37 °C. After diluting samples with 100 mM ammonium bicarbonate buffer to a final urea concentration of 1.6M, proteins were digested by incubation with sequencing-grade modified trypsin (1/50, w/w; Promega, United States) overnight at 37°C. After acidification using 5% TFA, peptides were desalted on C18 reversed-phase spin columns according to the manufacturer's instructions (Macrospin, Harvard Apparatus, United States) and dried under vacuum. The setup of the μRPLC-MS system was as described previously (Ahrné et al., 2016). Chromatographic separation of peptides was carried out using an EASY nano-LC 1000 system (Thermo Fisher Scientific, United States), equipped with a heated RP-HPLC column (75 μm x 37 cm) packed in-house with 1.9 μm C18 resin (Reprosil-AQ Pur, Dr. Maisch). Sample aliquots of 1 μg total peptides were analyzed per LC-MS/MS run using a linear gradient ranging from 95% solvent A (0.15% formic acid, 2% acetonitrile) and 5% solvent B (98% acetonitrile, 2% water, 0.15% formic acid) to 30% solvent B over 90 minutes at a flow rate of 200 nl/min. Mass spectrometry analysis was performed on Q-Exactive HF mass spectrometer equipped with a nanoelectrospray ion source (both Thermo Fisher Scientific, United States). Each MS1 scan was followed by high-collision-dissociation (HCD) of the 10 most abundant precursor ions with dynamic exclusion for 20 seconds. Total cycle time was approximately 1 s. For MS1, $3*10^6$ ions were accumulated in the Orbitrap cell over a maximum time of 100 ms and scanned at a resolution of 120,000 FWHM (at 200 m/z). MS2 scans were acquired at a target setting of 10^5 ions, accumulation time of 100 ms and a resolution of 30,000 FWHM (at 200 m/z). Singly charged ions and ions with unassigned charge state were excluded from triggering MS2 events.

The normalized collision energy was set to 27%, the mass isolation window was set to 1.4 m/z and one microscan was acquired for each spectrum. The acquired raw-files were imported into the Progenesis QI software (v2.0, Nonlinear Dynamics Limited, United States), which was used to extract peptide precursor ion intensities across all samples applying the default parameters. The generated mgf-files were searched using MASCOT against a decoy database containing normal and reverse sequences of the predicted S. pombe proteome (Uniprot, release date 2019/1/14) and commonly observed contaminants (in total 11,068 sequences) generated using the SequenceReverser tool from the MaxQuant software v.1.0.13.13). The search criteria were set as follows: full tryptic specificity was required (cleavage after lysine or arginine residues, unless followed by proline); 3 missed cleavages were allowed; carbamidomethylation (C) was set as fixed modification; oxidation (M) ; mass tolerance of 10 ppm (precursor) and 0.02 Da (fragments). The database search results were filtered using the ion score to set the false discovery rate (FDR) to 1% on the peptide and protein level, respectively, based on the number of reverse protein sequence hits in the datasets. The relative quantitative data obtained were normalized and statistically analyzed using an in-house script as above (Ahrné et al., 2016).

We used intensity based absolute quantification (iBAQ) levels to quantify proteins. For each sample we computed the 10% and the 90% quantile of the peptide counts and computed the mean of all counts between them (m_i for sample i). The average mean $\overline{m}$ was computed and each m_i was divided by it, resulting in m_i'. All abundances c in sample i and for protein j were then divided by m_i' to produce the normalized abundances.

$$c_{i,j}' = c_{i,j} \cdot \frac{m_i}{\overline{m}}$$

We tested proteins for differential expression by applying two-sided two-sample t-tests to the normalized levels of the proteins from the two groups of samples for each comparison. As for the transcriptome, we did not perform a two-factor test, but performed comparisons between the two strains separately in each condition and tested the effect of the treatment separately in each strain. We performed an FDR

correction for each comparison separately.

2.3 Results

2.3.1 Multi-omics characterization of a fission yeast panel

To investigate the influence of natural genetic variation on transcript- and protein-levels, as well as cellular fitness we generated a panel of 130 strains of *Schizosaccharomyces pombe* that was derived from three parental strains (JB50, JB759 and JB760). As this study aimed to study the effects of genetic variations on the response to oxidative stress, we chose founder strains that differed in this aspect. The parental strains included a strain that was sensitive to oxidative stress (JB759), a strain that was resistant (JB760) and a strain with an intermediary phenotype (JB50, the laboratory strain). We grew a total of 286 samples for 130 strains spread across two experimental conditions, including biological replicates for some strains. Samples were grown either on a full medium (normal condition, 'N') or a full medium with the addition of 0.5 mM H202 (stress-condition or 'H', exposed for one hour), a source of oxidative stress. RNA from these samples was extracted and sequenced with a median coverage of 44X. The transcriptome of all 130 strains was quantified at least once in both conditions. 152 separate samples for a total of 61 strains were grown for proteomics measurements and quantified via mass spectrometry (MS). For 59 of these strains the proteome was quantified at least once in both conditions.

We were able to measure a total of 7160 transcripts with no missing values in both conditions for all strains, corresponding to 92% of the annotated transcriptome. This included 2134 noncoding transcripts. The abundances of 16648 peptides was measured with less than 50% of missing values and successful measurements in at least two experimental batches. These peptides corresponded to 2184 unique proteins. For the proteins with multiple measured peptides, we combined positively correlated peptides to peptide groups. Our intention was to average out independent technical noise among the different peptides and to generate traits that are affected by technical

noise to a lesser degree than the individual peptide levels (Methods). The levels of 16648 peptides were combined to 6853 peptide groups. At least one peptide group was generated for all 2184 quantified proteins, corresponding to 42% of the proteome. 1433 proteins had multiple peptide groups. The median number of peptides that form an individual peptide group was 2.

We explored the relationship between the transcript- and protein levels of each gene by computing the Pearson's correlation between the transcript level and the level of each peptide group separately across the panel of our strains. Most peptide groups were positively correlated with the transcript level of the same gene (Figure 2.1). The levels of transcripts and peptide groups for the same gene correlated better in the normal condition than in the stress-condition (one-sided paired Wilcoxon's rank sum test $p < 10^{-16}$).
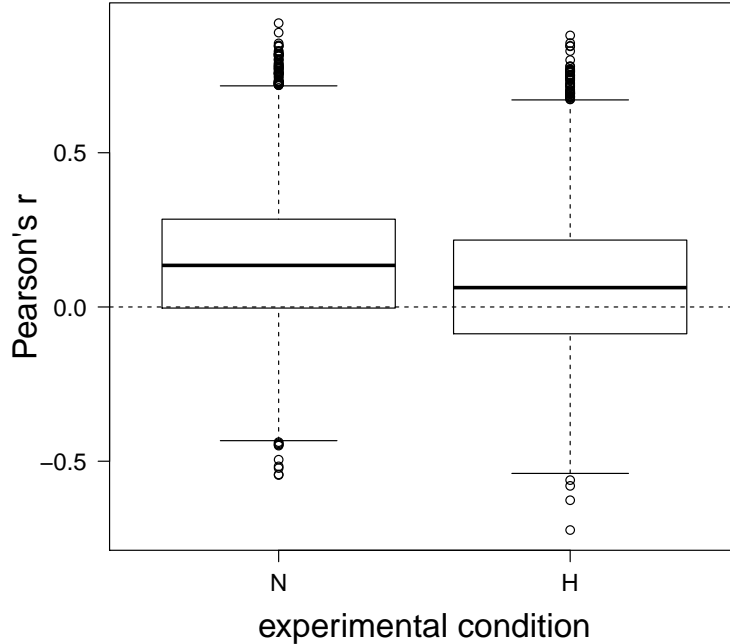
Figure 2.1: The Pearson's correlation between the levels of transcripts and levels of peptide groups of the same gene is shown separately per condition. N and H refer to the normal condition and the stress-condition respectively. All genes with available data on the transcript and peptide group level are shown. Proteins with multiple peptide groups are shown multiple times.

To be able to investigate the consequences of changes on the transcriptome and proteome layers for cellular fitness, we generated growth curves by measuring the optical density of liquid cultures. From these growth curves we extracted the total change in optical density (ΔOD). As the cultures were grown until they entered the stationary phase, ΔOD should be considered as the growth efficiency rather than growth speed. We estimated ΔOD for 114 strains in three growth conditions, differing in the concentration of hydrogen peroxide (no, medium and high H2O2). In order to explore the relationship between genetic variation and quantitative traits we genotyped our panel of strains. We used a previously described strategy to call

variants within coding regions of the genome from RNA-seq data and the parental genomes (Clement-Ziza et al., 2014). In total we called 7548 distinct variants for our panel of strains. These variants were in high linkage disequilibrium with their direct neighbors. To improve the power and interpretability of the subsequent QTL-mapping we combined the characterized variants to 1693 genetic markers that were not identical to their neighbors in the full panel of 130 strains. These markers had an average of 0.8% of missing values.

Notably, a large region of Chromosome I is inverted in JB50, but not the other two parental strains. This led to reduced recombination and increased linkage disequilibrium in that region, complicating the identification of causal variants.

2.3.2   Stress response

Organisms react to environmental changes with a complex response that includes widespread changes in gene expression. The response of a strain isogenic to JB50 to oxidative stress was described before (Chen, 2003). We estimated differential expression in JB50, the laboratory strain, by computing the log2-differences between expression levels in the normal and stress conditions. 2859 transcripts had an absolute log2 fold change of at least 1, of which 1388 increased and 1471 decreased in expression in the samples exposed to oxidative stress compared to those in the normal condition. Overall, noncoding transcripts were strongly enriched among the differentially expressed transcripts (Fisher's exact test: $p < 10^{-16}$, $OR = 2.17$). We observed both induced and repressed noncoding transcripts (595 and 539 respectively).The coding transcripts whose levels increased at least twofold were strongly enriched in cellular functions related to detoxification, and especially in functions related to the oxidative stress response (Supplementary Table 5.2). Repressed transcripts were strongly enriched in translation and ribosome biogenesis (Supplementary Table 5.1). Previous work has demonstrated the repression of ribosomal protein as part of the core environmental stress response (Chen, 2003; Gasch et al., 2000). The log2 fold changes of transcripts correlated strongly with previously published data ($r = 0.84$) (Chen, 2003), confirming that the observed expression differences are

mainly due to the stress-treatment.

We also computed the log2 fold changes for all peptide groups between normal and oxidative stress conditions in samples of JB50, the laboratory strain. For proteins with multiple peptide groups, we averaged the log2 fold changes per protein. 265 proteins had an absolute log2 fold change that was larger than 1, with 170 stress-induced proteins and 95 proteins that were repressed by oxidative stress. The stress response on the two molecular layers was positively correlated ($r = 0.29$), indicating that the transcriptomic changes caused by the exposure to oxidative stress were transmitted to the proteome to some degree (Figure 2.2). While proteins with up-regulated transcripts also increased in abundance in the stress condition (negative log2 fold changes, $r = 0.44$), proteins whose transcripts were repressed in the stress condition were not affected in the same way (positive log2 fold changes, $r = 0.04$).
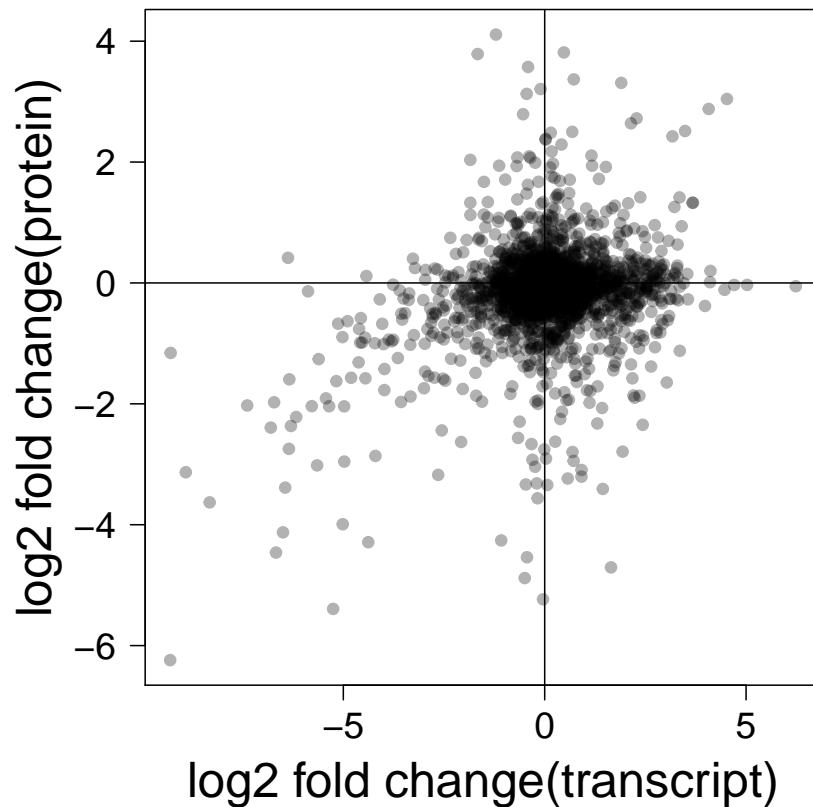
Figure 2.2: Log2 fold changes of transcript and protein levels between the normal and stress-conditions in JB50. If a protein had multiple peptide groups, they were averaged. Positive values indicate higher trait levels in the normal condition, and negative values indicate induction in the stress-condition.

2.3.3 The effects of genetic variants on the transcriptome often differ between experimental conditions

We investigated the effects of genetic variance on molecular traits with a QTL-mapping approach that we developed earlier (Clement-Ziza et al., 2014; Michaelson et al., 2010). We mapped QTL for transcripts, peptide groups and growth traits in normal and oxidative stress conditions. We detected at least one QTL for 47% of transcript-levels and for 1.3% of the peptide groups at FDR<10% in the normal condition. After stress treatment we detected at least one eQTL for 85% of the transcript-traits and at least one pQTL for 3.7% of the peptide-groups at FDR<10%. We combined all QTL with a FDR<10% across both conditions, resulting in 12729 eQTL and 303 pQTL. QTL that were significant in both conditions, were combined to a single one.

To be able to classify the activity of QTL across conditions in an unbiased way, we tested for an interaction between the marker of the QTL with the highest significance and a dummy variable representing the condition, on the trait values, using an ANOVA as described in Ackermann et al. (2013). Small p-values for the interaction between the QTL and the condition indicate that the effect of the QTL was not the same in both conditions. This could be due to the QTL affecting the trait in only one condition or it affecting the trait in different ways in the two conditions. That difference could be in effect size or in effect direction. To identify the conditions in which the QTL is active, the significance of the effect of the QTL in each condition was assessed separately, also with an ANOVA. QTL without a significant interaction effect with the condition were classified as 'static', meaning that their effects were very similar in both conditions. Those QTL that had a significant interaction effect were classified as 'conditional'. We defined several classes of conditional QTL: Those that were only active in the normal condition were classified as 'conditionally active in N', while those only active in the stress-condition were termed 'conditionally active in H'. QTL that had strong but different effects in both conditions are referred to as 'conditionally active in N and H' in the following.

We tested all 12729 eQTL and 303 pQTL at FDR<10% for conditional activity

(Figure 2.3, Figure 2.4). 40.9% of all eQTL were called as static (Figure 2.5a), 41.4% of eQTL were called as being only active in the stress-condition, while only 12.2% of eQTL were called to only affect transcription in the normal condition. In contrast to the eQTL, the vast majority of pQTL was called as being static (94%, Figure 2.5b). Possible reasons for this are considered later (Discussion).

The effects of eQTL across the conditions showed clear differences between the eQTL-classes we called (Figure 2.6). The effects of static eQTL were very similar in both conditions (r=0.59). Conditionally active eQTL showed larger effects in the condition they were called for. For those classes of eQTL the correlation between effects in N and H was weak (conditionally active in N: r=0.08, and conditionally active in H: r=-0.1). Notably, conditional eQTL that were active in both conditions affected gene expression in the same direction for most targets, but did so with a difference in effect size (r=0.6). eQTL with strong effects in both conditions and opposite effect directions between the conditions, were very rare.
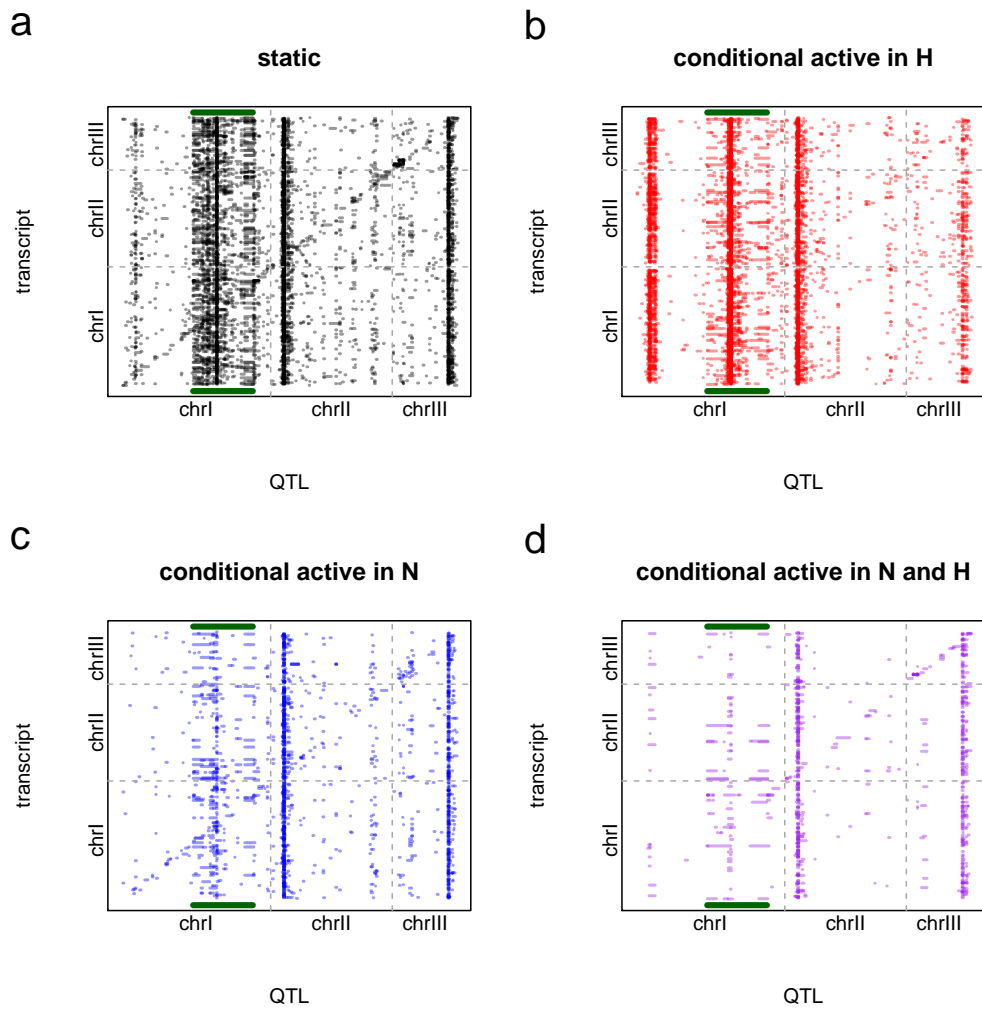
Figure 2.3: Significant eQTL at FDR<10%, shown separately for each class of eQTL. The inversion on Chromosome I is marked with green bars. **a** depicts the static eQTL, which are active in both conditions in a similar way. **b** shows those eQTL that only have a significant effect in the stress-condition. **c** shows all eQTL that are only active in the normal conditions. **d** describes the eQTL, that are active in both conditions but have effects that are different in size or direction.
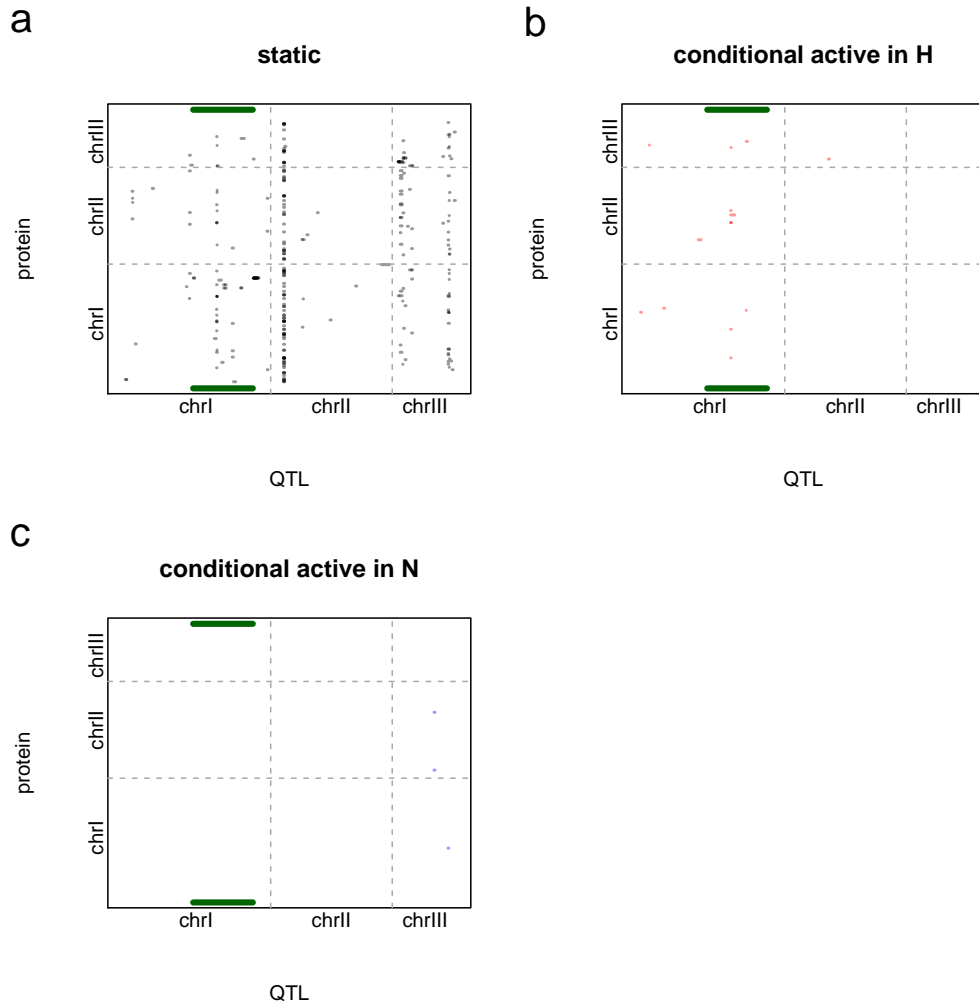
Figure 2.4: Significant pQTL at FDR<10%, shown separately for each class of pQTL. The inversion on Chromosome I is marked with green bars. **a** depicts the static pQTL, which are active in both conditions in a similar way. **b** shows those pQTL that only have a significant effect in the stress-condition. **c** shows all pQTL that are only active in the normal conditions.

a

eQTL

b
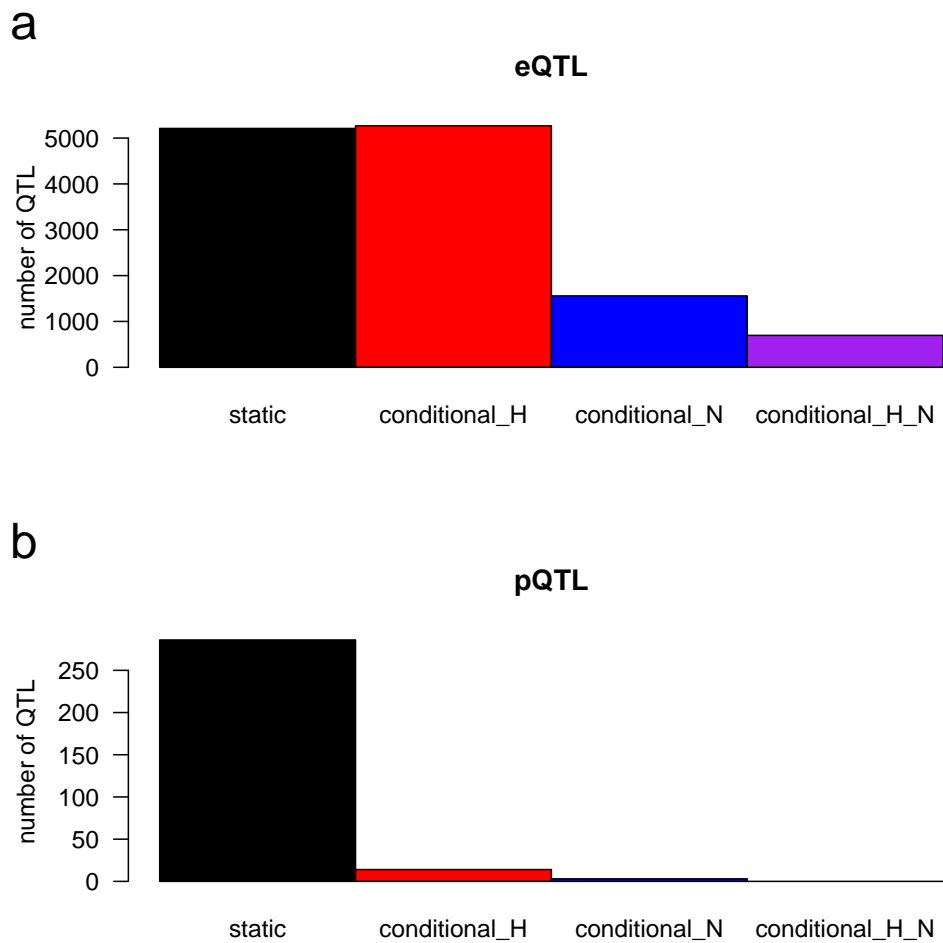
pQTL

Figure 2.5: Number of QTL at FDR<10% in each class. **a** shows the distribution of eQTL across classes and **b** depicts the distribution of pQTL.

Transcripts that were affected by conditional eQTL did not have a stress response that was biased in one direction, i.e. some targets were up-regulated and some were down-regulated upon the exposure to oxidative stress (Figure 2.7a), but they did have an overall stronger stress response than transcripts which were affected by static eQTL (Figure 2.7b). This was especially true for transcripts that were affected by an eQTL that was only active in the normal condition (Wilcoxon's rank sum tests;

N: $p < 10^{-16}$, H: $p = 0.03$).
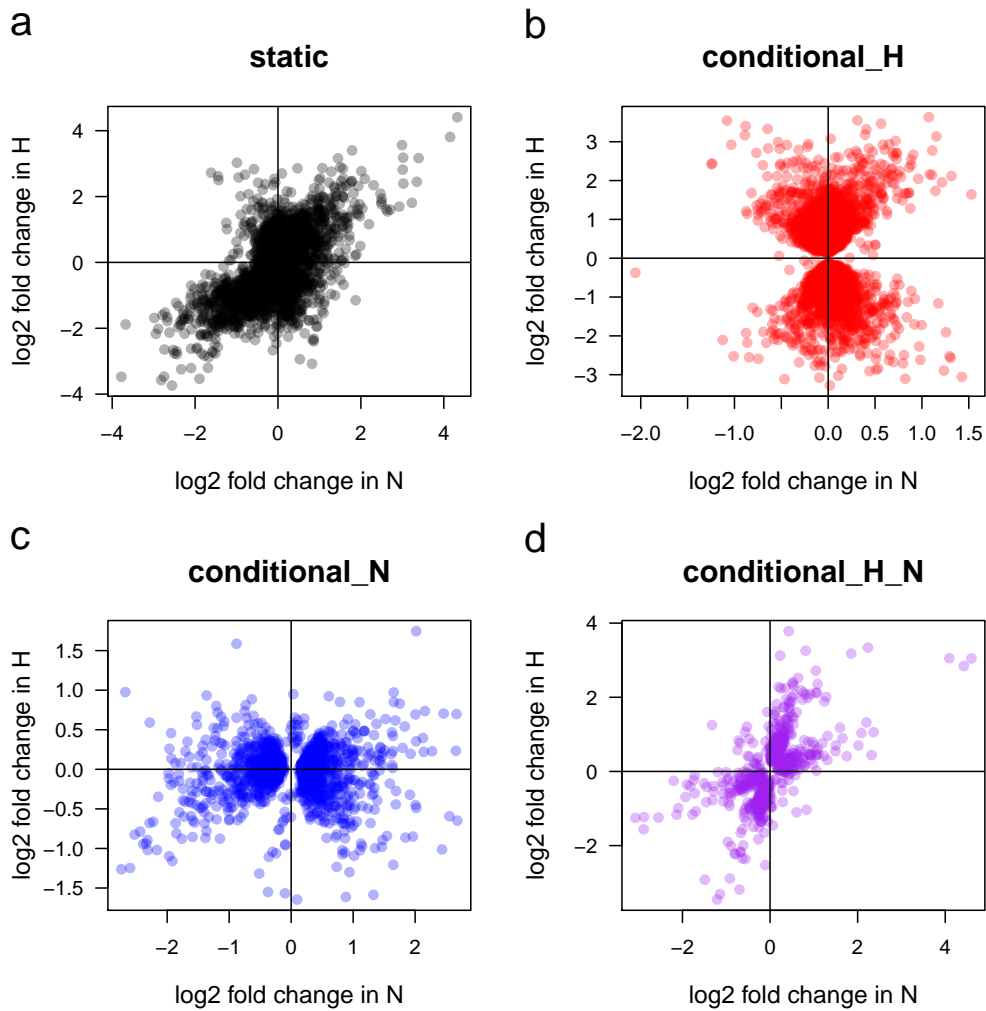
Figure 2.6: The effects of eQTL on the abundances of their targets in the normal and stress-conditions are shown separately for each class of eQTL. Each dot represents an eQTL. Static eQTL with similar effects in both conditions (**a**), conditional eQTL with large effects in the stress condition but no significant effects in the normal condition (**b**), conditional eQTL that only affect trait levels in the normal condition but not the stress-condition (**c**), and eQTL with significant but different effects in both conditions (**d**).
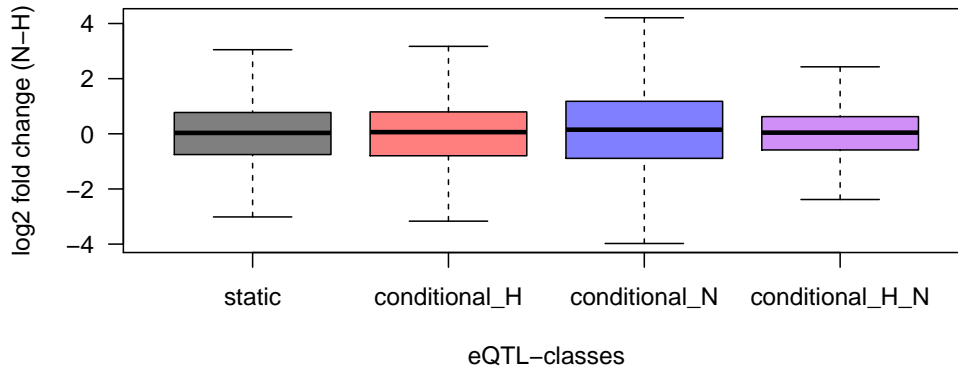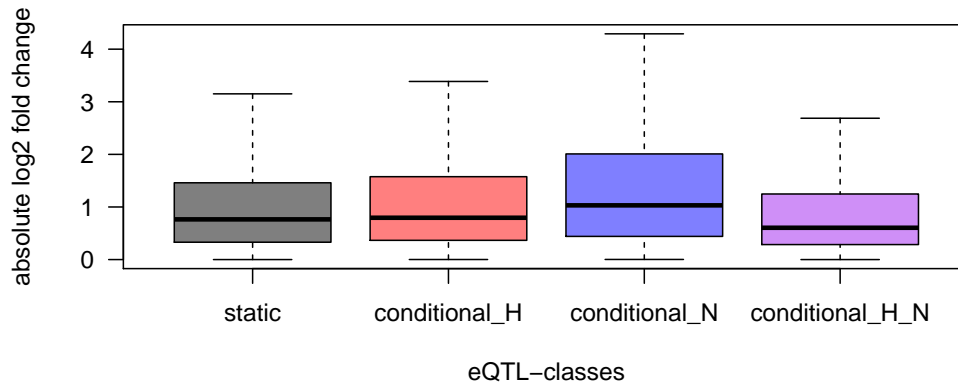
Figure 2.7: Log2 fold changes of transcript levels between the normal and stress-conditions in JB50, shown separately for each activity pattern of eQTL. The same transcript can be affected by multiple eQTL with different activity patterns, and can be included in different boxes.

QTL that affect products of genes, which are encoded close to the QTL, are often considered to be local, while QTL that are further away from the affected gene are considered to be distant (Albert and Kruglyak, 2015). We classified QTL to be local if they included a marker that had an absolute correlation of 0.8 or higher with one of the markers closest to the target gene, all other QTL were classified as distant. 8.7% of all eQTL were local, affecting a total of 1002 transcripts. The amount of local eQTL was larger than expected if genes were targeted by eQTL independently of their position in the genome (p=0.005), suggesting the effects of local variation on the expression of nearby encoded molecular traits are especially important. We found at least one local pQTL for 24 proteins at FDR<10% (32 peptide groups with local pQTL). As for local eQTL, this was higher than expected by chance (p=0.005). Genes with a local pQTL and a local eQTL overlapped significantly (Fisher's exact test, $p=10^{-14}$). 20 out of 24 genes with a local pQTL also had a local eQTL, indicating that most local pQTL are caused by local eQTL in this cross.

Since we observed different activity patterns of eQTL across the two conditions, we asked whether local and distant eQTL differed in this regard. When all eQTL were considered, the relationship between the activity pattern of an eQTL across conditions and it being local or distant was significant (Chi-squared test: p=0.02). As there was high linkage disequilibrium within the inverted region on Chromosome I, it was harder to accurately call local QTL in this region. For this reason we excluded all eQTL and targets within the inversion from this analysis. After restricting the eQTL in this fashion, the relationship between the activity pattern of an eQTL across conditions and it being local or distant became even more distinct (Chi-squared test: $p=10^{-11}$). *Post hoc* tests revealed that especially eQTL which were only active in the stress-condition differed in their probability to also be local to their target from other eQTL (Fisher's exact test: $p < 10^{-10}$, $OR = 0.5$). Compared to eQTL with other activity patterns, eQTL that were only active in the stress-condition were much less likely to be local to their target (Figure 2.8). Conditional eQTL that were active in both conditions were more likely to be local to their target ($p < 2 * 10^{-4}$ and OR=1.87). 65 out of 66 local eQTL with significant but different effects in the two conditions affected the expression of their targets in the same direction in both

conditions. These results indicate that local eQTL are often active in all conditions.
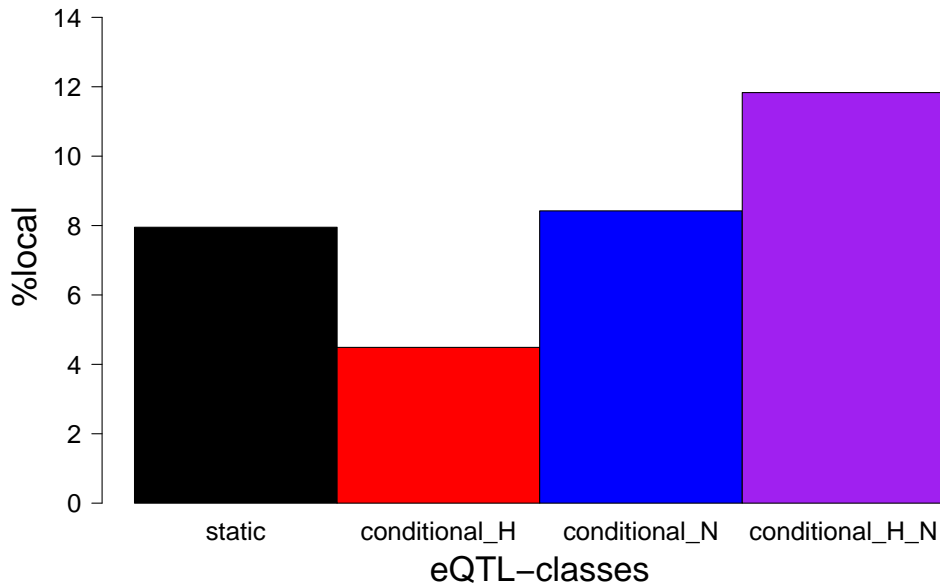
Figure 2.8: Proportion of eQTL for each activity-class, that are also local. eQTL within the inversion of Chromosome I were excluded from this analysis.

2.3.4 Most eQTL-hotspots have a strong effect on the proteome

While we found a large number of distinct QTL that regulated molecular traits, some loci served as QTL for a large number of molecular traits (visible as vertical lines in Figures 2.3 and 2.4). Such loci are often referred to as 'QTL hotspots' (Smith and Kruglyak, 2008). These hotspots often result in physiological phenotypes and can be characterized in detail (Brem et al., 2002; Smith and Kruglyak, 2008; Clement-Ziza et al., 2014).

We identified hotspots by binning the genome and comparing the observed number of associations for each bin with one that would be expected if QTL were distributed

across the bins according to a Poisson-distribution, as proposed before (Brem et al., 2002). Consecutive bins on the same chromosome that were called as hotspots were combined to a single one. We identified four eQTL hotspots (Table 2.1). The hotspots were large in size (mean=827kb). In the case of the eQTL-hotspot e-chrI:2, it encompassed the whole genomic inversion on Chromosome I (Figure 2.9). Three of these hotspots (e-chrI:1, e-chrI:2, and e-chrIII:1) were described in a previous study on a part of this cross, derived from the JB50 and JB759 strains (Clement-Ziza et al., 2014), while the hotspot e-chrII:1 was described in this study for the first time. Interestingly, the activity of the eQTL across conditions differed between the different hotspots and often also within the same hotspot (Figure 2.10). Especially the two eQTL-hotspots on Chromosome I were more active in the stress-condition. The hotspot on Chromosome II contained eQTL of all activity patterns, highlighting the importance of environmental influences on the effects of genetic variants on gene expression. The targets of the eQTL-hotspots were strongly enriched in shared functions (Supplementary Tables 5.3 - 5.6).

The robust identification of hotspots requires large numbers of QTL (Wu et al., 2008). As the pQTL at FDR<10% were extremely sparse, we considered all pQTL at FDR<25% to more accurately identify pQTL-hotspots (Figure 2.10). Using this criterium, we identified eight pQTL hotspots (Table 2.2). Aside from the eQTL-hotspot e-chrI:1 all eQTL-hotspots overlapped with pQTL-hotspots. In addition we identified two hotspots that affected the proteome independently of the transcriptome (p-chrI:2, and p-chrIII:1). A previously reported eQTL-hotspot, caused by a frameshift within *swc5* (p-chrIII:2), also affected a large number of proteins at FDR<25%. The targets of most hotspots were enriched in particular functions (Supplementary Tables 5.7 - 5.12)

A hotspot within the inversion on Chromosome I which was reported in a previous study (e-chrI:2) had more targets that were only regulated after the induction of oxidative stress than targets of static eQTL. Strains with the JB759-allele at this locus had higher levels of transcripts of ribosomal proteins and lower levels of transcripts related to autophagy and acospore formation in the stress-condition (Supplementary Tables 5.13 - 5.15). These changes in the transcriptome are consistent with a reduced

stress response in the strains that carry the JB759-allele.

Table 2.1: eQTL hotspots.

name	chr	first marker	last marker	start (bp)	end (bp)	targets (unique)	targets (static)	targets (H)	targets (N)	targets (H and N)
e-chrI:1	I	35	72	450023	760993	992	160	818	10	19
e-chrI:2	I	270	866	2658236	5030169	3716	1655	2120	213	61
e-chrII:1	II	915	981	376806	747051	3196	1239	1250	572	310
e-chrIII:1	III	1556	1664	2047332	2298791	1620	831	221	366	208

Table 2.2: pQTL hotspots.

name	chr	first marker	last marker	start (bp)	end (bp)	targets (unique)	targets (static)	targets (H)	targets (N)	targets (H and N)
p-chrI:1	I	29	34	396286	448771	14	12	2	0	0
p-chrI:2	I	243	253	2526720	2569788	31	28	4	0	0
p-chrI:3	I	351	452	3011981	3356530	29	22	6	1	0
p-chrI:4	I	480	701	3451791	4199795	151	135	20	0	0
p-chrI:5	I	797	810	4549139	4602196	15	14	1	0	0
p-chrII:1	II	916	927	387522	549453	182	179	4	0	0
p-chrIII:1	III	1438	1463	254886	714880	159	156	1	3	0
p-chrIII:2	III	1556	1680	2047332	2349859	160	152	5	3	0
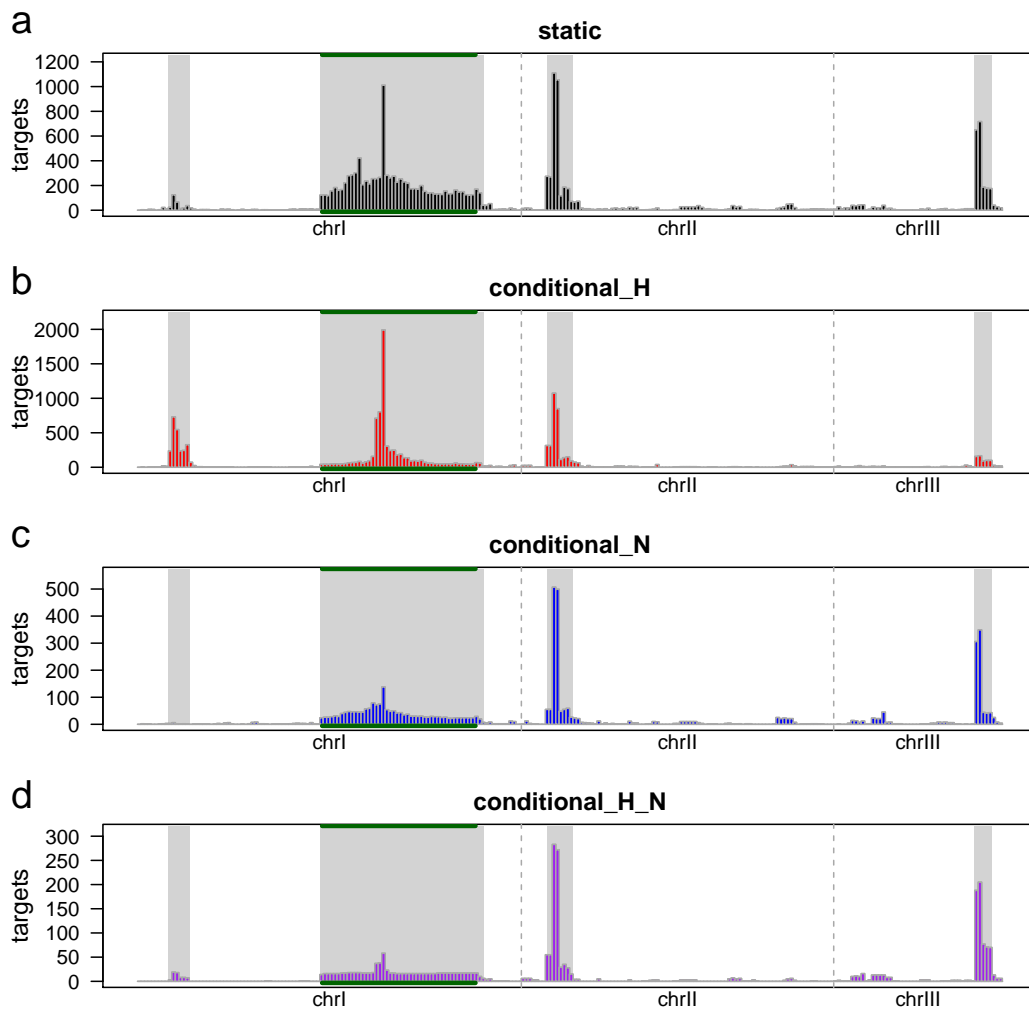
Figure 2.9: The number of transcripts that are affected by eQTL at FDR<10% in each 50kb-bin of the genome is depicted as vertical bars. These counts are shown separately for transcripts affected by static eQTL (**a**), conditional eQTL that are only active in the stress-condition (**b**), conditional eQTL that are only active in the normal condition (**c**), and conditional eQTL that are active in both conditions (**d**). Hotspot regions are marked by grey rectangles. The chromosomal borders are shown by dotted vertical lines. The inverted region on Chromosome I is indicated by green horizontal lines.
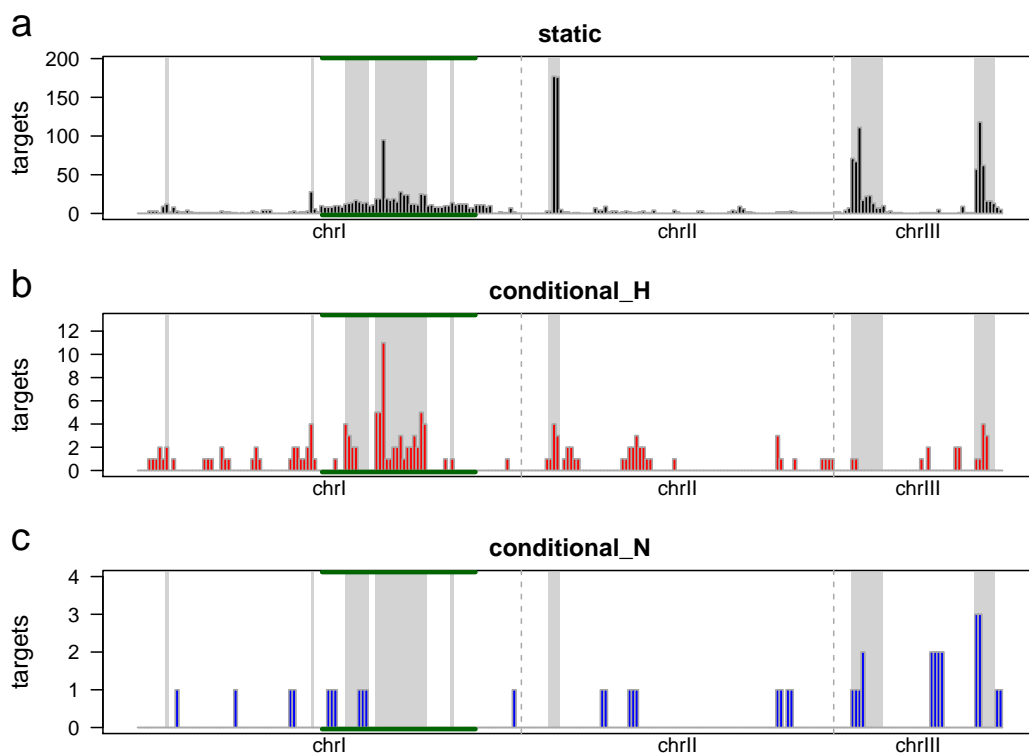
Figure 2.10: The number of proteins that are affected by pQTL at FDR<25% in each 50kb-bin of the genome is depicted as vertical bars. The number of targets is shown separately for proteins affected by static pQTL (**a**), conditional pQTL that are only active in the stress-condition (**b**), and conditional pQTL that are only active in the normal condition (**c**). Hotspot regions are marked by grey rectangles. The chromosomal borders are shown by dotted vertical lines. The inverted region on Chromosome I is indicated by green horizontal lines.

2.3.5 An eQTL-hotspot on Chromosome I affects the response to oxidative stress

The eQTL-hotspot e-chrI:1 affected the abundance of 992 different transcripts (Figure 2.9, Table 2.1). The parental strain JB759 had a different allele at this locus than the other two parental strains JB50 and JB760. 82% of the targets of this locus were affected in a stress-specific way. While the hotspot had large effects on

the transcriptome, we detected only 2 pQTL at FDR<10%, 15 pQTL at FDR<25%, and no significant growth QTL for this locus. Most of the eQTL-targets of this locus that were only affected in the stress-condition had a decreased stress response in the strains with the JB759-allele of the marker with the most targets at this locus (74%, Figure 2.11, Methods).

Overall, the eQTL targets of e-chrI:1 were strongly enriched for functions related to ribosome biogenesis (Supplementary Table 5.3). To further elucidate how this hotspot affects different groups of targets, we separated the transcripts that were targeted by this hotspot through stress-specific eQTL into four groups, based on the direction of the stress response, and for which allele it was stronger. The direction of the stress response was the same between both alleles for most transcripts (78%). We tested each group of targets for enrichments compared to the other three groups. Genes whose transcripts were stress induced in the strains with the JB50/JB760-allele and whose stress response was stronger in the strains with the JB50/JB760-allele at this locus were enriched for general stress response functions, autophagy and membrane localization (Supplementary Table 5.16). Genes that were repressed in the stress-condition in strains with the JB50/JB760-allele and whose stress response was more extreme in the strains with the JB50/JB760-allele were strongly enriched for ribosome biogenesis (Supplementary Table 5.17). Genes whose stress response was less extreme in strains with the JB50/JB760-allele and up-regulated in the same strains in the stress-condition, on the other hand, were enriched for growth related functions, such as alpha-amino acid synthesis and macromolecule biosynthesis (Supplementary Table 5.18). This was also true for transcripts that also responded less strongly to the stress treatment and were down-regulated in the strains with the JB50/JB760-allele (Supplementary Table 5.19).

The strains with the JB50/JB760-allele were able to induce the expression of stress response genes and repress ribosome biogenesis on the transcript level in response to oxidative stress while the strains with the JB759-allele were less able to do so. The JB759-allele at this locus confers a reduced stress response.
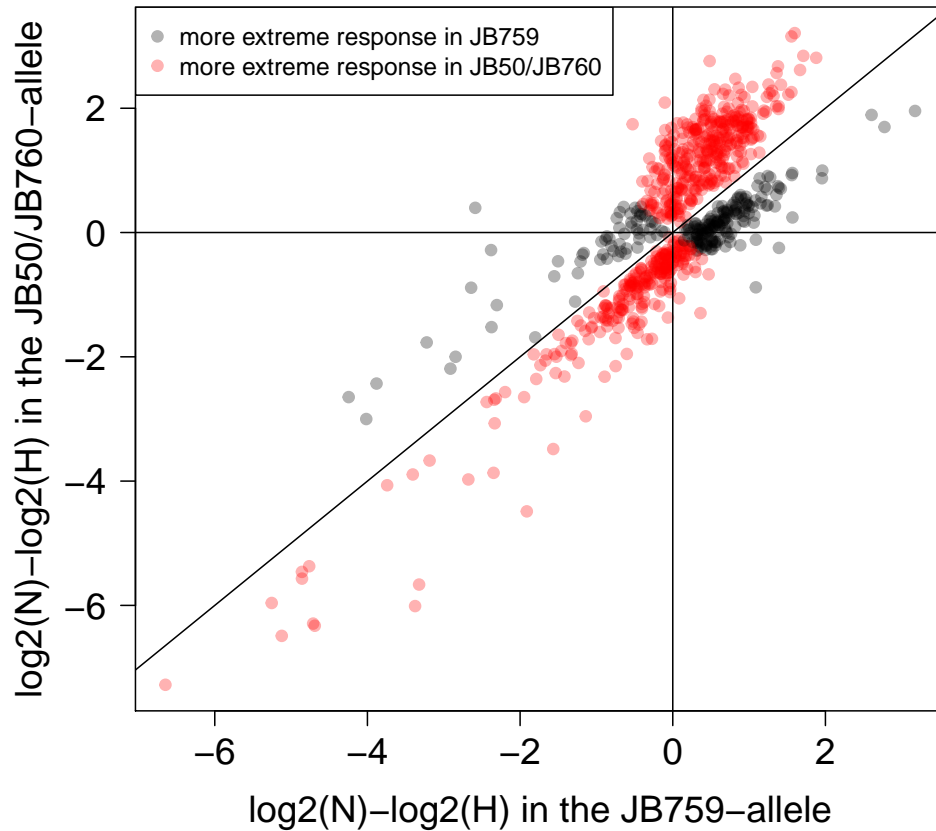
Figure 2.11: Effects of HS e-chrI:1 on the stress response of those genes that a had an eQTL at this locus that was only active in the stress condition. The axes are shown as vertical and horizontal lines. The diagonal is shown as a solid black line. Red dots represent genes whose transcripts have a larger absolute log2 fold change between the normal and the stress-condition in the strains with the JB50/JB760-allele. Black dots represent the genes whose transcript responded more strongly in the strains with the JB759-allele.

Although we mapped protein-level traits for many of the genes whose stress response was affected by the hotspot on Chromosome I, we did not observe a pQTL hotspot at this locus. A direct comparison of the effects of this hotspot on the transcript and protein levels of genes that had stress-only eQTL at this locus revealed that the levels of the peptide groups belonging to the respective proteins were affected in the same direction as their transcripts (Figure 2.12). The effects of the locus on the levels of mRNAs and peptide groups was significantly positively correlated ($r = 0.45$, $p < 10^{-16}$). Most peptide-groups were affected to a lesser degree than the transcript of the same gene (87.5%). This could reflect buffering of protein levels against variation on the transcript level, or it could be related to the general lack of response of the proteome to transcriptional repression after the treatment with oxidative stress.

We analyzed the genes in this hotspot to identify a causal gene. The genomic region of the hotspot contained 185 genes, of which 124 were coding. To identify a potentially causal gene, we concentrated on those genes that either had local eQTL, local pQTL or coding polymorphisms between the parental strains that produced the subcrosses which showed the effects of the hotspot. 38 genes satisfied these criteria. Of these, the most promising candidates were *cph2*, *akr1*, and *nrd1*. All three genes were in close vicinity of the markers with the highest number of targets and had polymorphisms that changed the amino acid sequence, with JB50 and JB759 sharing one allele and JB760 having a different allele. In addition, *akr1* had a local eQTL. *Cph2* is a member of the RPD3S-complex which is involved in the repression of spurious transcription through epigenetic modifications (Lickwar et al., 2009). While some members of the RP3DS-complex are involved in forming the related RPD3L-complex, which is also involved in the response to oxidative stress, *cph2* is not one of them (Alejandro-Osorio et al., 2009; Nicolas et al., 2007). *Akr1* is a palmitoyl-transferase involved in signaling and protein localization (Zhang et al., 2013). The third candidate gene, *nrd1*, is known to localize to several kinds of RNA-granules, including granules containing PABp, and to bind to mRNAs directly (Satoh et al., 2012; Kobayashi et al., 2013). The localization of *nrd1* was reported to be partially dependent on oxidative stress (Satoh et al., 2012). Further experiments would be

56

necessary to determine a causal gene. Due to its known link to the stress response, *nrd1* was a strong candidate.
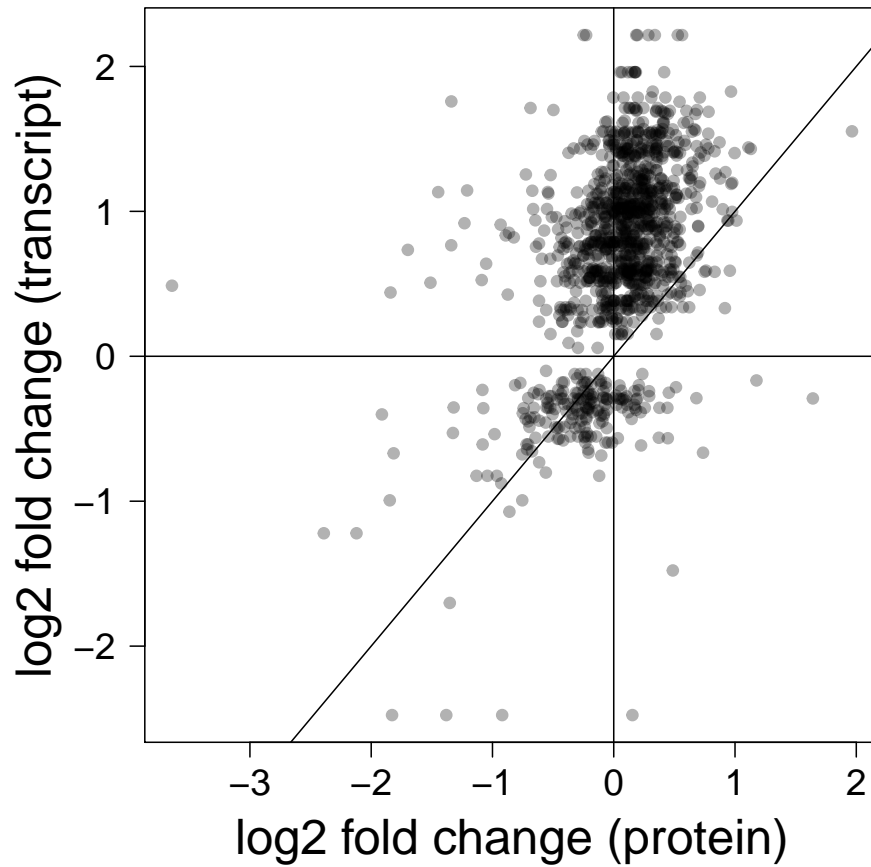
Figure 2.12: Effects of HS e-chrI:1 on the levels of peptide groups whose transcripts had an eQTL, that was only active at the stress condition, at the same locus. Shown are the levels of the peptide groups after stress treatment. The axes and the diagonal are shown as solid black lines. Negative log2 fold changes identify genes with higher expression in the strains with the JB50/JB760-allele.

2.3.6 Growth efficiency is affected by genetic variation

To investigate the effects of genetic variation on cellular fitness, we mapped the growth efficiency for 112 strains in three conditions that differed in the intensity of the oxidative stress. We identified three growth QTL for the normal condition, one for the intermediary stress-condition and two for the harsher stress-condition (Table 2.3, Figure 2.13). Several gQTL affected ΔOD in multiple conditions. The QTL-hotspot on Chromosome II (e-chrII:1/p-chrII:1) also affected ΔOD in the normal condition and the harsher stress-condition signficantly. A second locus on Chromosome II (bp 767643-808514) significantly affected the growth trait in the normal condition. ΔOD in the harsher stress condition was also affected by a third locus on Chromosome II (bp 2304929-2549939). A fourth QTL on Chromosome II (bp 3799420-3827447) significantly affected ΔOD in the normal condition and the intermediary stress condition. While these gQTL were only signficant in a subset of growth-conditions, they were among the most significant in all conditions (Figure 2.13).

ΔOD was highly correlated across the three conditions (for all pairs of conditions: r>=0.83). We estimated the average effect of each of these four loci in all three growth conditions by computing the mean differences between strains from the same cross with each allele (Methods). The QTL had very similar effects across all three

Table 2.3: gQTL that affect ΔOD. Conditions in which the gQTL is significant, are indicated with a 1.

name	first marker	last marker	chromosome	start (bp)	end (bp)	N (0 mM)	H (0.5 mM)	H (1 mM)
gQTL1	917	919	II	470768	521267	1	0	1
gQTL2	985	998	II	767643	808514	1	0	0
gQTL3	1201	1201	II	2304929	2549939	0	0	1
gQTL4	1362	1378	II	3799420	3827447	1	1	0

conditions, indicating that other causes than conditional effects could be responsible for the QTL to be significant only in a subset of these conditions (Figure 2.14). Notably, the JB759-allele was associated with smaller ΔOD for all four gQTL.
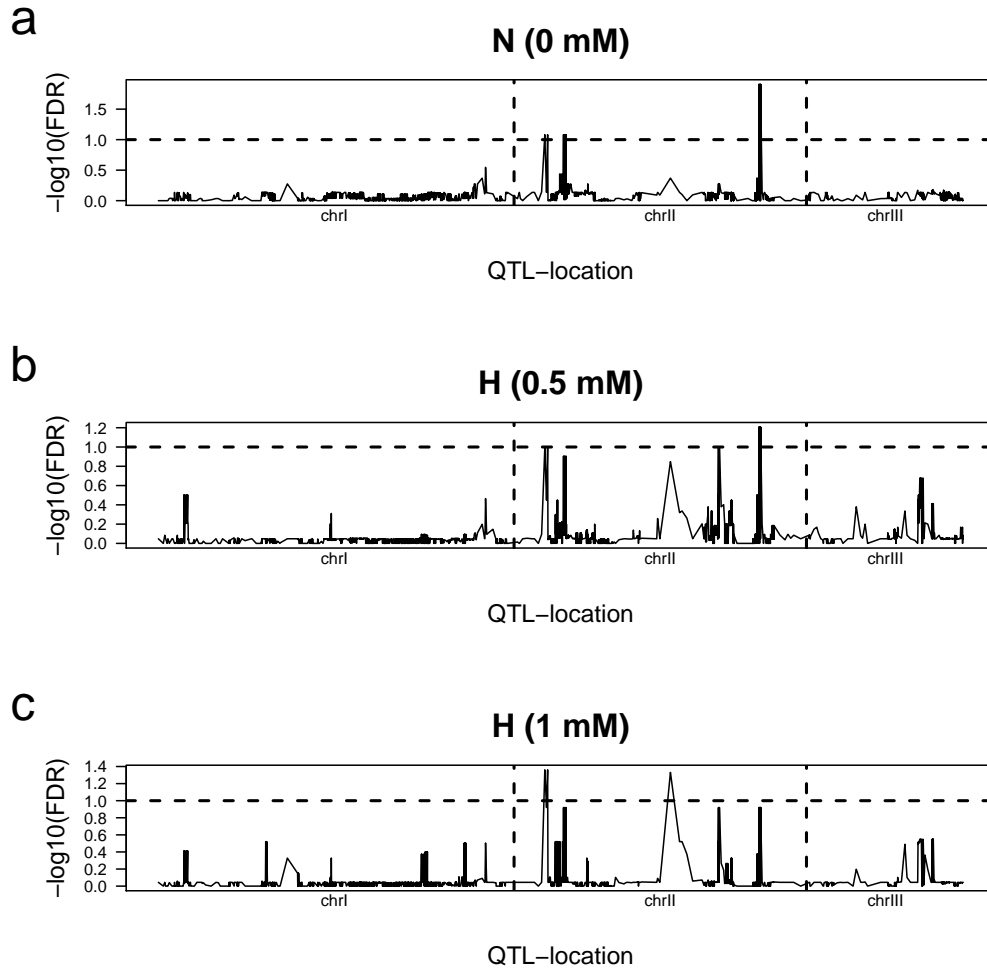
Figure 2.13: Significance of all genetic markers on the ΔOD growth trait in the normal condition (**a**) and two stress-conditions that differ in the concentration of H2O2 (**b** and **c**, respectively). The dotted vertical lines mark chromosome ends and the horizontal dotted line represents the significance threshold at FDR<10%.
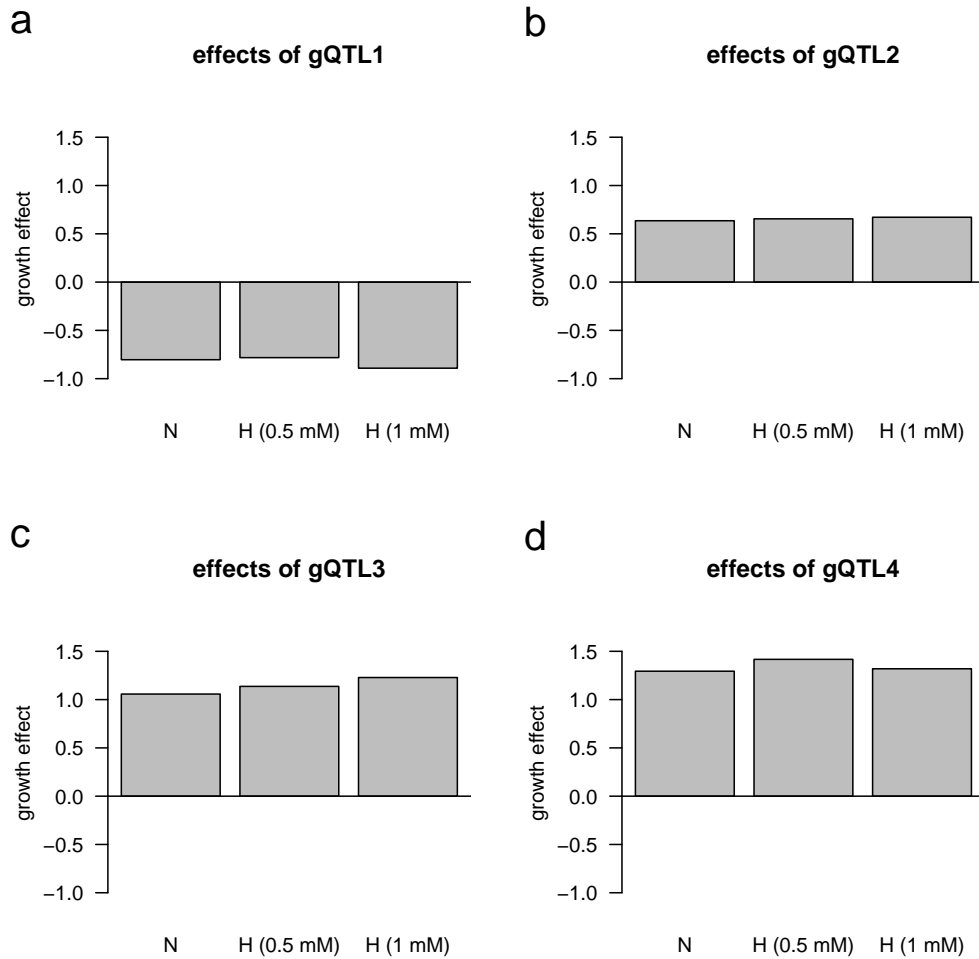
Figure 2.14: Average effects of the most significant marker of each gQTL in all three growth-conditions. The y-axis shows the difference in growth of strains with the JB50-allele at the respective locus compared to strains with the second allele. **a** to **d** show the effects of the gQTL shown in Table 2.3 in the same order.

2.3.7 A polymorphism in *pka1* explains effects on the transcriptome, proteome and cellular fitness

A locus on Chromosome II (chrII:376806-747051) had a significant effect on the expression of 3196 transcripts and 182 proteins, across the two conditions. As this was the largest eQTL- and pQTL-hotspot and it regulated growth, we investigated this locus in more detail. The marker region within this hotspot with the most linkages (chrII:470768 to chrII:521267) significantly affected the levels of 2645 transcripts and 113 peptide groups at FDR<0.1. In the following we concentrated on this marker when the effects of the hotspot are considered. We characterized the effects of this locus by performing a functional enrichment analysis of its targets in the transcriptome and proteome. Genes with an eQTL at this locus were enriched in functions related to ribosome biogenesis, glycolysis, alpha-amino acid biosynthesis, and proteins localizing to the mitochondria (Supplementary Table 5.5). We saw similar enrichments for genes with pQTL (Supplementary Table 5.10). We further compared the effects of the hotspot on the transcriptome with previously published data (Malecki and Bähler, 2016). Malecki et al. grew samples of JB22, a strain that differs only in the mating type from JB50, on both glucose and glycerol. In contrast to using glucose as a primary carbon source, fission yeast has to metabolize glycerol through aerobic respiration (Chiron et al., 2008). A direct comparison of the transcriptomic response to the induction of respiration and the effects of the hotspot revealed a strong agreement (r=0.81, Figure 2.15). Strains with the JB760 allele of this locus resembled samples growing on glycerol on the transcriptomic layer more than strains with the JB50/JB759 allele of *pka1*. Transcripts that were affected only in the normal condition by this hotspot were enriched for functions related to ribosome biogenesis and mitochondrial respiration (Supplementary Table 5.20). Specifically, transcripts related ribosome biogenesis were down-regulated in strains with the JB760-allele and transcripts related to meiosis and mitochondrial function were up-regulated (Supplementary Tables 5.21 and 5.22). Coding genes that were targeted by this locus exclusively in the normal condition, had a stress response that was highly correlated with the effects of e-chrII:1 on their abundances in this condition (r=0.82). e-chrII:1

affects these transcripts in way that is very similar to the stress response.

To identify a causal variant for this hotspot we analyzed the genes in the region. 106 genes were located in the overlap of the eQTL- and pQTL-hotspots at this locus. Of these genes, 69 were coding. We further concentrated on the genes that had a local eQTL or had missense-mutations that segregated in the R2- and R3-crosses. 37 genes were affected by a local eQTL and only two genes had polymorphisms resulting in a difference in amino acid sequence for which the JB760-allele differed from that of both JB50 and JB759. One of the genes in this region was *pka1*. *pka1* was a candidate causal gene for multiple reasons. It had both a local eQTL and was polymorph between the parental strain, with the JB760-allele differing from that of the other two parental strains. As the hotspot had a large number of targets the causal gene likely was well annotated and might have a functional connection to energy metabolism or a signaling role. *pka1* is well documented to have a crucial role in the PKA/RAS pathway, which integrates signals on glucose availability and stress exposure (Maeda et al., 1994). The enrichments of the eQTL-targets of this locus in ribosome biogenesis and mitochondrial respiration were consistent with differences in RAS-signaling causing the molecular effects of this hotspot, as these functions are often regulated in response changes in glucose availability (Malecki and Bähler, 2016).

pka1 had a polymorphism at position 1073 of the coding sequence, changing the respective codon from TGT in JB50/JB759, coding for a cysteine residue, to TTT in JB760, coding for a phenylalanine residue. The neighboring threonine residue at position 356 was reported as a phosphosite before, and its phosphorylation was shown to partially control the localization of *pka1* (Gupta et al., 2011). Using strains with phosphomimetic alleles of *pka1*, Gupta *et al.* showed that the phosphorylation at T356 increases the concentration of *pka1* in the nucleus. The inability to be phosphorylated at residue 356 was shown to result in a partial loss of *pka1*-function (Gupta et al., 2011). Taken together, this evidence motivated us to validate if the mutation in *pka1* was causal to the hotspot observed on Chromosome II. To investigate the causal mutation for this hotspot, we generated a strain that differed from JB50 only in the amino acid at position 356 of the *pka1* protein, using a previously published

approach utilising the CRISPR/Cas9 methodology (Rodríguez-López et al., 2017). This allele-replacement strain is referred to as 'PKA1Rep' in the following. This strain, together with JB50, allowed us to investigate the effects of the polymorphism in *pka1* without the confounding effects of polymorphisms at other genetic loci. We investigated the effects of this polymorphism through transcriptomic and proteomic profiling. Three samples each of JB50 and PKA1Rep were grown in the normal condition and with the addition of 0.5 mM H2O2. Each sample was divided and subjected to quantification of the transcriptome and proteome. In total, we were able to quantify 7749 transcripts. An analysis with DESeq2 revealed that 5335 transcripts were differentially expressed between the strains in the normal condition at FDR<10%. 4795 transcripts were differentially expressed in the stress-condition at FDR<10%. We also quantified the levels of 3378 proteins in both conditions. 2029 proteins were differentially abundant between JB50 and PKA1Rep in the normal condition and 1504 proteins were differentially abundant between the two strains in the stress-condition at FDR<10%. The changes occuring between JB50 and PKA1Rep in the transcriptome and in the proteome in the normal condition were compared directly, showing a strong positive correlation (r=0.79, Figure 2.16a). Transcriptomic and proteomic changes between the two strains in the stress-condition also showed strong similarity, but were less strongly correlated than those in the normal condition (r=0.5, Figure 2.16b).

The log2 fold changes at the transcriptomic level between JB50 and PKA1Rep correlated strongly with those observed for the hotspot in the full panel of strains when all coding genes with transcript levels in both experiments were considered (r=0.68 in the normal condition, r=0.34 in the stress-condition). The correlation between changes in the transcriptome was highest when only genes with available protein levels (r=0.82 in the normal condition, r=0.39 in the stress-condition, Figure 2.17a and b) were considered.
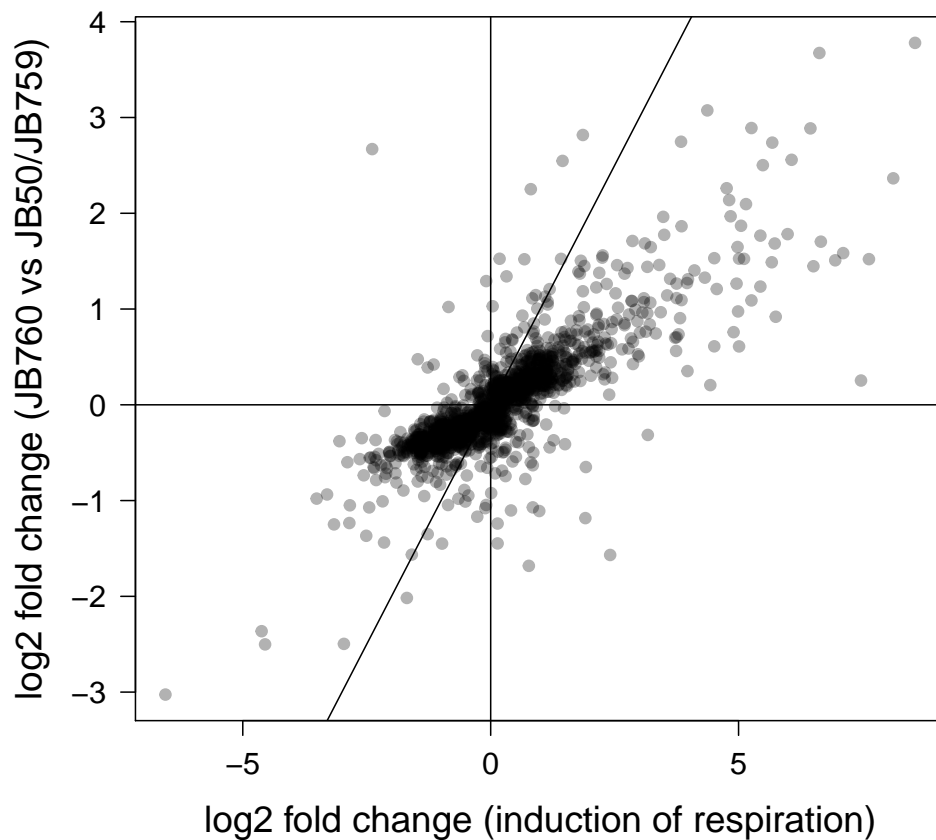
Figure 2.15: Changes in transcript levels upon the induction of respiration, using glycerol as the primary carbon source, compared to changes caused by the hotspot e-chrII:1. Each dot corresponds to a gene that is regulated by the hotspot in the normal condition. The diagonal and the axes are represented by black lines.
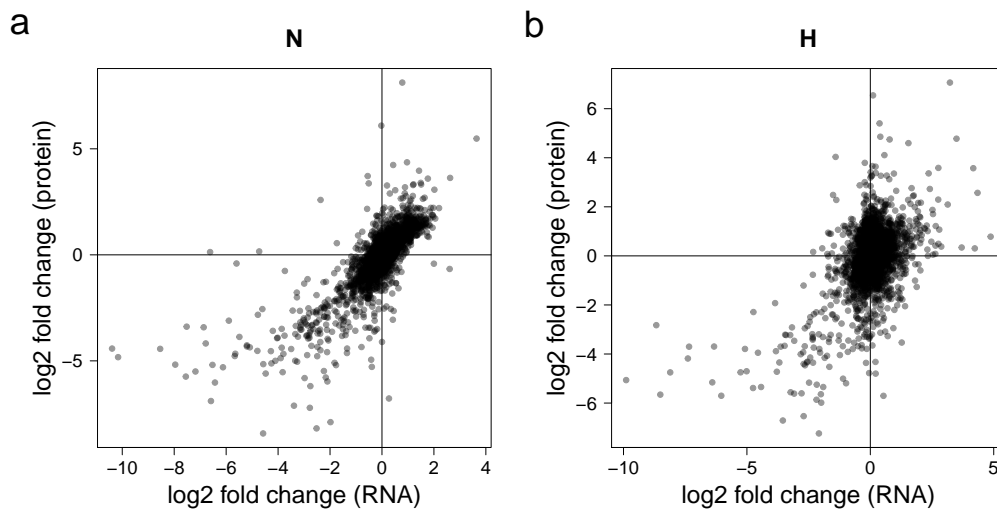
Figure 2.16: Differential expression between JB50 and PKA1Rep on the transcriptomic and proteomic layers for all genes with measurements on both layers. Differential expression is shown separately for the normal (**a**) and stress-condition (**b**). Axes are represented by solid black lines.
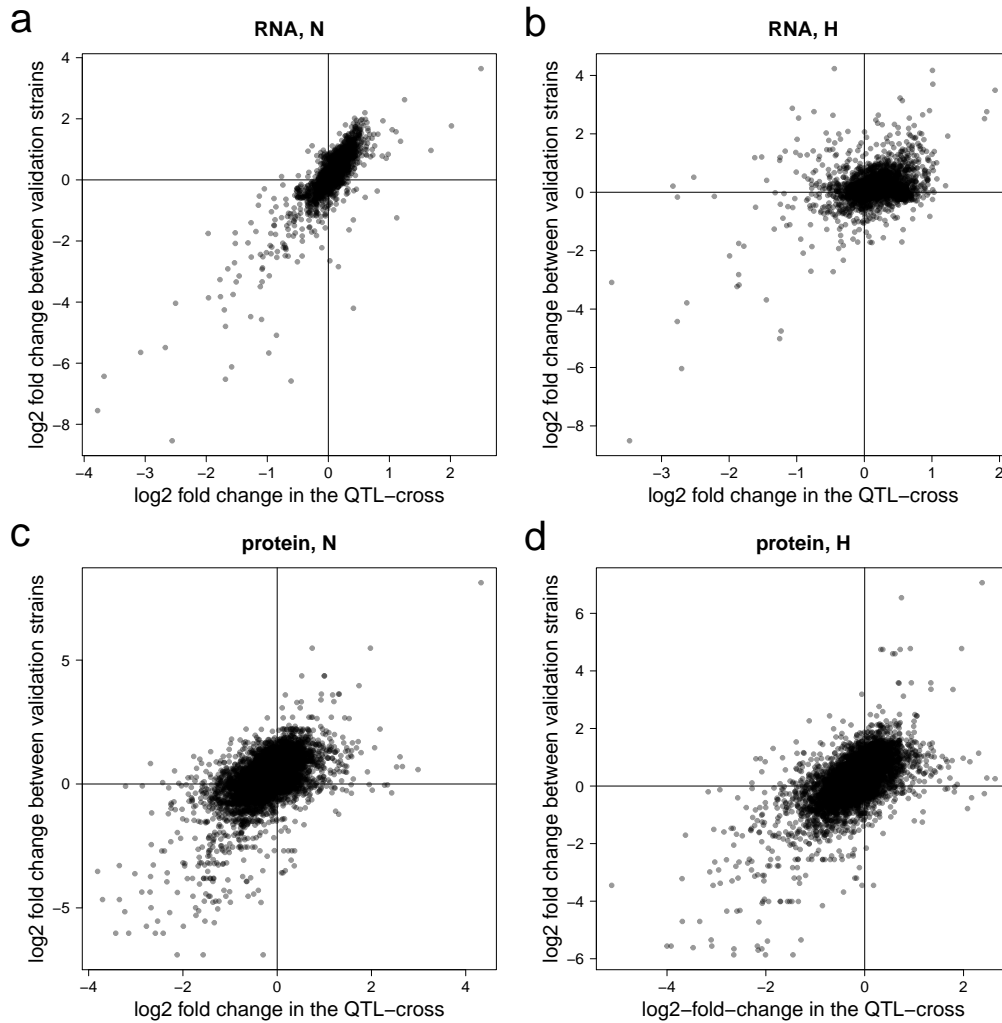
Figure 2.17: The effects of the allele replacement in JB50 are compared to the expression differences of strains differing in the *pka1*-allele in the cross in N (**a** and **c**) and H (**b** and **d**). Each dot represents a gene with available protein levels in both experiments. Plots **a** and **b** show the effects on RNA levels, while **c** and **d** show the effects on the protein levels. Proteins with multiple peptide groups correspond to multiple points in **c** and **d**.

The differences of the protein levels between JB50 and PKA1Rep correlated strongly with those observed in the cross in the normal condition (r=0.59, Figure 2.17c) and the stress-condition (r=0.64, Figure 2.17d). We were able to replicate the relationship between conditional effects of the hotspot in the normal condition and the stress response in the independent data generated with JB50 and PKA1Rep. 88% of coding genes that were targeted by e-chrII:1 in the normal condition only were also differentially expressed between JB50 and PKA1Rep in the normal condition at FDR< 10%. Their log2 fold changes between JB50 and PKA1Rep correlated strongly with the stress response of JB50 in the same experiment (r=0.69).

The mutation within *pka1* caused the differential expression of 86% of the target transcripts of the e-chrII:1 hotspot in the normal condition. 93% of the transcripts, that were differentially expressed between the two alleles of *pka1* in both experiments, changed their expression in the same direction. 69 out of 77 proteins that had a pQTL at the *pka1*-locus were also differentially expressed between the validation strains. The abundance of 68 out of 69 of these proteins changed in the expected direction. In the normal condition the strains with the JB760-allele at the *pka1*-locus have a lower expression of transcripts related to ribosome biogenesis and a higher expression of transcripts related to mitochondrial respiration. Here we show that a single amino acid substitution in *pka1p* causes widespread changes in gene expression, protein abundance and growth. It does so in condition-specific way and induces changes in the transcriptome in the normal condition that resemble the oxidative stress response in the strains carrying the same allele as JB760, the parental strain with the highest stress resistance.

2.4   Discussion

The aim of this study was to elucidate the genetic control of the oxidative stress response in fission yeast. This was accomplished by generating a genetically diverse population of fission yeast strains and studying the combined effects of oxidative stress and genotypic variation on a large number of molecular traits. We characterized the transcriptome of all strains and the proteome of a subset of strains on both

69

molecular layers, using RNA-seq and mass spectrometry, respectively. The growth efficiency of a subset of strains was also measured. We identified eQTL and pQTL, and classified these QTL depending on their effects on the different growth conditions (Ackermann et al., 2013). We identified multiple polymorphisms that affect the response to oxidative stress on the transcriptomic level.

At a FDR<10%, we identified eQTL for 90.7% of all transcripts that were measured and found pQTL for 11% of all proteins which were measured in our study. We improve on the number of eQTL reported for a subset of this cross and detect pQTL for the first time in fission yeast (Clement-Ziza et al., 2014). Most eQTL were found to have different effects in the normal condition and the stress-condition and were therefore classified as conditional. Thus, the perturbation of the cellular network through the application of oxidative stress has greatly increased the information that could be extracted through QTL-mapping. We report similar proportions of transcripts being affected by conditional eQTL as a similar study in budding yeast (Smith and Kruglyak, 2008). The conditional eQTL that affected transcript levels in both conditions usually had the same effect direction in both conditions, speaking to the robustness of gene regulation across conditions.

We detected much more eQTL that were only active in the stress-condition than eQTL that were only active in the normal condition. This is likely due to a combination of biological and technical reasons. Several loci that affected many targets did so in a stress-specific manner (e.g. e-chrI:1 and e-chrI:2). Conversely, we did not identify eQTL hotspots that are specific to the normal condition. As the FDR-correction was performed separately for the two datasets, the existence of these hotspots can affect the FDRs of all other QTL in a condition-specific manner. We correct for this with the *post hoc* calling of conditionality and detect static eQTL that would not have reached significance in the normal condition independently. As we only test eQTL that reached the significance threshold in at least one of the conditions for conditional effects, we might miss some eQTL that were specific to the normal condition and did not reach significance, but could have done so if their p-value was corrected together with those of eQTL called for the stress-condition. For future experiments, the FDR-correction might be performed for p-values of different datasets

simultaneously to avoid this possible bias.

Variants that are present in our panel of strains were fixed in at least one parental strain. Variants with strong negative effects on fitness likely would have been removed through negative selection. If the parental strains were exposed to negative selection primarily under conditions other than oxidative stress, variants with effects on the response to oxidative stress would be allowed to accumulate if they did not affect growth in other conditions simultaneously. Since fission yeast was heavily influenced by human activity, a reduced adaption to stressors in some parental strains seems plausible (Jeffares et al., 2015).

We detected much more eQTL than pQTL (12,729 eQTL vs 303 pQTL). Per trait that was mapped, we found around 40 times as many eQTL as pQTL. The transcriptomic and proteomic data differed in important technical aspects, making it harder to detect pQTL. The sample size for the pQTL study was around half of that of the eQTL study, decreasing the statistical power of the approach. Further, some peptides were only measured in a subset of samples. We chose to combine peptide levels to peptide groups to avoid removing meaningful variation in the trait levels. However, this also increased the multiple testing burden. An approach that combines peptide levels to levels for proteins in a more accurate and sparse way could have increased the number of significant pQTL. In addition mass spectrometry is overall more prone to technical noise (Collins and Aebersold, 2018).

Since the samples were taken an hour after the application of the oxidative stress, the cultures were likely not in steady state. The relative lack of proteins that were repressed after the induction of stress is an indication of the ongoing adaption to stress on the proteomic level. This could have led to increased variation in protein levels, making it harder to map them.

Earlier work indicated that protein levels are under stricter selection than transcript levels and are therefore more conserved (Khan et al., 2013). Genetic variants that solely affect the levels of transcripts but not protein levels, would therefore be expected to be under less strict selection than those variants that have a strong effect on protein levels. This would explain some reduction in pQTL, compared to eQTL. We detected only a small number of conditional pQTL. We observed a lack of re-

pression of proteins corresponding to transcripts that are down-regulated upon the induction of oxidative stress. The likely reason for this is, that the samples were taken one hour after the stress was first induced. As many proteins have a half-life that is much longer than one hour, only those proteins that are actively degraded would show a drastic decrease in levels (Christiano et al., 2014). As we computed FDRs for the interaction effect between the QTL and the condition, the lack of conditional pQTL that affect how the levels of proteins decrease after stress treatment could have made it more difficult to correctly classify those pQTL that increase protein levels in the stress condition as conditional. Some static pQTL might start to lose their effect on the proteome upon the induction of stress but might still be detected as significant in both conditions due their target proteins not being degraded fast enough. As we use univariate linear tests to assess the effect of the QTL in each condition, the statistical power to detect its effect might be reduced (Michaelson et al., 2010). If QTL were found not to have conditional effects because of a lack in power, they would be misclassified as static QTL. An additional test assessing the similarity of effects, as opposed to their difference, could be added to make the identification of static QTL more reliable. Weak QTL for which the difference in effects between the conditions cannot reliably determined could be classified as an additional 'unknown' class.

Many molecular traits were affected by nearby variants. Both the levels of transcripts and peptide groups were targets of these local QTL more often than expected by chance. Notably, local eQTL were less often specific to the stress-condition than expected by chance. This suggests that local eQTL often affect gene expression in fundamental ways that are affected by the response to environmental stimuli. Most local pQTL overlapped with local eQTL, suggesting that local pQTL in this cross are mostly caused by local eQTL. In agreement with previous QTL-studies in yeast, we found most eQTL and pQTL to be distant (Clement-Ziza et al., 2014; Albert et al., 2018; Foss et al., 2011; Brem et al., 2002).

We identified hotspots that affected large numbers of traits on both molecular layers. These hotspots largely overlapped between the molecular layers. Notable exceptions were the hotspots e-chrI:1 and p-chrIII:1, which were hotspots only affecting the

transcriptome and proteome, respectively. For e-chrI:1 we showed that the effects on the transcriptome were in fact transmitted to the proteome but were buffered to a large degree. Several hotspots were more important for one experimental condition than the other. Many targets of e-chrI:1 and e-chrI:2 were only affected in the stress-condition. The allele that was associated with a weaker response to oxidative stress was inherited from JB759 in both cases. JB759 was isolated from an alcoholic beverage, and might therefore be adapted to a niche with abundant glucose in which fast growth coupled with fermentation is preferable. Oxidative stress would likely play a small role in that environment. Conversely, many targets of e-chrIII:1 were only affected in the normal condition. e-chrII:1, on the other hand, included eQTL of all classes. Through the reduction of RAS-signaling and the induction of aerobic respiration, the JB760-allele of e-chrII:1 caused transcriptomic changes in the normal condition that resemble the response to oxidative stress.

Earlier work in budding yeast indicated that the proteome is regulated independently from the transcriptome to a large degree (Foss et al., 2007). We observed eQTL that effect the levels of proteins less than expected (e-chrI:1), eQTL that affect protein levels as expected (e-chrII:1) and pQTL that are not caused by eQTL (p-chrIII:1). This suggests that the transmission of eQTL effects to the proteome is not the same for all causal variants.

The eQTL- and pQTL-hotspot with the most targets was located on Chromosome II (e-chrII:1, p-chrII:1). We identified and validated the causal variant within *pka1*, a protein kinase central to RAS-signaling (Maeda et al., 1994). The variant causes a missense mutation close to a phosphorylation site that controls the activity of *pka1* through its localization (Gupta et al., 2011). A possible mechanism through which this variant might affect the function of *pka1* is by affecting the accessibility of this site for kinases. We quantified the proteome of JB50, the lab strain, and PKA1Rep, the strain in which the *pka1*-allele was replaced. Many of the proteins that were significantly differentially expressed between them also were affected by the hotspot in the QTL-panel but were mostly not called as significant pQTL. This indicates that our QTL-study was underpowered to detect these associations. Notably, the JB760-allele of *pka1* conferred a positive effect on growth efficiency. As it also up-regulated

aerobic respiration, cultures could grow to a higher density by metabolizing available glucose by respiration as opposed to aerobic fermentation.

In summary, we identify three major eQTL-hotspots that affect the expression of stress response genes. Notably the JB759 parental strain, which is more susceptible to oxidative stress, carries the allele for all three loci that is associated with a reduced induction of stress response genes in the respective condition. Further, we show that these hotspots also affect the proteome, although the response on the proteomic layer is weakened in the case of e-chrI:1. The same parental strain (JB759) carries those alleles that reduce growth efficiency at all gQTL. JB760, the parental strain that is most resistant to oxidative stress, carries all alleles at the eQTL-hotspots with an effect on the stress response that increase the induction of stress response genes. JB760 also carries the allele associated with higher growth density at three out of four gQTL. We identified and validated a mutation within a key player of RAS-signaling, *pka1*, for the e-chrII:1 hotspot. Further work is required to investigate the causal polymorphisms at other hotspots.

Our results have implications for the fission yeast community and the field of QTL-mapping. The hotspots and gQTL that we identified can explain some of the phenotypic differences between different fission yeast isolates and may help to select the right isolate for a specific research question. With this panel of fission yeast strains, we provide a resource that can be used for further studies. Through our analysis, we add to the understanding of how transcriptomic changes are transmitted to the proteome, and identify a hotspot that regulates the levels of proteins without affecting their transcript levels. We show that hotspots often affect different molecular targets in different conditions, emphasizing the need to generate omics-data in the most relevant condition. Local eQTL were often active in all conditions, as observed before (Smith and Kruglyak, 2008), suggesting that data from multiple conditions could be used to differentiate local *cis*-eQTL and local *trans*-eQTL that might occur around hotspots.

The detected QTL may also be used to investigate directed selection in the parental strains, as proposed previously (Orr, 1998; Riedel et al., 2015).

2.5 Contributions

Not all scientific work in this project was performed by me. I did not contribute to the generation the QTL-cross, the wet-lab experiments or the preprocessing of the generated data, including the generation of the genotype-matrices. I performed the combination of sample-wise measurements to strain-wise data and performed the QTL-mapping. I identified the QTL-hotspots, called conditions for eQTL and pQTL, assessed the stress response in JB50 and identified local QTL. The analysis on QTL-hotspots and their effects on the stress response was also performed by me. The identification of *pka1* as the candidate gene for e-chrII:1 was done by me, as well. Generation of the allele replacement strains, the preparation of the cultures and follow up measurements were done by others. Processing of the generated data and integration with the omics-data for the QTL-cross were done by me.

Chapter 3

Effects of genetic variation on the phosphoproteome of *Saccharomyces cerevisiae*

3.1   Introduction

Genetic polymorphisms are important modifiers of many physiological traits, such as body height and disease susceptibility. Variations of these traits are caused by alterations of the underlying molecular networks (Emilsson et al., 2018). As most cellular functions are performed by proteins, their abundances and activities are of special relevance when studying this network. Genetic variants might either act indirectly through multiple layers to affect protein levels, e.g. by first changing transcript levels, which then influences protein levels, or directly on individual layers, e.g. by directly changing protein levels through altered protein stability (van der Sijde et al., 2014). Changes in the sequence of an mRNA can affect translation rates, hence altering protein abundance, while the concentration of the transcript from the same gene stays constant (reviewed in Bali and Bebok (2015)). Genetic variants might also affect protein activities directly through post-translational modifications (PTMs) such as phosphorylation without affecting the abundance of the protein or its transcript.

Varying concentrations of a kinase can affect phosphorylation levels of its targets, which, in turn, might lead to their degradation (Henchoz et al., 1997), activation or inhibition (Ardito et al., 2017), or changes in their subcellular localization (Miller and Cross, 2002).

A comprehensive view of the effects of genetic perturbations on integrated molecular networks requires accurate quantitative measurements at different molecular layers from the same samples (Civelek and Lusis, 2014). For example, with current RNA sequencing (RNA-seq) technologies it is possible to detect eQTL for virtually all transcripts present in a cell (Albert et al., 2018). Likewise, mass spectrometry has been used to identify hundreds pQTL (Foss et al., 2011; Holdt et al., 2013; Picotti et al., 2013; Wu et al., 2013) and QTL affecting metabolic traits (mQTL; reviewed in Nagana Gowda and Djukovic (2014)). A shortcoming of many existing studies is that either only one layer was measured, typically transcripts or, if multiple layers were measured, transcripts and proteins were not isolated from the same cultures. Such issues have limited the comparison and integration of data from different molecular layers. For example, the extent to which eQTL also affect protein levels has been a matter of debate and previous studies have come to seemingly contradictory conclusions (Albert et al., 2018; Chick et al., 2016; Foss et al., 2011). Studies that quantified both the transcriptome and proteome covered a much larger part of the transcriptome and included only a small proportion of the proteome (Foss et al., 2007, 2011; Albert et al., 2014a). Another limitation of the existing body of research is the focus on the concentrations of biomolecules, i.e., transcript and protein levels, whereas PTMs of proteins, such as phosphorylation, which are key mediators of cellular signalling and modulators of enzyme activities, were largely neglected. Its essential role for cellular organization suggests that variable protein phosphorylation serves as a central 'relay' mediating the effects of genetic polymorphisms to complex cellular and organismal traits, such as cellular growth rates.

In order to investigate the effects of genetic variation on these protein traits we designed a multi-omics QTL study in recombinant offspring of a cross of two budding yeast strains with the following four key distinguishing features: (i) We quantified transcripts, proteins, and phosphorylation levels at genomic scale; (ii) by using the

77

SWATH-MS technology we could reproducibly quantify a large number of proteins consistently across virtually all samples; (iii) RNA, protein, and phosphoprotein samples were obtained from the same yeast culture, greatly facilitating the comparability of the data; and (iv) the data integration scheme that we developed for this study enabled us to investigate interdependencies of the molecular patterns between the layers.

To address these questions, we utilised a cross of two *Saccharomyces cerevisiae* strains (BY4716 and RM11-1a) that has been studied with different technologies and in different conditions (Brem et al., 2002; Foss et al., 2007; Smith and Kruglyak, 2008; Foss et al., 2011; Albert et al., 2014a,b, 2018). Budding yeast is generally well annotated, including information on gene functions, protein-complexes and phenotypes. As the causal variants are known for a number of loci in this cross, QTL that affect phosphorylation traits have the potential to be readily interpreted and analyzed. Further, budding yeast has very low levels of splicing, which simplifies the investigation in the relationship of transcript and protein levels.

We used this approach to study the effects of genetic variation on transcript levels, protein levels, protein phosphorylation simultaneously.

3.2   Methods

3.2.1   Sample preparation

All media were prepared in a single batch to limit experimental variability. The yeast strain collection was originally derived from a cross between the two parental strains: BY4716, an S288C derivative (MATα lys2Δ0), and RM11-1a (MATa ,leu2Δ0 ura3Δ0 ho::KAN) (Brem et al., 2002) . Strains were picked in random series of 16 (see supplementary information), pre-cultured in synthetic dextrose medium (S.D., containing 2% glucose), and then grown in 120 ml fresh S.D. medium at 30°C until a maximal OD600 of 0.8 (+/- 0.1). The cultures were then harvested and processed essentially as described in (Bodenmiller and Aebersold, 2010) . In short, at an OD600 of 0.8, trichloroacetic acid (TCA) was added to the culture media to a final concentration

of 6.25% and the cells were harvested by centrifugation at 1500 g for 5 min at 4 °C and washed three times with cold acetone. The final cell pellets were transferred into 2-ml Eppendorf tubes and frozen before further processing. Altogether, 180 cell pellets were collected for cultures which matched the abovementioned conditions and processed further.

3.2.2   RNA seq

Total RNA was isolated from deep frozen aliquots of yeast pellets using RiboPure RNA Purification Kit, yeast (Ambion, Germany), which includes a DNaAse treatment to eliminate contamination. RNA quality was assessed using RNA ScreenTape assay (Agilent, United States). All RNAs were of very high quality (median RIN 9.8, minimal RIN 9.1). cDNA libraries were prepared from poly(A) selected RNA applying the Illumina TruSeq protocol for mRNA using a total of 1 μg RNA per sample and 14 PCR cycles (Illumina, United States). The cDNA libraries were sequenced on a HiSeq2000 with 20 samples per lane (8 million reads per sample). The generated reads were stranded, single-end, and had a length of 100 bp.

3.2.3   Genotype calling

The BYxRM yeast cross has been widely used for QTL mapping. Microarray-based genotype information of the segregants is available (Brem et al., 2005). However, deep sequencing enables a more accurate genotyping of recombinant lines. To this end we have exploited published resequencing data of the parental strains (Bloom et al., 2013) together with the RNA-seq data generated here to infer the genotype of the segregants of the BYxRM cross using a method that was previously published (Clement-Ziza et al., 2014). Sequence variation information of the parental strains was obtained from `http://genomics-pubs.princeton.edu/YeastCross_BYxRM/`; only calls with a MQ>30 where considered, which represents 42,769 polymorphic sites. RNA-seq data were mapped to the S.cerevisiae reference genome (SaCer3) using TopHat 2 (Kim et al., 2013)) with the following options: `--min-intron-length 10 --min-segment-intron 10 --b2-very-sensitive --max-multihits 1`

`--library-type fr-secondstrand`. Read group information was added, and BAM files were sorted using Picard utilities (`http://broadinstitute.github.io/picard/`). RNA-seq data were further processed using the GATK pipeline (version 3.4-46-gbc02625) following the best practice guide (Van der Auwera et al., 2013). First reads containing exon-exon junction were split using the following options: `-T SplitNCigarReads -rf ReassignOneMappingQuality -RMQF 255 -RMQT 60 -U ALLOW_N_CIGAR_READS`, then the variants were called using the UnifiedGenotyper at the 42,769 polymorphic sites identified in the resequencing data of the parental strains (Bloom et al., 2013). Genotype calls where the GTAK genotype score was below 40 (GQ<40) or that were covered with less than 5 reads (DP<4) were considered as missing values.

For every polymorphic site between the progenitors, we compared the polymorphisms in the segregants and the parental strains to infer which allele was inherited. We further excluded polymorphisms (i) that could not be called (or correctly called) in the parental strains based on RNA-seq data, (ii) that could be called in less than 70% of the segregants, and (iii) with a lower allele frequency of less than 20%. As previously discussed in Clement-Ziza et al. (2014), genotypes called differing from the two direct flanking markers in more than one segregant (93 cases) probably denote erroneous genotype calls; the corresponding polymorphisms were excluded from the analysis. This resulted in 25,590 polymorphisms that were considered as genetic markers. Finally, missing genotype values were inferred from when the informative flanking sites showed the same segregation patterns and when they were < 20 kilo bases (kb) away, the same genotype as the flanking one was assigned to the missing value, assuming that no recombination event took place (Clement-Ziza et al., 2014). Adjacent markers with the same segregation pattern across all segregants were collapsed into one unique marker, resulting in a set of 3,593 unique genetic markers for QTL-mapping. Thus, each marker represents a genomic interval in which all polymorphisms are in full linkage disequilibrium in the cross.

3.2.4 Transcriptome quantification

In previous work, we have shown that accounting for individual genome variations for RNA-seq alignment improved gene expression quantification and deflated the number of falsely detected local eQTL. Therefore we used the strategy we had previously developed to map RNA-seq reads. It consists of generating a strain-specific genome, for each segregant, against which the corresponding reads are aligned. First, we generated both strain-specific genome sequences and strain-specific annotations from the reference genome sequence (SaCer3), the reference genome annotation and the available information about genetic polymorphisms (VCF files) previously generated with RNA-seq data. Then RNA-seq reads were aligned to the corresponding strain specific genomes using STAR (ver 2.5.0a) (Dobin and Gingeras, 2015) with the following options: `--alignIntronMin 10 --quantMode GeneCounts`. The gene specific read counts (strand specific) generated by STAR were used to quantify gene expression. RNA-seq coverage was computed by dividing the sum of the length of all reads by the sum of the length the coding regions of the quantified transcripts per sample. Raw read counts were normalized using rlog method of DESeq2(Love et al., 2013). Normalized data were further corrected for effects due to culture batches using the non-parametric empirical Bayes framework ComBat (Johnson et al., 2007). Normalized, log transformed and batch corrected read counts c were corrected for gene length as follows for gene i in sample j:

$$c'_{i,j} = log2(2^{c_{i,j}} \cdot \tfrac{1000}{l_i})$$

, where l_i is the length of the coding region of gene i, excluding intronic regions.

3.2.5 Proteomics

Protein extraction

Cell pellets were resuspended in lysis buffer containing 8 M urea, 0.1 M NH4HCO3, and 5 mM EDTA and cells were disrupted by glass bead beating (5 times 5 min at 4°C, allowing the samples to cool down between cycles). The total protein amount

from the pooled supernatants was determined by BCA Protein Assay Kit (Thermo Fisher Scientific, United States). Three milligrams of extracted yeast proteins were reduced with 5 mM TCEP at 37 °C for 30 min and alkylated with 12 mM iodoacetamide at room temperature in the dark for 30 min. The samples were then diluted with 0.1 M NH4HCO3 to a final concentration of 1 M urea and the proteins were digested with sequencing-grade porcine trypsin (Promega, United States) at a final enzyme:substrate ratio of 1:100 (w/w). Digestion was stopped by adding formic acid to a final concentration of 1%. Peptide mixtures were desalted using 3cc reverse phase cartridges (Sep-Pak tC18, Waters) and according to the following procedure: washing of column with one volume of 100% methanol, washing with one volume of 50% acetonitrile, washing with 3 volumes of 0.1% formic acid, loading acidified sample, reloading flow-through, washing column with sample with 3 volumes of 0.1% formic acid, and eluting sample with two volumes of 50% acetonitrile in 0.1% formic acid. Peptides were dried using a vacuum centrifuge and resolubilized in 100 μl of 0.1% formic acid. Retention time standard peptides (iRT-Kit, Biognosys, Switzerland) were spiked into the samples before they were analyzed by LC-MS for total protein abundances ('non-enriched samples').

Enrichment for phosphopeptides

The remaining 95 μl were supplemented with 300 μl of an overnight re-crystalized and cleared up phthalic acid solution prepared by carefully dissolving 5 g of phthalic acid in 50 ml of 80% acetonitrile before adding 1.75 ml of trifluoroacetic acid. The samples were then enriched for phosphopeptides by incubating for 1 hour under rotation with 1.25 mg of TiO2 resin (GLscience, Germany) pre-equilibrated twice with 500 μl of methanol, and twice with 500 μl of phthalic acid solution. Peptides bound to the TiO2 resin were then washed twice with 500 μl phthalic acid solution, twice with 80% acetonitrile, 0.1% formic acid, and twice with 0.1% formic acid. The phosphopeptides were eluted from the beads twice with 150 μl of 0.3 M ammonium hydroxide at pH 10.5 and immediately acidified again with trifluoroacetic acid to reach pH 2. The enriched phospho-peptides were desalted on microspin columns with the protocol described above, dried using a vacuum centrifuge, and resolubi-

lized in 10 μl of 0.1% formic acid. Again, retention time standard peptides (iRT-Kit, Biognosys, Switzerland) were spiked into the samples before they were analyzed by LC-MS for total peptide abundances ('phosphopeptides samples').

Measurements with SWATH-MS

The peptide concentration in all samples was measured on a NanoDrop at OD280 and normalized to allow injection of approx. 1 μg of material into the mass spectrometer. The samples were randomized and then either injected individually for SWATH-MS acquisition or pooled and injected in technical duplicates for shotgun acquisition (see supplementary information). The LC-MS acquisitions were performed on an AB Sciex 5600 TripleTOF coupled to a NanoLC2Dplus HPLC system. The liquid chromatographic separation and mass spectrometric acquisition parameters were essentially as described earlier (Selevsek et al., 2015). The peptide separation was performed on a 75 μm diameter PicoTip/PicoFrit emitter packed with 20 cm of Magic C18 AQ 3 resin using a 2-35% buffer B at 300 nl/min (buffer A: 2% acetonitrile, 0.1% formic acid; buffer B: 98% acetonitrile, 0.1% formic acid). For shotgun experiments, the mass spectrometer was operated with a 'top 20' method, with a 500-ms survey scan followed by a maximum of 20 MS/MS events of 150 ms each. The MS/MS selection was set for precursors exceeding 200 counts per second and charge states greater than 2. The selected precursors were then added to a dynamic exclusion list for 20 s. Ions were isolated using a quadrupole resolution of 0.7 amu and fragmented in the collision cell using the collision energy equation 0.0625 x m/z - 3.5 with a collision energy spread of 15 eV. For SWATH-MS acquisition, a 100-ms survey scan was followed by a series of 32 consecutive MS/MS events of 100 ms each with 25 amu precursor isolation with 1 amu overlap. The sequential precursor isolation window set-up was as follows: 400-425, 424-450, 449-475 ... 1174-1200 m/z. The collision energy for each window was determined based on the collision energy for a putative doubly charged ion centered in the respective window (equation: 0.0625 x m/z - 3.5) with a collision energy spread of 15 eV. All the MS data files were visually inspected and curated at this stage for low total ion chromatogram intensities, and the corresponding samples re-injected, if possible. This resulted in a final set of 179

SWATH data files for non-enriched and 179 SWATH data files for phospho-enriched samples that were used for data extraction. Similarly, 40 DDA files for non-enriched and 30 DDA files for phospho-enriched samples were selected for database searching and library generation.

LC-MS database searching

The shotgun data was searched with Sorcerer-Sequest (TurboSequest v4.0.3rev11 running on a Sage-N Sorcerer v4.0.4) and Mascot (version 2.3.0) against the SGD database (release February 2011, containing 6,750 yeast protein entries, concatenated with 6,750 corresponding 'tryptic peptide pseudo-reverse' decoy protein sequences). For the search, we allowed for semi-tryptic peptides and up to two missed cleavages per peptide. For the non-enriched samples, we used carbamidomethylation as a fixed modification on cysteine residues and oxidation as variable modification on methionine residues. For the phospho-enriched samples, we additionally allowed for phosphorylation as variable modification on serine, threonine, and tyrosine residues. The Sequest and Mascot search results were converted to pep.xml-files and then combined using iProphet (included in TPP version 4.5.2) both for the non-enriched and for the phospho-enriched samples. Both search results were filtered at 1% FDR by decoy counting at the peptide spectrum matches (PSM) level, resulting in a total of 698,652 identified spectra, 26,893 unique peptides, and 4,310 proteins for the non-enriched sample set; in the phospho-enriched sample set there were a total of 22,4551 identified spectra, 16,515 unique peptides (thereof 14,466 unique phosphopeptides), and 2,333 proteins (thereof 1,911 phosphoproteins). Those data were compiled into two spectra libraries (one 'non-enriched' and one 'phospho-enriched') using SpectraST (included in TPP 4.5.2) essentially as described earlier (Schubert et al., 2015) , including the specific splitting of the consensus spectra when MS/MS scans identifying the same peptide sequence were recorded more than 2 minutes apart, also described earlier (Schubert et al., 2015). Those 'split peptide assays' were given different protein entry names labeled `Subgroup_0_ProteinX` to `Subgroup_N_ProteinX`, respectively. The fragment ion coordinates for the peptides contained the top 6 most intense (singly or doubly charged) y or b fragment ions for each spectrum, excluding

those in the SWATH precursor isolation window for the corresponding peptide. The non-enriched assay library comprised assays for 19,473 peptides (thereof 18,074 proteotypic peptides matching a total of 3,119 unique proteins). The phospho-enriched assay library comprised assays for 14,339 peptides (thereof 12,969 phosphopeptide sequences) or assays for 13,786 proteotypic peptides (thereof 12,678 proteotypic phosphopeptides, matching a total of 1,676 unique phosphoproteins).

The SWATH-MS data extraction was performed using the iPortal workflow manager (Kunszt et al., 2015) calling OpenSWATH (openMS v. 1.10) (Röst et al., 2014) and pyProphet (Teleman et al., 2015). The precursors were then re-aligned across runs using TRIC (Röst et al., 2017). The two resulting SWATH identification result files contained a total of 18,273 identified peptides (thereof 16,922 proteotypic peptides matching a total of 2,940 proteins) for the non-enriched datasets; in the phospho-enriched datasets there were 13,748 identified peptides (thereof 12,412 phosphopeptides) or 13,218 proteotypic peptides (thereof 12,139 proteotypic phosphopeptides matching a total of 2,247 unique phosphoproteins). After alignment, we used a set of in-house scripts to compare the chromatographic elution profiles of the various isobaric phosphopeptide isoforms matching a same delocalized peptide form (peptide sequence + number of phosphorylations) within each single run and to eventually group those co-eluting phosphopeptide assays into the proper corresponding number of phospho-peak clusters (labeled `_cluster0` to `_clusterN`, respectively). The phospho-peak clusters were then consistently re-numbered across runs and those were used as input to mapDIA (Teo et al., 2015) to select for the best suitable transitions and peptides for quantification. This resulted in the final peptide and protein quantification matrices for the non-enriched and phospho-enriched datasets that were used for further processing.

Processing of peptide levels to protein level data

First, features detected after 7000 s, corresponding to decoys, reverse proteins, or not unique peptides were removed. We also removed fragments with oxidized methionine and their corresponding non-oxidized fragments. Next, native retention times were converted to iRTs (Escher et al., 2012). Fragments corresponding to pep-

tides, whose sequence was existing only in the reference proteome (i.e. the BY1416 background) and not in RM11-1a background were excluded from subsequent analysis. Normalization of the fragment-level data and aggregation into peptides and protein level data was performed using mapDIA (Teo et al., 2015) with the following options: `NORMALIZATION = RT 10, MIN_CORREL = 0.3, MIN_FRAG_PER_PEP = 2, MIN_PEP_PER_PROT = 1` and a maximum of 20% missing data for each fragment. Finally, abundance data were further corrected for effects due to culture batches and proteomics measurement batches using the non-parametric empirical Bayes framework ComBat (Johnson et al., 2007).

Processing of phosphopeptide levels

As for the protein level data, only unique peptides were considered. First features that were detected after 6000 seconds and peptides corresponding to decoys or reverse proteins were removed. As for the non phospho-enriched samples, we replaced native retention time by iRT (Escher et al., 2012). We removed peptides that were polymorph between the parental strains. Using mapDIA we combined the fragment levels for each peptide to one abundance for the whole phosphopeptide with the following options: `NORMALIZATION = RT 10, MIN_CORREL = 0.3, MIN_FRAG_PER_PEP = 2`. Again we allowed for up 20% of missing values per fragment. Correction for culture batches was performed with ComBat (Johnson et al., 2007).

3.2.6 Computation of derived traits

In order (i) to separate the changes in protein abundance due to RNA changes from changes that originate on the protein level through post-transcriptional processes, and (ii) to distinguish changes in phosphopeptide abundance due to protein abundance changes from changes of relative phosphorylation levels, we have generated regression-traits, as already proposed in (Foss et al., 2011). This approach was used to generate regression-traits that represent protein-level variation that is not due transcript variation and to generate traits that represent variation in the abundance of phosphopeptides that is not due changes of protein levels of the same gene. First,

for each pair of corresponding traits (e.g. a phosphopeptide and the protein that it is derived from) relative abundances across sample were normalized using a modified transformation to standard score (i.e. centering and scaling). To compute mean and standard deviation for this normalization, only the values corresponding to the samples, in which measurements were available for both phosphopeptides and protein were used. Then, for each pair, normalized phosphopeptide data were regressed on protein data using a robust linear regression using a MM-estimate (Yohai, 2007), initialized by an S-estimate using Hubber's weights function and using an M-estimator as final estimate using a Tukey's biweigh function as implemented through the MM-estimation option in the rlm function of the MASS R package. The residuals of these regression were then used to as trait in the following analyses.

Traits representing variance in protein levels that is not due to variance in the transcript levels of the same gene are term pt-traits and can be regulated by ptQTL. Traits representing variance in phosphopeptide levels that is not due to variance in the protein levels of the same gene are term phRes-traits and can be regulated by phResQTL.

3.2.7   Heritability

Broad-sense heritability estimates were computed based on replicate measurements of some strains, as described elsewhere (Bloom et al., 2013). We used six different segregants with three replicates each, as well as the parents with six (BY) and eight (RM) replicates each. In short, the `lmer` function from the lme4 R package was used to create a linear mixed-effects model with the phenotype as the response and the segregant labels as random effects. The variance components $\sigma^2 G$ (the variance due to genetic effects, i.e., different segregants), and $\sigma^2 E$, (the error variance), were extracted, and broad sense-heritability was calculated as

$$H^2 = \frac{\sigma^2 G}{\sigma^2 G + \sigma^2 E}.$$

Standard errors were calculated using the delete-one jackknife procedure, as proposed previously (Bloom et al., 2013). Random distributions of heritability estimates for

each molecular layer were generated by permuting the strain labels and compared to the real distributions using a Mann-Whitney test. In order to be able to compare heritability estimates between the different molecular levels, we restricted the analysis to the set of 402 proteins where we had measurements for expression, protein, and at least one phospho-peptide.

3.2.8   QTL mapping

QTL-mapping was performed as described in the previous chapter. We mapped trait data that was averaged by strain for each molecular layer considered. We only considered those 112 strains for which data on all molecular layers was available. Therefore the same genotype was used for all molecular layers. We combined the 25,990 genetic markers to 3593 aggregated genetic markers for QTL-mapping by fusing identical neighboring markers. Within each aggregated genetic marker all genetic markers are in perfect linkage disequilibrium within our cross. We included 7 predictors representing the population structure, computed as described in the last chapter. These covariates explain more than 25% of the genotypic variance. We generated 2,171,600 permutations for the transcript data, 520,800 permutations for the pt-traits, 524,000 permutations for the protein traits, 152,800 permutations for the phRes-traits, and 464,000 permutations for the phosphopeptides.

3.2.9   Identification of QTL-hotspots

Hotspots were identified essentially as described in the previous chapter. We divided the genome into 293 bins. Each bin was 40kb, only bins at the end of chromosomes were smaller. At $\alpha = 0.01$, a bin would have to have 30 eQTL, 9 ptQTL, 13 pQTL, 4 phResQTL, or 11 phQTL. Consecutive bins that were called as hotspots for any molecular layer were combined to a single hotspot.

3.2.10   Local and distant QTL

Local QTL were identified as described in the last chapter.

3.2.11 Functional enrichment analysis

GO-enrichment was performed as described in the last chapter, usign the topGO package (Alexa, 2007). For this study we used annotations form SacCer3.

3.2.12 Integration of growth QTL

QTL affecting growth under various conditions were taken from (Bloom et al., 2013). If the reported peak position of the growth QTL was within 50 kb of the middle of the positions of the loci that affected the most traits on each of the molecular layers, we considered them to be the same QTL.

3.2.13 Enrichment of targets of kinases and phosphatases among *HAP1*-targets

We tested the phosphoproteins targeted by the HAP1-locus for enrichments in the previously annotated targets of a large number of kinases and phosphatases (Bodenmiller et al., 2010). Here we only considered target proteins that were reported to be phosphorylated or dephosphorylated at serine residues. We also considered proteins that were measured in (Bodenmiller et al., 2010) but not found to be regulated by any genetic perturbation. Among the 315 proteins that were measured in both studies, a total of 45 proteins had at least one phResQTL at the HAP1-locus and were measured in (Bodenmiller et al., 2010). We used one sided Fisher's exact tests to assess the significance of the enrichments of targets of a kinase or phosphatase among the genes that were also targeted by the HAP1-locus.

3.2.14 Annotation of protein complexes

We used previously published annotations for protein complexes (`ftp://ftp.ebi.ac.uk/pub/databases/intact/complex/current/complextab/saccharomyces_cerevisiae.tsv`, Meldal et al. (2015)). We extracted the participating molecules per complex. Non protein members where not considered. We also removed all complexes with

only one protein member. To investigate the influence of complex stoichiometry on the protein levels, we computed the correlation between levels of proteins in the same complex. Here we only considered genes for (i) which protein and transcript levels were available, and (ii) which were annotated for complexes that did not have any overlap with other complexes. We excluded nucleolar and ribosomal proteins, as genes of these functions were observed to be buffered (Results) which might bias the results from the analysis of complex protein pairs. In total we considered 286 unique pairs of proteins from the same complex. These pairs corresponded to 188 genes. Some genes were part of more than one pair. For the same set of genes, we investigated if QTL were shared between pairs of genes coding for proteins that either share a complex or are members of different complexes, by checking if there was any overlap between the QTL of both genes.

3.2.15   Polymorphisms in and around genes

To investigate the cause of local QTL, we counted polymorphisms in the upstream region, downstream region, 3' and 5' UTRs, coding sequence and amino acid sequence of each coding gene between the BY and RM genomes. We considered all SNPs reported by (Bloom et al., 2013). We excluded all genes with insertions and deletions from this analysis. UTR-annotations were downloaded from `www.yeastgenome.org` in October 2017. If the UTR was reported multiple times with differing lengths, we used the largest annotation. The up- and downstream regions of a gene spanned 2kb each and began at the outer borders of the UTRs. The number of coding polymorphisms was defined as the number of amino acid changes in a protein between BY and RM. Multiple SNPs in the same codon were only counted once. Figure 3.7 shows only genes with available UTR annotations and available measurements on the protein level. Genes with indels are not included. For each phosphosite we identified the closest polymorphism in the amino acid sequence space by computing the absolute difference of the position of the modified serine and all polymorphic amino acids within the protein and using the minimum of those distances. Note that phosphopeptides were only included in this study if they did not have any

polymorphisms.

3.3   Results

3.3.1   Available data

The BYxRM *Saccharomyces cerevisiae* yeast segregant panel used in this study results from a cross of a wild isolate (RM11-1a) and a laboratory strain (BY4716) isogenic to the reference strain S288C. This cross has previously been used to study the genetic contribution to molecular traits - including RNA and protein levels - and to test novel systems genetics approaches (van der Sijde et al., 2014). We grew the two parental haploid yeast strains and 110 of their recombinant offspring under tightly controlled conditions, with multiple replicates for some of the strains (Figure 3.1). Each of the 170 cultures underwent transcriptome sequencing at high coverage (38-186x) allowing for the quantification of 5,429 transcripts. The RNA-seq data was also used to infer the genotypes of the strains, using a strategy that was previously developed (Clement-Ziza et al., 2014), which significantly improved the localization of recombination sites compared to previous studies (Brem et al., 2005). In samples from the same yeast cultures, we used SWATH-mass spectrometry (MS) to obtain abundances of 1,862 proteins with less than 1.8% missing values across all samples (Figure 3.2). This represents a four-fold increase in the number of quantified proteins compared to previous studies in the same cross (Foss et al., 2007, 2011; Picotti et al., 2013; Albert et al., 2014a). We also quantified the phosphorylation state of the proteins by SWATH-MS and after stringent filtering we obtained abundances of 2,116 phosphopeptides from 988 proteins with less than 2.4% missing values across all samples. Only a subset of the host-proteins of the quantified phosphopeptides was measured (40.6%).
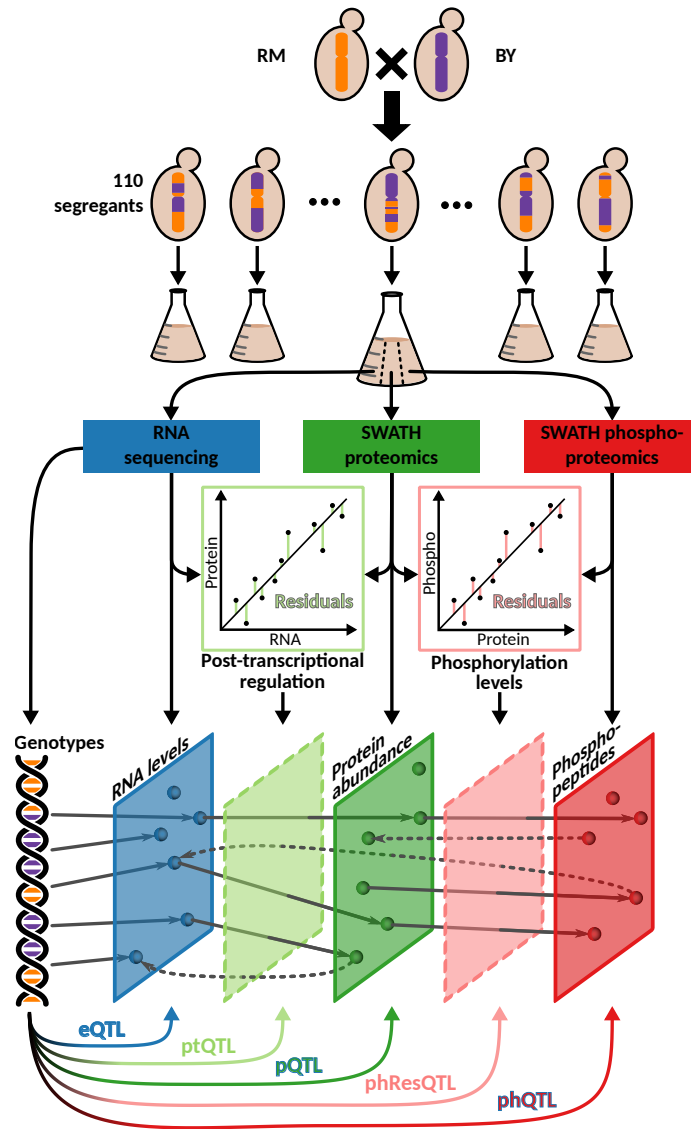
Figure 3.1: Experimental design and study overview. Yeast segregants and the parental strains they were derived from, BY and RM, were grown, and characterized with different omics approaches. Their transcriptome, proteome, and phosphoproteome was directly quantified from the same cultures. Regression traits (lighter colors) representing the disparities between transcript and protein levels (i.e., measuring post-transcriptional regulation, light green), and those between protein and phosphopeptide levels (i.e., measuring relative phosphorylation, pink) were computed. QTL analyses were performed to elucidate genetic regulation within each of these molecular layers.
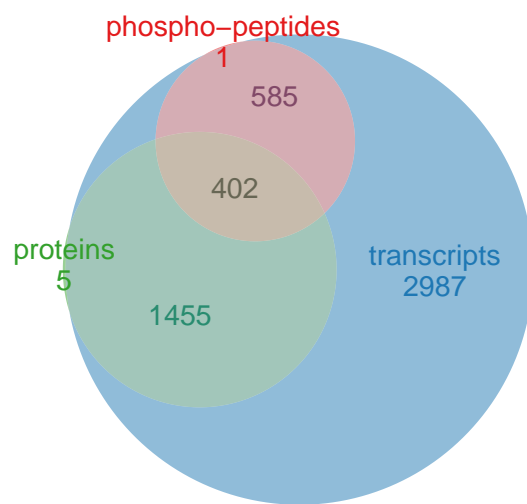
Figure 3.2: The overlap of quantified transcriptome, proteome, and phosphoproteome is shown through colored areas and additionally indicated by numbers next to the labels.

3.3.2   Derived traits

As expected, we observed that in most cases, protein abundances were positively correlated with their transcript levels (average $r = 0.23$, Figure 3.3a), and most phosphopeptides were positively correlated with their proteins of origin (average $r = 0.29$, Figure 3.3b). However, each layer can also be affected independently of the genetic effects on the other layers. We aimed to investigate genetic variation that affected protein levels and the levels of phosphopeptides directly. For this purpose we computationally generated traits that for a given gene only include variation that can not be explained by a matched molecular layer (Foss et al., 2011). First, we estimated the contribution of transcript changes to protein-level changes by regressing the concentration of a given protein against the concentration of its encoding transcript across all strains. Deviation from this regression (residuals) can either result from noise in the data or from effects on protein levels that are independent of transcript-level changes. We used the residuals as estimates of post-transcriptional regulation, resulting in 1857 post-transcriptional (pt) traits (Foss et al., 2011). Likewise, we regressed phosphopeptide levels against levels of their proteins of origin and used the residuals as estimates of differential phosphorylation, resulting in 879 phospho-residual (phRes) traits. As we could only compute phRes-traits for phosphopeptides for which the levels of the protein of origin were also available, phRes-traits were computed only for 42% of phosphopeptides. The QTL obtained by mapping these residual traits were termed post-transcriptional QTL (ptQTL) and phospho-residual QTL (phResQTL), respectively.

94
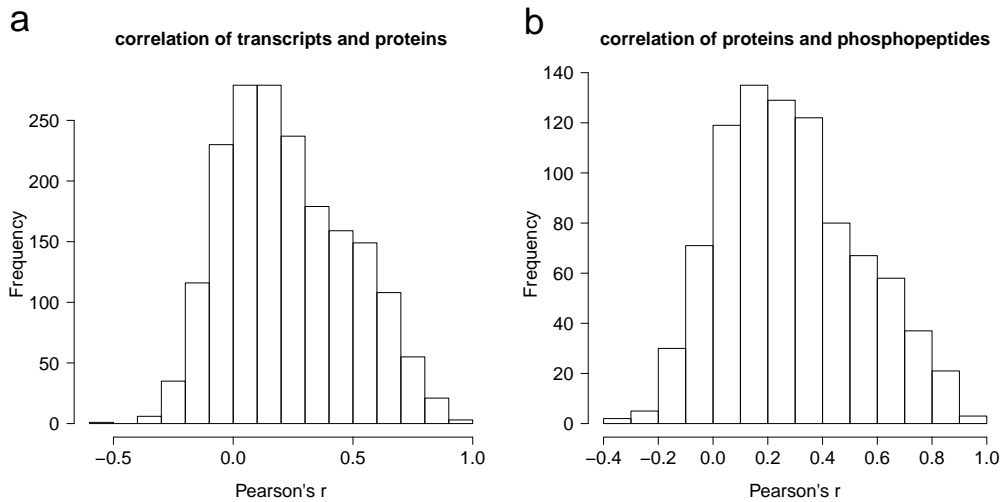
Figure 3.3: Correlation of traits for the same gene across different molecular layers. (**a**) shows the correlation of transcript and protein levels for the same gene for all genes with available measurements on both levels. (**b**) shows the correlation between phosphopeptides and their proteins of origin for all genes with available measurements on both levels. As one protein can host multiple phosphopeptides, the same gene can be shown multiple times.

3.3.3 Traits at all molecular layers are affected by genetic variation

In order to quantify the fraction of trait variation that can be attributed to genetic differences, we quantified broad sense heritability by leveraging available replicates, as proposed before (Bloom et al., 2013). Broad sense heritability was computed by comparing the variation between replicates from the same strain (which is non-genetic) to the total variation between strains. When the intra-strain variation is small compared to the inter-strain variation, one can conclude that the genetic contribution to trait variation is large under the experimental conditions tested (Bloom et al., 2013). In order to ensure comparability between the molecular layers, only the 402 genes for which measurements at each level were available are shown (Figure 3.4). The traits at all molecular layers had heritabilities greater than expected by chance (Wilcoxon rank sum test, $p < 2.2 \cdot 10^{-16}$). For genes with available data on all

95

molecular layers, transcript-levels were on average more heritable (mean $H^2 = 0; 47$) than protein (mean $H^2 = 0.33$) and phosphopeptide levels (mean $H^2 = 0.3$), and the heritability of the residual traits was on average lower than that of the directly measured traits (mean $H^2 = 0.29$ for pt-traits and mean $H^2 = 0.24$ for phRes-traits). Note that the heritabilities for some genes were likely underestimated, because only few strains had enough replicates to be used for the estimation, therefore under-representing the genetic diversity. As not all genes are regulated by the same loci, a QTL for some genes might segregate in the strains used to estimate the heritability, while the QTL for other genes might not segregate. Further, the parental strains could have been exposed to balanced selection in the past, generating allele combinations in these strains that lead to similar trait values. For those genes heritability might also be underestimated. Accordingly, for many traits that were estimated to have a broad-sense heritability close to zero, we were still able to identify significant QTL. Remarkably, 525 (28%) pt-traits and 165 (18%) phRes traits had heritabilities greater than 50%, which means that the majority of the variance of these derived traits was due to genetic variation. The lower heritability of the regressed traits compared to the directly measured layers can be partly attributed to the fact that they were derived from multiple measurements, thus adding up noise.
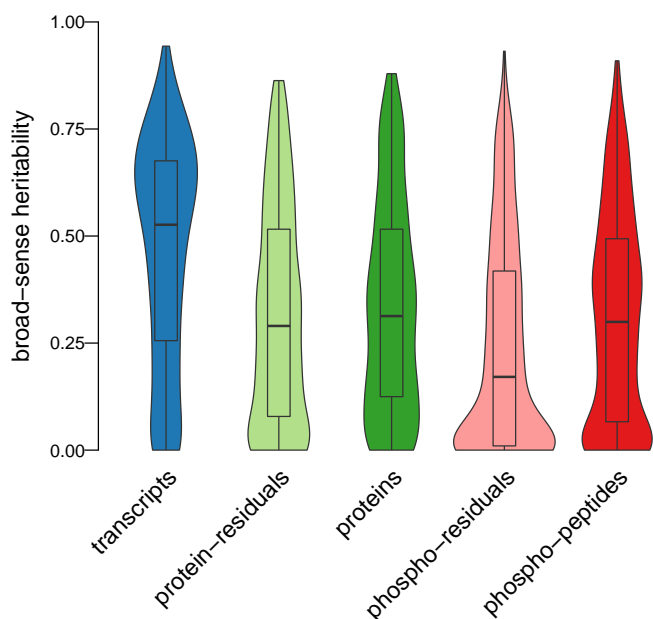
Figure 3.4: Broad-sense heritability of traits belonging to the 402 genes for which measurements at each molecular level were available. As one protein can host multiple phosphopeptides, the same gene can be shown multiple times for the phospho-layers.

3.3.4 Detection of QTL for all molecular layers

We utilized a mapping strategy based on Random Forest to identify QTL. This approach was previously shown to outperform traditional QTL mapping methods, because of its capability to account for complex (epistatic) interactions between genetic loci (see Michaelson et al. (2010) and Methods). In total, we detected 5,776 eQTL, 2,078 pQTL, and 1,327 ptQTL at FDR<10%. (Table 3.1). The same fraction of transcripts and proteins had at least one QTL (77% at FDR<10% in both cases; Figure 3.5). The high quality of the proteomics data is underlined by this large proportion of proteins with at least one pQTL (Foss et al., 2011). We also detected 1,595 phQTL and 466 phResQTL affecting 1,266 phosphopeptides (60%) and 389 phospho-residuals (44%), respectively (Table 3.1).

97
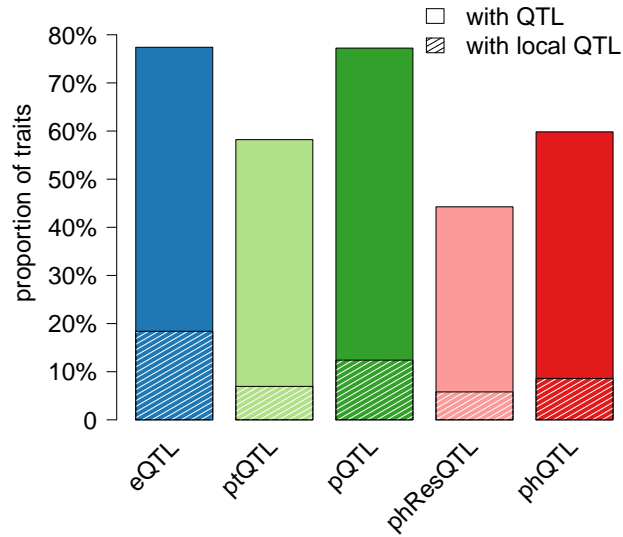
Figure 3.5: Proportion of traits affected by at least one QTL are shown separately for each molecular layer. Shaded areas indicate the proportion of traits for which a local QTL was found.

Table 3.1: Number of QTL and affected traits at each molecular layer at FDR<10%.

type	QTL	traits
eQTL	5776	4202
ptQTL	1327	1081
pQTL	2078	1438
phResQTL	466	389
phQTL	1595	1266

3.3.5 Local variation frequently affects protein levels directly

To understand where these QTL are located with respect to their target genes, we classified QTL as either local or distant based on their linkage disequilibrium with the genetic marker that is closest to the affected gene (Methods). We detected local QTL for all molecular layers, including the protein-residuals and phospho-residuals. QTL-maps for all molecular layers reveal an enrichment of local QTL as a high density of dots on the diagonal (Figure 3.6). The fraction of molecular traits with a local QTL was in a similar range for all directly measured traits (10-20%), with transcripts being most strongly enriched for local QTL (Figure 3.5). The residual-derived traits (pt and phRes) had the smallest fraction of local QTL, which may be due to biological reasons or increased noise (as discussed above) or a combination of the two. Next, we asked whether local eQTL and pQTL could be attributed to changes in the sequence of the respective transcript, which might influence transcription and translation rates. We found that genes with local QTL affecting concentrations of biomolecules (i.e. local eQTL, pQTL, ptQTL), compared to genes only affected by distant QTL, had an increased number of polymorphisms in non-coding regions (e.g. 5' untranslated regions (UTRs) or 3' UTRs; Figure 3.7). The existence of local ptQTL (129 local ptQTL for 6.9% of all pt traits), and the fact that ptQTL were enriched for polymorphisms outside coding regions, suggests that variants in non-coding parts of the genome can influence protein levels independently of their coding transcripts (Foss et al., 2011) for example through polymorphisms in ribosomal binding sites affecting translation initiation or variants altering mRNA capping and looping.
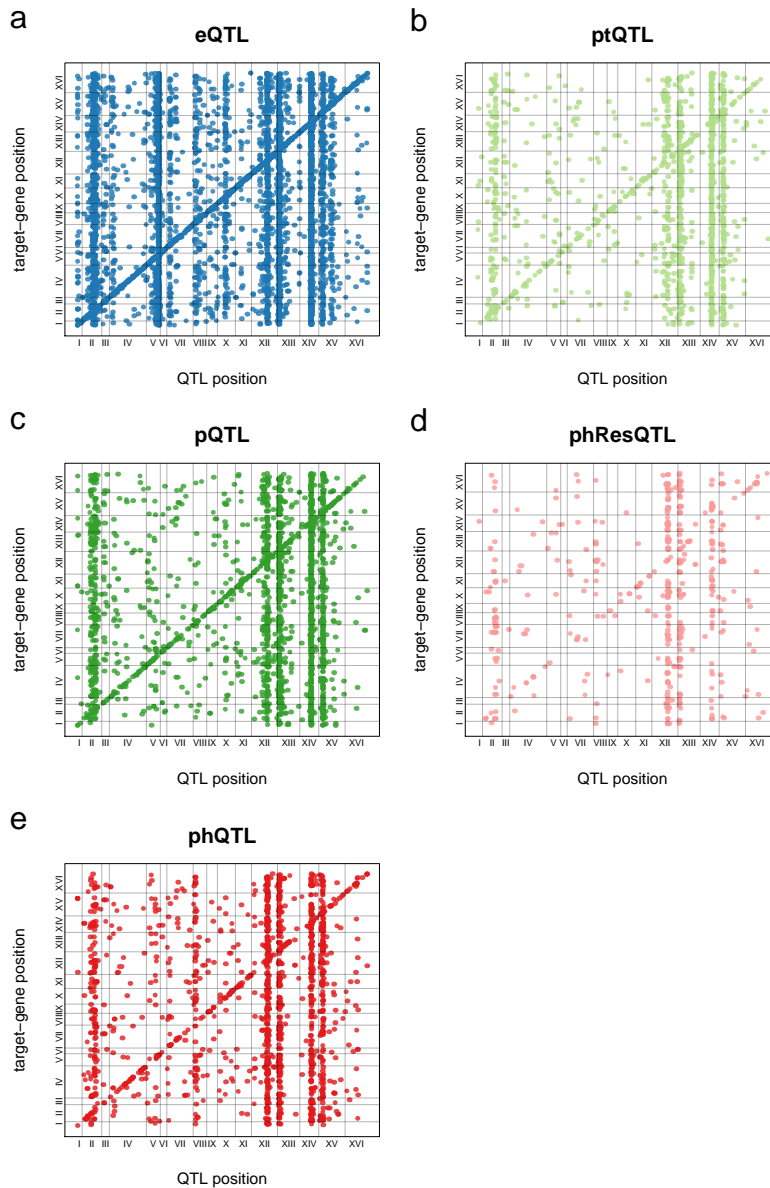
Figure 3.6: The locations of QTL and their targets are shown. Each dot represents a QTL. Its coordinates on the map represent the location of the QTL (x-axis) and the target (y-axis). In the case of phosphopeptide-traits, the location of the protein of origin is shown. Separate maps are shown for eQTL (**a**), ptQTL (**b**), pQTL (**c**), phResQTL (**d**), and phQTL (**e**). Roman numerals on the axes indicate the chromosomes and chromosomal borders are shown with solid lines.
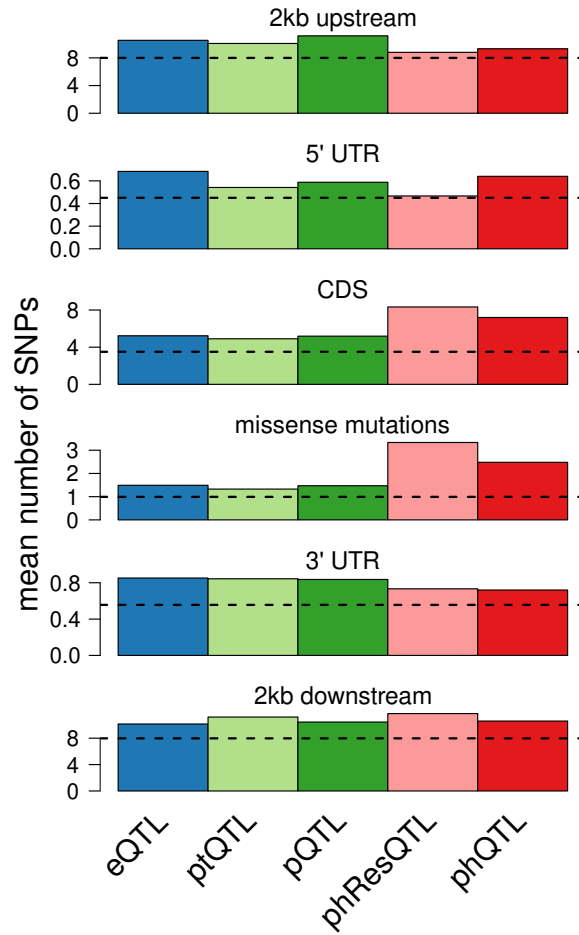
Figure 3.7: Average number of SNPs in and around genes affected by local QTL at the respective molecular layer. The horizontal dashed black line shows the average number of SNPs across all annotated genes in the respective regions. See Methods for which genes were considered.

3.3.6 Several QTL-hotspots affect all molecular layers

Regions in the genome that affect significantly more traits than expected by chance are referred to as QTL hotspots (Smith and Kruglyak, 2008). Since these loci affect large numbers of traits they are often assumed to act through master regulators such as transcription factors or kinases (Yvert et al., 2003; Albert et al., 2018). We tested for regulatory hotspots as proposed before (Brem et al., 2002) (Methods) and detected between 9 and 15 significant hotspots for each molecular layer, with the largest number of hotspots being detected for the eQTL layer (Figure 3.8 , Table 3.2). Previous work found that most distant eQTL act from within hotspots (Albert et al., 2018). Our data shows that this is the case for all five molecular layers. Many of the detected hotspots have been reported as eQTL hotspots for this yeast cross before, and for some of them a causal gene has been validated, e.g., *HAP1* (chrXII:2), *IRA2* (chrXV:1), and *MKT1* (chrXIV:1)(Brem et al., 2002; Smith and Kruglyak, 2008; Zhu et al., 2008). While most of the hotspots affected multiple molecular layers simultaneously, we also observed hotspots that predominantly impacted the transcriptome (e.g., chrV:2), the proteome (e.g., chrXII:1) or the phosphoproteome (e.g., chrVIII:1), suggesting that the causal variant influences the effect transmission between the different molecular layers.
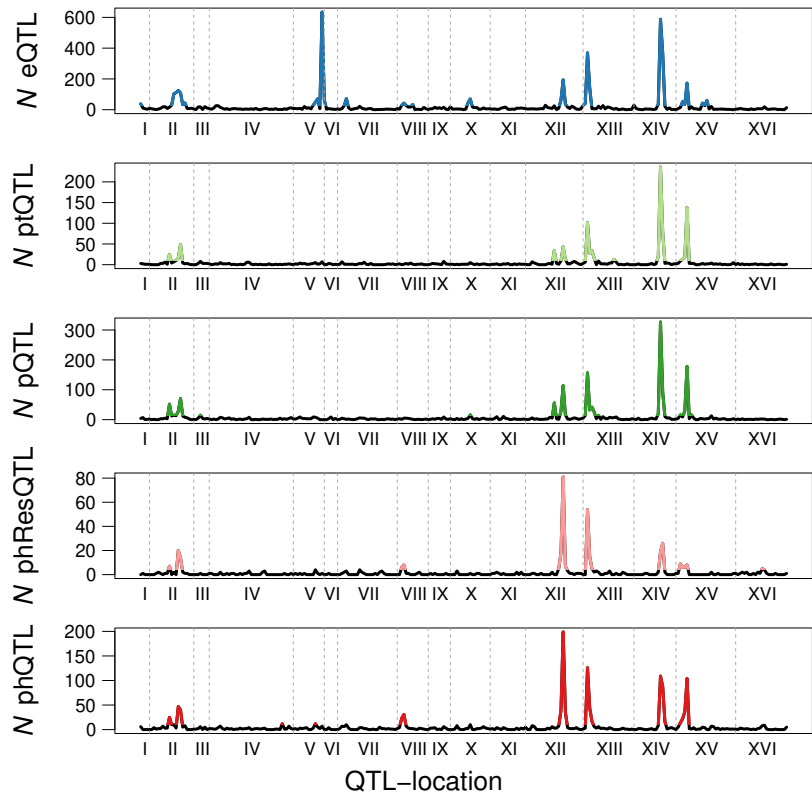
Figure 3.8: Number of traits affected by each locus at FDR<10% for each molecular layer. Regions with more QTL than expected by chance (QTL hotspots) are shown in color.

Table 3.2: QTL hotspots.

name	chr	start (bp)	end (bp)	start (bin)	end (bin)	eQTL targets	ptQTL targets	pQTL targets	phResQTL targets	phQTL targets	hotspot	causal gene	reference
chrI:1	chrI	1	40001	1	1	42	3	6	0	3	e	unknown	-
chrII:1	chrII	320001	360001	14	14	22	25	57	7	25	pt, p, phRes, ph	unknown	-
chrII:2	chrII	400001	640001	16	21	473	83	133	27	81	e, pt, p, phRes, ph	AMN1	Yvert et al. 2003
chrIII:1	chrIII	80001	120001	28	28	19	10	17	1	4	p	LEU2	Brem et al. 2002
chrIV:1	chrIV	1280001	1320001	65	65	1	0	1	0	3	ph	unknown	-
chrV:1	chrV	360001	440001	80	81	119	6	16	5	18	e, phRes, ph	unknown	-
chrV:2	chrV	480001	576874	83	84	696	7	0	1	4	e	unknown	-
chrVII:1	chrVII	120001	160001	94	94	82	1	2	3	7	e	unknown	-
chrVII:2	chrVII	360001	400001	100	100	32	5	9	3	3	phRes	unknown	-
chrVIII:1	chrVIII	40001	120001	119	120	69	1	8	12	41	e, phRes, ph	GPA1	Yvert et al. 2003
chrVIII:2	chrVIII	240001	280001	124	124	33	1	1	0	2	e	unknown	-
chrX:1	chrX	280001	360001	149	150	118	6	20	1	11	e, p	unknown	-
chrXII:1	chrXII	480001	520001	188	188	31	35	57	0	5	pt, p	unknown	-
chrXII:2	chrXII	560001	720001	190	193	289	64	155	89	216	e, pt, p, phRes, ph	HAP1	Brem et al. 2002
chrXIII:1	chrXIII	1	200001	202	206	490	161	230	58	152	e, pt, p, phRes, ph	unknown	-
chrXIII:2	chrXIII	240001	280001	208	208	27	9	16	3	5	pt, p	unknown	-
chrXIII:3	chrXIII	520001	560001	215	215	8	15	8	0	0	pt	unknown	-
chrXIII:4	chrXIII	880001	924431	224	224	31	0	2	0	1	e	unknown	-
chrXIV:1	chrXIV	400001	560001	235	238	985	319	426	37	149	e, pt, p, phRes, ph	MKT1	Zhu et al. 2008
chrXV:1	chrXV	40001	200001	245	248	271	171	225	21	133	e, pt, p, phRes, ph	IRA2	Smith and Kruglyak. 2008
chrXV:2	chrXV	240001	280001	250	250	19	3	16	0	1	p	unknown	-
chrXV:3	chrXV	440001	480001	255	255	46	3	4	0	5	e	unknown	-
chrXV:4	chrXV	520001	560001	257	257	70	3	4	0	0	e	unknown	-
chrXV:5	chrXV	600001	640001	259	259	22	5	14	2	8	p	unknown	-
chrXVI:1	chrXVI	440001	480001	282	282	9	1	2	2	6	phRes	DIG1	Smith and Kruglyak. 2008

3.3.7 The effects of eQTL on protein levels differ between functional groups of target proteins

The deep coverage of the transcriptome and proteome enabled us to investigate to what extent transcript level changes transmit to levels of corresponding/cognate proteins. First, we observed that local eQTL and local pQTL were significantly overlapping (one sided Fisher's exact test , $OR = 11$, $p < 2.2 \cdot 10^{-16}$), which is expected in the absence of major post-transcriptional or post-translational regulation. In order to estimate QTL effect sizes, we split the population based on the alleles at a linked locus and computed the log fold change of the transcript and protein levels between the two sub-populations. When comparing transcript fold changes with the respective protein fold changes, we found that they were strongly correlated (FDR<10%, r=0.73, Fig. 3b). Previous work suggested that local and distant eQTL affect the proteome in different ways (Foss et al., 2011; Chick et al., 2016). Using our data, a regression of transcript and protein fold changes had a slope close to 1 for both local and distant eQTL, implying that changes in transcript levels tend to cause similar changes in protein levels regardless of the eQTL being local or distant (Figure 3.9). This also held true when we only considered eQTL that did not overlap with ptQTL for the same gene. In addition, for most eQTL hotspots, transcript variation was propagated to protein levels, as for example at the *HAP1* locus (chrXII:2). Effects of this hotspot on protein concentrations were highly correlated with those at the transcript level (Person's correlation coefficient r=0.94 for the 289 eQTL at FDR<10%).

Despite widespread concordance between eQTL effects on transcripts and proteins, we also detected many eQTL exhibiting effects on the protein level that were different from those on the transcript level. We classified targets of eQTL into three groups based on the difference of effects at the transcript and protein levels (Methods and Figure 3.10). The first group contained genes for which the effect of an eQTL on its transcript and protein fold changes was similar. The second group contained genes for which the effects of an eQTL were repressed or even entirely buffered on the protein level. The third group contained genes for which proteins showed en-
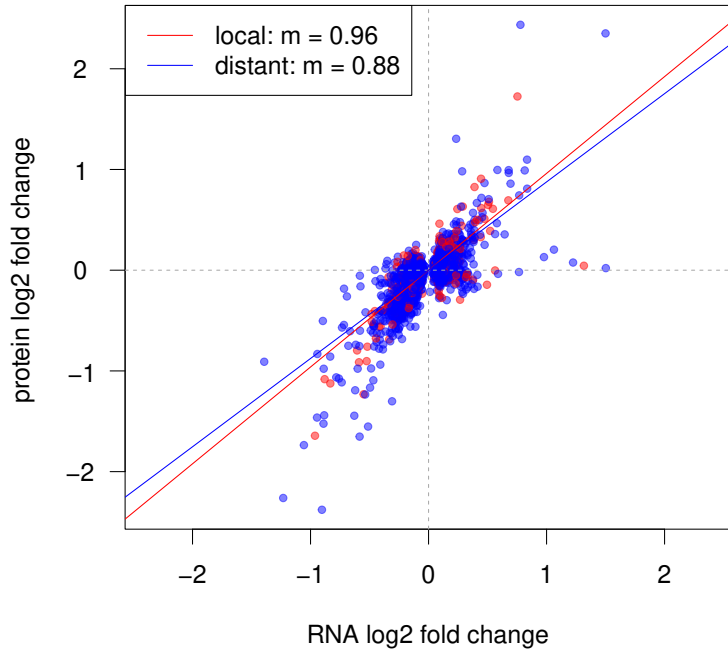
Figure 3.9: Effects of local (in red) and distant (in blue) eQTL on the proteome. The effect on transcript levels is shown on the x-axes and the effect on protein levels is shown on the y-axis. The colored lines represent regression lines for the respective set of eQTL.

hanced responses compared to their corresponding transcript. As a fourth group we added genes that were affected on their protein level by ptQTL without a significant effect on the corresponding transcript (protein only). GO-analysis revealed that genes with eQTL that affected their protein level to a similar extent as the transcript levels were enriched for genes related to protein import into the mitochondrial matrix (Supplementary Table 5.23). Genes with buffered eQTL effects were strongly enriched for terms related to cytoplasmic translation, including the large subunit of the ribosome, and proteins localizing to the nucleolus (Supplementary Table 5.24). Although buffering of ribosomal proteins has been observed before (Foss et al., 2011),

it could not be excluded that this was due to technical issues in protein quantification. However, this was unlikely the case here: first, because ribosomal proteins are relatively highly expressed, i.e. easy to quantify, and second, because these proteins were affected by other pQTL at a similar rate as the rest of the proteome (73% of proteins with buffered eQTL had at least one pQTL).
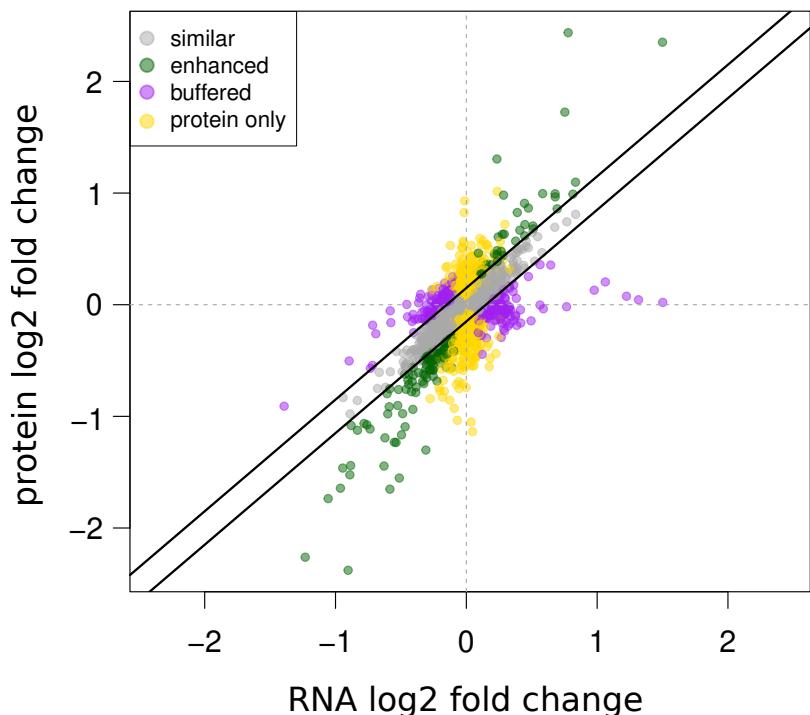
Figure 3.10: Effects of QTLs on transcript and protein levels. Each dot represents an association of a QTL with a gene. The relationship between effect size and direction on the transcript and protein levels are colorcoded based on the four effect classes described in the main text. Axes show the log2transformed fold changes of BY versus RM alleles.

Genes affected by enhanced eQTL were strongly enriched for mitochondrial ribosomes and other terms related to mitochondrial translation (Supplementary Table 5.25). Interestingly, the correlation between transcript- and protein levels was higher for mitochondrial ribosomes than for cytoplasmic ribosomes or nucleolar proteins (Figure 3.11). Like buffered proteins, proteins affected by 'protein only' effects were also enriched for functions related to cytoplasmic translation (Supplementary Table

5.26). This unexpected functional similarity between buffered proteins and proteins subject to 'protein only' effects raised the question whether the same proteins could be subject to both phenomena. Indeed, we found that genes affected by a buffered eQTL were more likely to also be affected by a 'protein only' QTL (219 genes; Fisher's exact test: $p = 7 \cdot 10^{-4}$). Thus, specific groups of proteins seem to require extensive post-transcriptional fine-tuning of their cellular concentrations, which decouples protein levels from transcript levels.
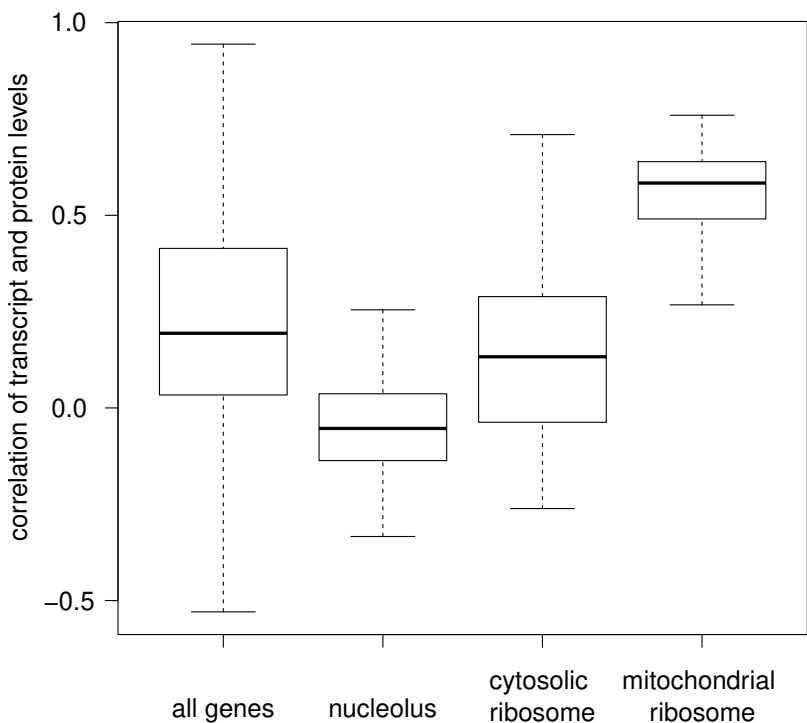
Figure 3.11: Correlations between transcript and protein levels of the same gene are shown for different groups of genes. The groups are based on GO-annotations and not exclusive. Outliers are not shown.

This notion was further supported by detailed investigation of individual hotspots with heterogeneous eQTL and pQTL effects. For example, the effects of the *IRA2* hotspot (XV:1) on transcript levels of cytoplasmic ribosomal genes was not transmitted to protein levels, but the same locus affected protein levels of genes related to mitochondrial respiration without changing their transcript levels. While the ef-

fects of the *MKT1* hotspot (XIV:1) on the protein levels of cytosolic ribosomes were buffered, the effects of the same locus on the protein levels of mitochondrial ribosomes were enhanced. Thus, our analysis revealed complex post-transcriptional QTL effects, especially for genes involved in translation, with marked differences between cytoplasmic and mitochondrial translation.

3.3.8 Protein levels are regulated through protein-complex stoichiometry

The buffering of cytoplasmic ribosomal proteins is a well-described phenomenon: excess proteins not incorporated into ribosomes get degraded (Tsay et al., 1988). Recently, buffering of components of other protein complexes has been reported (Jüschke et al., 2013; Liu et al., 2017, 2019), suggesting that the need to maintain protein complex stoichiometry drives post-transcriptional protein abundance regulation. Indeed, we observed extensive co-regulation of genes whose products participate in common complexes on multiple molecular layers: both, transcript levels and protein levels of protein complex members were correlated across the strains (average Pearson's correlation on the RNA and protein levels: $r = 0.8$ and $r = 0.46$, respectively). Furthermore, pairs of genes of the same complex were affected by the same eQTL or pQTL at a significantly higher rate than pairs from different complexes. Whereas this phenomenon was particularly strong for ribosomal proteins, we still observed significant enrichment after excluding ribosomal proteins from the analysis (eQTL: $OR = 3.45$, Fisher's exact test: $p < 7 \cdot 10^{-13}$; pQTL: $OR = 3.06$, $p < 7 \cdot 10^{-9}$). Strikingly, proteins in the same complex were even more often affected by the same ptQTL ($OR = 6.18$, $p < 2.2 \cdot 10-16$), indicating that their levels are co-regulated independently of their RNA-levels to a significant degree. Indeed we observed that most proteins from complexes correlated better with other proteins in the same complex than with their own transcript (75%), confirming that complex stoichiometry is a major driver controlling protein levels. One example for this is the Gcn1p-Gcn20p complex, which plays a role in translation initiation. Whereas both proteins were affected by a common pQTL, only *GCN20* transcripts, but not

GCN1 transcripts were affected by the respective eQTL (Figure 3.12, Garcia-Barrio (2002)). Hence, the protein level effect on Gcn1p is likely an indirect response to the Gcn20p change, thereby re-establishing stoichiometric ratios. Taken together, our data supports protein complex stoichiometry as an important post-transcriptional effector of protein abundance that substantially contributes to ptQTL effects.
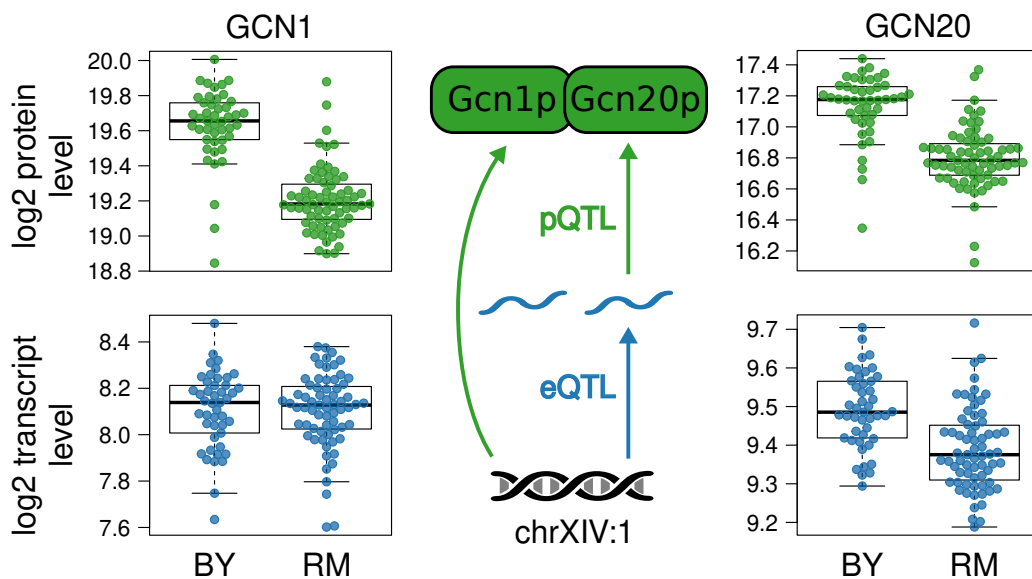
Figure 3.12: Effects of the hotspot chrXIV:1 on the Gcn1p-Gcn20p protein complex (y-axes show log2 -transformed read counts and protein abundances, respectively). Trait levels for each strain are shown separately for each allele as dots for each gene/protein. GCN20 is affected at the transcript and protein levels while the transcript levels of GCN1 did not change significantly.

3.3.9 Protein phosphorylation is often regulated separately from protein levels

To investigate effects acting directly on the phosphorylation state of proteins, we focused on the phRes traits for which the phosphorylation effects were corrected for abundance changes of the host proteins. After this correction we still detected 466 phResQTL (44% of all phRes traits), including multiple phResQTL hotspots (Figure 3.8). For example, we detected 89 phResQTL at the *HAP1* locus, affecting 22% of

all phosphoproteins but we detected QTL for only 6.5% and 2% of all transcripts and pt-traits of the 402 genes whose products were detected on all three molecular layers (Figure 3.13). phRes traits of the same protein (i.e., different phosphosites on the same protein) had a higher chance to be targeted by the same QTL than random pairs of phosphosites (one sided Fisher's exact test: $p = 3 \cdot 10^{-8}$; Figure 3.14). This is consistent with the notion that multiple phosphosites on the same protein are often targeted by a common kinase or phosphatase (Ben-Levy et al., 1995).
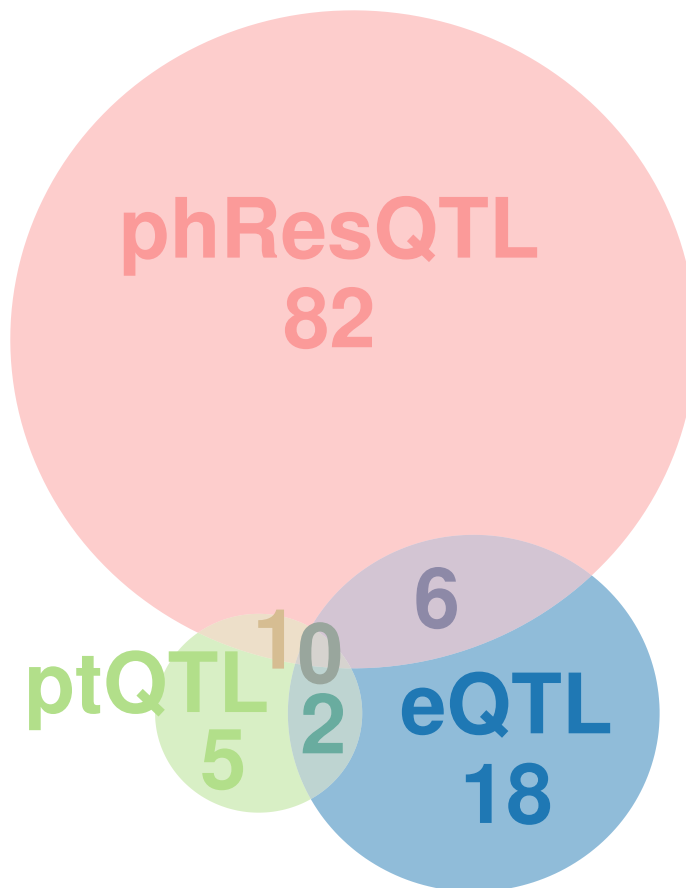
Figure 3.13: QTL targets of the *HAP1* locus (chrXII:2) on the transcript, protein residual, and phospho-residual level among the 402 genes whose products were detected on all layers are shown.
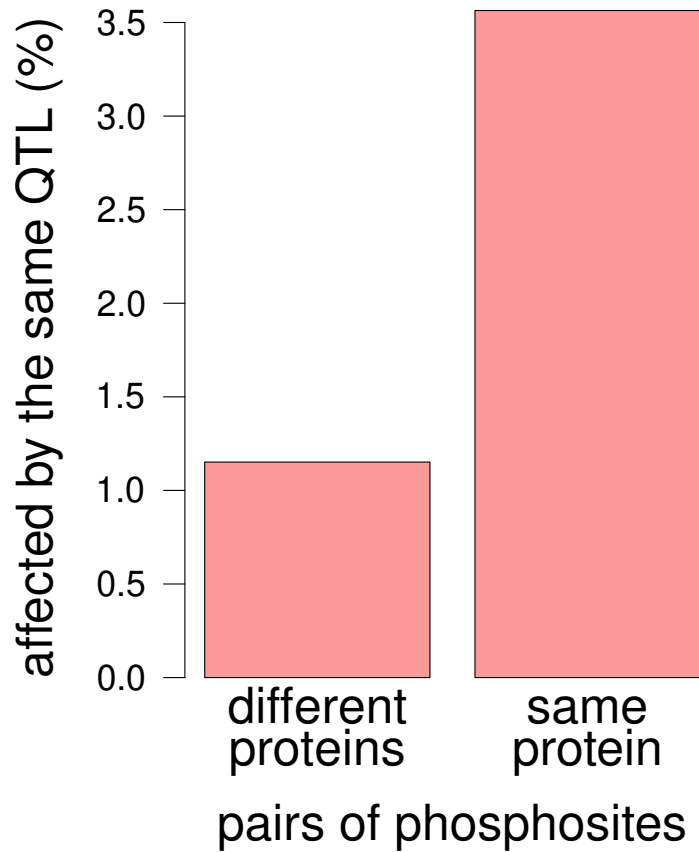
Figure 3.14: The proportion of pairs of phosphosites on different or the same protein that share a phResQTL is shown.

3.3.10 Protein sequence changes affect phosphorylation rates

Next, we asked whether local phResQTL might be caused by genetic variants directly affecting the phosphorylation state of a protein, for example through the modification of residues close to a kinase or phosphatase binding site. In support of that notion, we found that proteins that were affected by at least one local phResQTL had more missense polymorphisms than proteins that only had distant phResQTL (one-sided Wilcoxon rank sum test, $p = 1.3 \cdot 10^{-7}$). In addition, phosphosites with a local phResQTL were closer to a missense variant than those with a distant phResQTL ($p = 2.4 \cdot 10^{-5}$, Figure 3.15). Further, if two phosphosites on the same protein

were both affected by two different phResQTL - one local and the other distant - the phosphosite with the local phResQTL was on average closer to a polymorphism than the distant one (one-sided paired Wilcoxon rank sum test, $p = 0.005$, Figure 3.16). In summary, these results suggest that many local phResQTL are caused by missense polymorphisms acting by directly changing the folding of proteins or protein recognition motives that in turn affect binding of kinases and phosphatases.
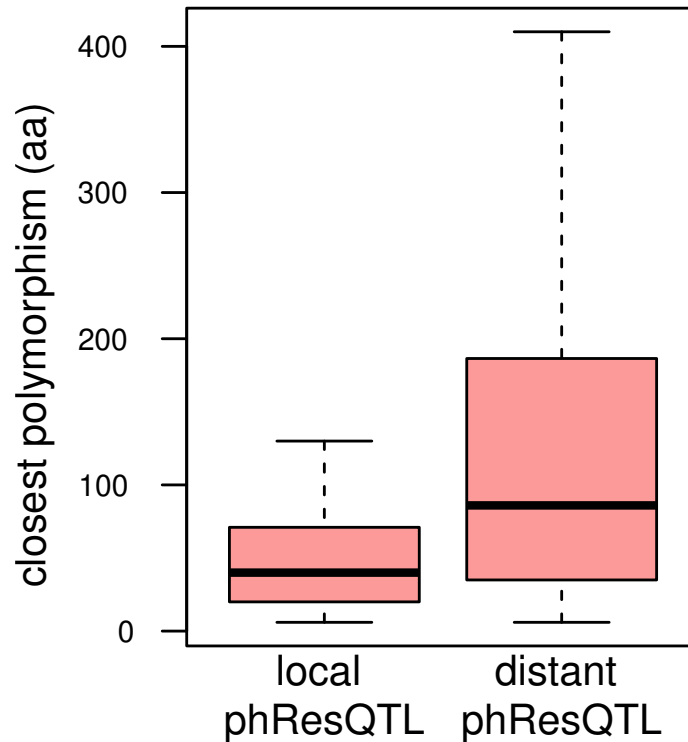
Figure 3.15: Distance of phosphosites affected by a local phResQTL or exclusively by distant phResQTL to the nearest missense mutation in the host protein are shown.
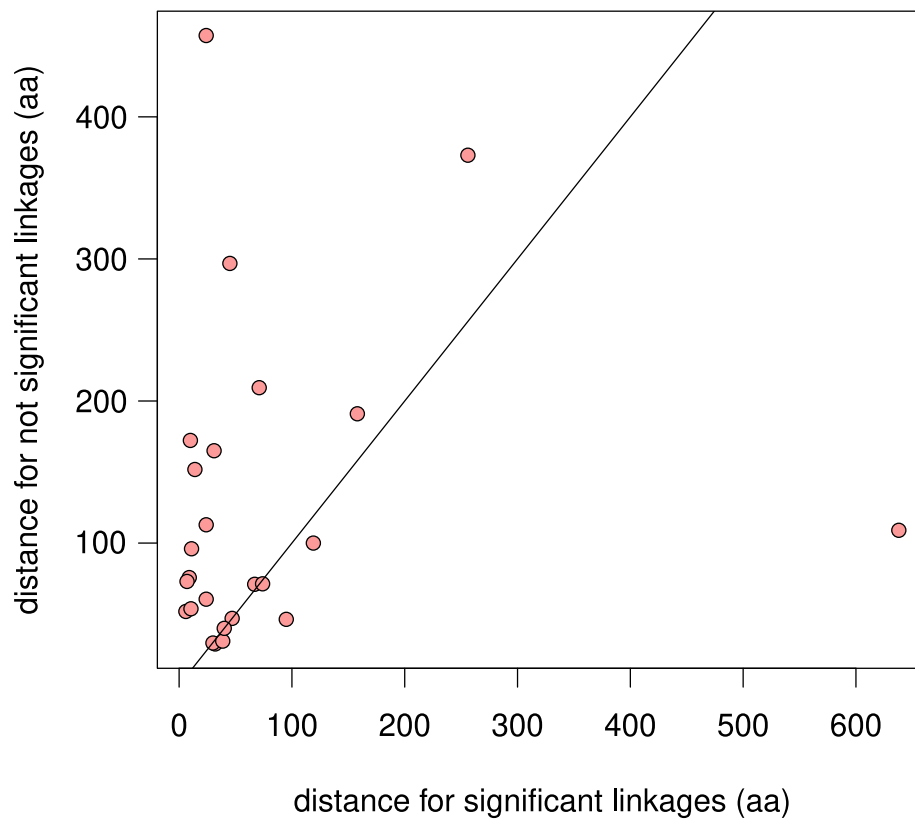
Figure 3.16: Distances to the closest missense mutations for multiple phosphosites on the same protein. For proteins with one phosphosite with a local phResQTL and a different phosphosite with a distant phResQTL the distance of both phosphosites to the closest missense mutation in the sequence space is shown. A solid black line shows the diagonal. Dots above the line represent proteins for which the phosphosite with the distant phResQTL is located further away from the closest missense mutation than the phosphosite with the local phResQTL. Dots under the line represent proteins where the reverse is true.

3.3.11 QTL effects on signaling networks and cellular fitness

In order to better understand the mechanisms of how genetic variation affects cellular fitness we integrated our molecular QTL data with 46 growth traits measured for the same yeast cross (Bloom et al., 2013). First, we found that QTL hotspots were more likely than other genomic regions to also affect growth rates (Figure 3.17a, Supplementary Table 5.27), which underlines the contribution of hotspots to complex growth trait variation. growthQTL that were close to molecular hotspots explained more variance than those growthQTL that were not (Figure 3.17b). Importantly, the number of growth traits linked to a hotspot was best predicted by the number of phospho traits targeted by the same hotspot, whereas all other molecular layers were less predictive (Figure 3.17c).

This unique role of post-translational responses to genetic variability demanded a closer inspection of state changes in regulatory networks. To study QTL effects on the 'regulome', we integrated the five molecular layers with regulatory network information. We found that while phResQTL targets were in general not enriched for particular functions, they were often functionally related to the putative causal genes of a given hotspot and/or targets of kinases that were affected by the same locus.

The *IRA2* hotspot (Smith and Kruglyak, 2008) is an example for such a case. This hotspot affected 36 physiological traits and it had targets on all five molecular layers (Figure 3.8). Ira2p inhibits the RAS/PKA pathway by promoting the GDP-bound form of Ras2p, which is crucial for the adaptation of cellular metabolism to diverse stress conditions (Tanaka et al., 1990). Earlier work showed that polymorphisms in the *IRA2* coding sequence affect the activity of the RAS/PKA pathway and that the RM-allele of *IRA2* inhibits this pathway more efficiently (Smith and Kruglyak, 2008). Several phQTL of this hotspot were functionally connected to *IRA2* including a local phQTL targeting Ira2p itself, a phQTL targeting Cdc25p, which promotes the GTP-bound form of Ras2p, and a phQTL targeting the Ras2p protein (Gross et al., 1992). In addition, we detected phQTL for numerous downstream targets of RAS/PKA signaling at the *IRA2* hotspot. This analysis implies that the differential
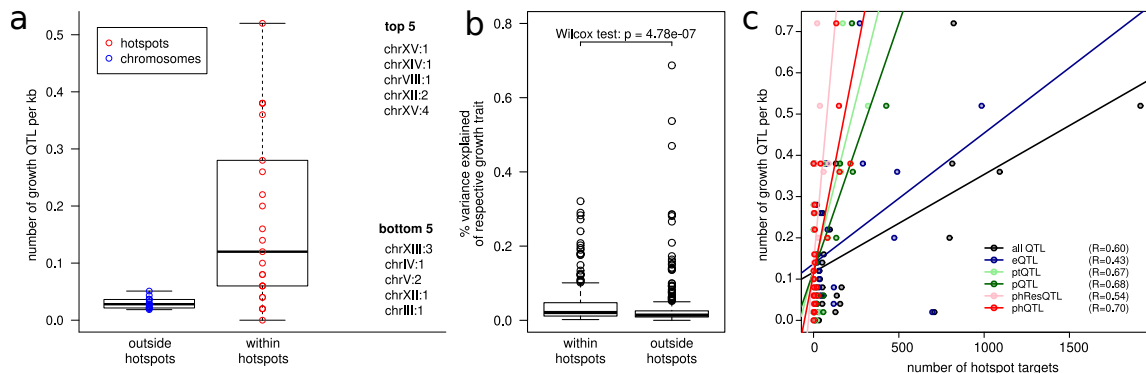
Figure 3.17: Hotspots for molecular traits have an effect on physiological pheno-
types. Growth QTL from Bloom et al. whose peak was within 50 kb of the loci
with the most targets in our hotspots were considered the same QTL. (**a**) Density
of growth QTL close to hotspots ('within hotspots') compared to the rest of the
genome ('outside hotspots'). Here, density refers to the number of growth QTL (any
growth condition) per kb. For hotspots, the points represent the individual hotspots
(red), for the rest of the genome the points represent the individual chromosomes
(blue). The hotspots with the highest and lowest growth QTL density are indicated.
(**b**) Distribution of the proportion of growth phenotype variance that was explained
by growth QTL that were close to hotspots, or not, respectively. (**c**) Relationship
between the number of targets at each layer and the number of growth QTL that
were affected by each hotspot. Linear regressions and Pearson correlation coefficients
are indicated. All correlations were significant with a corrected p-value below 0.05.

phosphorylation of Ras2p acts upstream of the effects on transcript and protein
abundances associated with this locus.

Another example illustrating the effect of genetic variants on cellular signaling is the
HAP1 locus on Chromosome 12 (chrXII:2) (Brem et al., 2002; Albert et al., 2014b),
which was shown earlier to impact on 19 physiological traits (Bloom et al., 2013).
Again, our analysis revealed that this locus had much more prominent effects on the
phospho layer than on transcript and protein abundances (Figure 3.13). Hap1p has
been shown to be involved in the regulation of respiration (Verdière et al., 1985). The
integration of the multi-layer QTL data with previously published kinase-substrate
networks suggested a new mode of effect for this hotspot: both the expression and

phosphorylation of Psk2p, a protein kinase and known regulator of carbohydrate metabolism (Rutter et al., 2002), were significantly regulated by the *HAP1* locus. Previously reported target sets of the kinase Psk2p (Bodenmiller et al., 2010) were enriched in proteins whose phosphorylation was affected by the *HAP1* locus. While the enrichment was not significant after correcting for multiple testing with the Bonferroni-method, *PSK2* was the kinase with the strongest enrichment ($OR = 2$, $p = 0.04$ before correction). This enrichment of phosphorylated Psk2p substrates strongly implies a directed signaling cascade, from the *HAP1* locus via altered Psk2p activity to the Psk2p substrates.

The examples of the *HAP1* and *IRA2* hotspots suggest that the phQTL mapping provides information on signaling networks that is orthogonal to transcript and protein abundance data: genetic variants often affect the phsophorylation states of gene products that are distinct from the genes affected by abundance changes (i.e. eQTL and pQTL targets) and those protein activity changes often act upstream of QTL abundance effects.

3.3.12 A signaling pathway is impacted by chrVIII:1

The added value of integrating biomolecule abundance data with protein phosphorylation traits was even more apparent from a hotspot on Chromosome 8 (chrVIII:1, Figure 3.8) with the pheromone response gene *GPA1* harboring the causal mutation (Yvert et al., 2003). The hotspot harbored only 69 eQTL, 8 pQTL, and 1 ptQTL, but 41 phQTL (Table 3.2). About 41% of the hotspot eQTL targets could be found downstream of the mating pheromone response pathway, which supports *GPA1* as a causal gene. However, several targets of the hotspot, especially targets on the phospho-layer such as Rck2p and Rsc9p, were involved in the osmotic stress response, which is unrelated to *GPA1*. This suggests that the polymorphisms in *GPA1* do not fully explain the effects of this hotspot and prompted us to look for other potential key effectors of the hotspot. The genetic marker with the most phQTL-targets within the hotspot lay within the coding region of *STE20*. Ste20p is the key activator of multiple mitogen-activated protein kinase (MAPK) pathways, including the mating

117

pheromone response, but also invasive growth regulation, regulation of sterol uptake, and, importantly, osmotic stress response (Figure 3.18). Indeed, for many of the target genes of this hotspot we found a link to *STE20* and/or *GPA1* (Supplementary Tables 5.28 and 5.29) and most of the phosphopeptides and phRes-traits targeted by the hotspot corresponded to proteins phosphorylated by components of MAPK pathways downstream of Ste20p (51%) (Figure 3.18). GO enrichment analysis of targets of the hotspot at the transcript layer revealed an enrichment for biological processes under the influence of *STE20* (Supplementary Table 5.30). Taken together, these results suggest that the effects of the chrVIII:1 hotspot are due to the combined effects of the polymorphisms in *GPA1* and *STE20*. Furthermore, we provide new potential components of the MAPK pathways that are regulated by this hotspot, but were not previously identified as downstream targets of Ste20p (Supplementary Table 5.31). Overall, we show that by quantifying the effects of sequence polymorphisms on multi-layer molecular networks our integrated approach can provide clues toward reconstructing the molecular architecture underlying complex traits and the chain of causality through multiple layers of regulation.
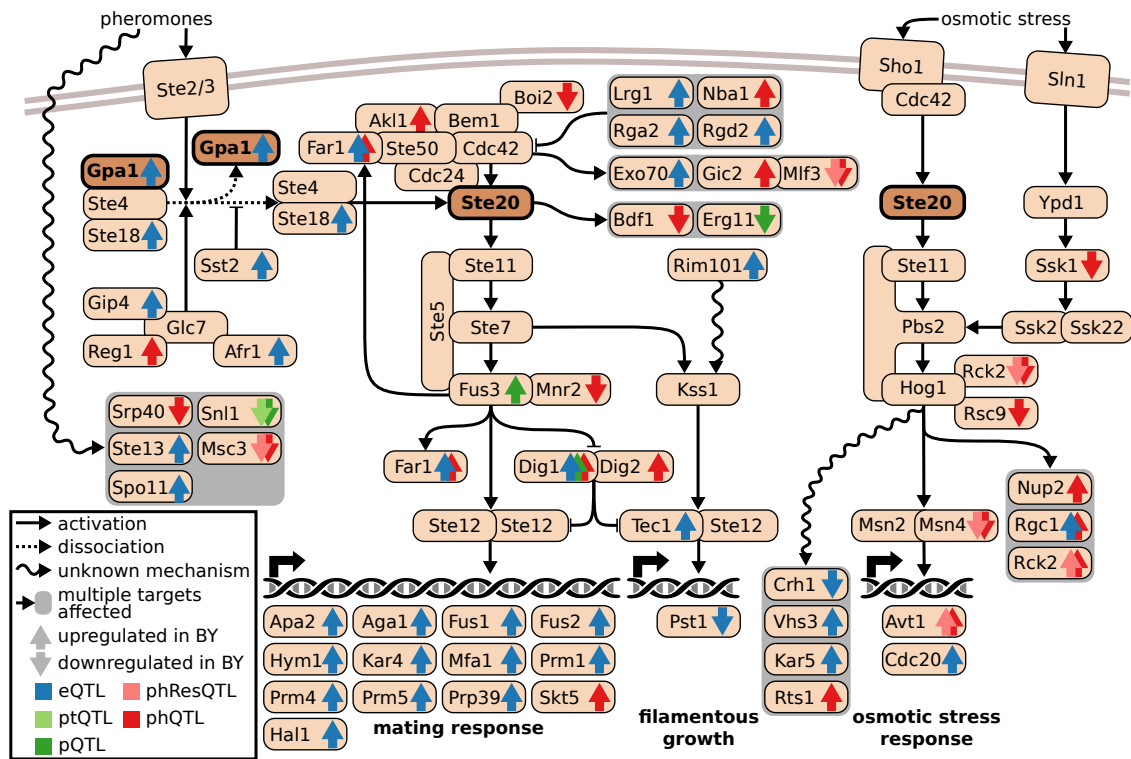
Figure 3.18: Multi-layer effects of the hotspot around *GPA1* and *STE20* on Chromosome VIII. Gpa1p and Ste20p play key roles in the mating response pathway (left, through transcription factor complex Ste12p/Ste12p) and filamentous growth regulation pathway (middle, through transcription factor complex Tec1p/Ste12p). Ste20p additionally regulates the osmotic stress response pathway (right, transcription factor Msn2p/Msn4p). Depicted are genes that are known to have a direct or indirect connection with *GPA1* and/or *STE20*. Genes that we found to be affected by the *GPA1/STE20* hotspot on any molecular layer are indicated with arrows in the respective color of each layer. The arrow orientation indicates the fold change direction (up: higher in BY, down: higher in RM).

119

3.4 Discussion

In this study, we present the first dataset that quantifies the effects of natural genomic variation concurrently on the transcript (eQTL), protein (pQTL), and protein phosphorylation (phQTL) layers. While previous 'omics' QTL studies have primarily focused on the abundances of biomolecules, our study integrates abundance with the states of biomolecules, i.e. protein phosphorylation. Hence, one conceptual advance of our study is the possibility to investigate the consequences of genomic variation comprehensively, for example the phosphorylation-mediated activation of a signaling pathway component, through the different layers of a molecular network. The three hotspots containing *IRA2*, *HAP1*, and *GPA1/STE20* exemplify this notion.

Our high-quality transcriptomic and proteomic data enabled us to study the relationship of transcript and protein levels under the influence of genetic variation in the same samples. We observed that eQTL effects were on average transmitted to the protein levels of the same gene with a 1:1 relationship (Figure 3.9), in contrast to previous observations (Foss et al., 2011). While many eQTL affected the proteome in expected ways, we also observed large numbers of eQTL that had either enhanced or buffered effects on protein levels (Figure 3.10). The targets of these eQTL with unexpected effects were often involved in translation (buffered: nucleolus, cytoplasmic ribosomes; enhanced: mitochondrial ribosomes; Supplementary Tables 5.24 and 5.25).

Earlier work suggested that local and distant eQTL affect the proteome in differing ways (Chick et al., 2016; Foss et al., 2011). We found local and distant eQTL to affect protein levels with the same strength, when the size of the effect on the transcript-levels was taken into account.

A large number of proteins was affected directly by genetic variants through ptQTL (Figure 3.6). Notably we found a large number of local ptQTL which affected protein levels of genes encoded in their vicinity, without affecting the transcript levels of the same genes (Foss et al., 2011; Albert et al., 2014b). As reported by others for a smaller part of the proteome, we identified hotspots that regulated protein levels, without having a similar effects on the respective transcript levels confirming the

120

existence of loci with highly pleoitropic effects that are not obvious on the transcript level (e.g. chrII:1 and chrXII:1, Albert et al. (2014b)). Many hotspots affected their targets in multiple modes: the same hotspots acted as buffered eQTL, enhanced eQTL, and ptQTL for different targets. This demonstrates that the transmission of the QTL-effect is highly dependent on its target. Our results show that protein levels are frequently affected by post-transcriptional regulation.

Protein complex membership has been proposed as a source of post-transcriptional regulation of protein levels (Jüschke et al., 2013). We confirm this, by showing that proteins in the same complex shared ptQTL at a more elevated rate than eQTL and that the levels of proteins in complexes agree more with other proteins in the same complex than their own transcripts. The example of the Gcn1p/Gcn20p-complex further illustrates how protein levels can be affected through complex stoichiometry. In this study, we generated genome-wide matched proteomic and phosphoproteomic data for a QTL-cross for the first time. This enabled us to correct the abundances of phosphopeptides for the abundances of their proteins of origin, producing traits that reflect relative phosphorylation. The absolute and relative phosphorylation of the phosphosites was often highly heritable. Indeed we identified phQTL and phResQTL that affected these phosphosites, demonstrating that that these traits are affected by natural genetic variation. Notably, we observed that the relative phosphorylation of many sites was affected by polymorphisms in their proteins of origin. Further, the distance of the phosphosites to the closest polymorphism was lower for phosphosites with a local phResQTL, suggesting that the physical proximity of a phosphosite to a polymorphism makes it more or less accessible for effector proteins like kinases and phosphatases. This implies that these sites are either under weak selection or selection on them differed between the parental strains.

We identified several hotspots that affected the phosphorylation of large numbers of proteins. These effects were consistent with previously reported effects of the loci (e.g. *IRA2*, *GPA1* and *HAP1*) (Brem et al., 2002; Smith and Kruglyak, 2008). For the *IRA2*-locus we were able to confirm that the *IRA2*-allele causes differences in phosphorylation in Ira2p itself, Cdc25p, and Ras2p, all of which have central roles in the RAS/PKA pathway. This suggests that the downstream effects of the *IRA2*-

121

polymorphisms are mediated through changes in phosphorylation of other pathway members. A hotspot at the *HAP1*-locus affected the highest number of phospho-targets overall. As its targets were enriched among reported substrates of Psk2p and *PSK2*-transcript levels and Psk2p-phosphorylation were affected by this locus, it is likely that the effects of this hotspot on the phosphoproteome are, at least in part, mediated through the kinase Psk2p. In the case of the hotspot near the *GPA1*-locus, our data indicates that the effects of this hotspot on the phosphorylation are not mediated solely by *GPA1* but also by *STE20*. The improved understanding of these QTL-hotspots illustrates the added value of directly measuring abundances of phosphopeptides and their proteins of origin.

Our results have implications for future work on human disease and genetic variation in the context of personalized medicine. First, we establish the feasibility of consistently measuring phosphopeptides across many samples and that a large part of the observed variance in these measurements is heritable. Second, we show that parts of the phosphoproteome, at least in yeast, are highly variable and that these differences in phosphorylation-traits can help to elucidate the effects of polymorphisms. Further, the phosphorylation changes were more predictive of cellular fitness than other molecular layers. This knowledge can be leveraged to better characterize disease states, develop biomarkers and to more accurately predict the effects of SNPs. Variants near phosphosites could be predictive of protein function. Our findings on QTL hotspots and local QTL have strong implications on QTL-studies. First, we show that local QTL are prevalent on all molecular layers that were analyzed here. Any studies that investigate the influences of genetic variants on specific gene products should give special attention to local variation, e.g. through a genotyping approach as demonstrated here. Second, we show that both the addition of proteomic and phosphoproteomic data improve our understanding of the effects of genetic variants. For each of these molecular layers we found hundreds of direct effects, that could not have been predicted through the transcriptome alone. Future studies should investigate as many molecular layers as possible, and not only the transcriptome. Our data could be used as a resource to study further topics such as the relationships between known kinases and the levels of phosphopeptides to identify new kinase-

substrate pairs. The multi-omic data across biological replicates could also be used to study the sources of biological variance in more detail, e.g. the correlation of molecular traits across replicates of the same strain. Our data on local phResQTL could inform a model that combines different data such as the distance to the closest polymorphism and the amino acid that is mutated to predict which polymorphisms will result in phResQTL. In conclusion, our study sets the stage for a truly comprehensive understanding of genomic effects on multi-layered cellular networks.

3.5 Contributions

As this was a large and diverse project, only part of the scientific work was performed by me. I did not contribute to the experimental work in the wet-lab, or the preprocessing of the molecular data and the computation of heritability. I performed the combination of sample-wise measurements to strain-wise data and performed the QTL-mapping and QTL-calling. The identification of QTL-hotspots and local QTL was done partly by me and I also contributed to the analysis of local phResQTL. I performed the analyses on local variation for other local QTL, the effects of eQTL on the proteome, the post-transcriptional effects of protein-complex membership, and shared QTL between phosphopeptides. Most of the analysis for both the *HAP1* and *IRA2* hotspots were also performed by me. I made only minor contributions to the analysis of the *STE20*-locus and the integration of hotspots and previously published growth QTL.

Chapter 4

Outlook and Conclusions

We investigated the genetic control of molecular traits in two yeast models. Despite the evolutionary separation of these species by millions of years, we were able make highly consistent observations between them. In both crosses, multiple eQTL- and pQTL-hotspots impacted on the transcript and protein levels of ribosomal proteins, underlining the sensitivity ribosome biogenesis to genetic and environmental perturbations. Notably, the effects of eQTL were often not transmitted to the protein-level for these genes. pQTL hotspots often affected the protein levels of ribosomal proteins without affecting their transcript levels. These observations demonstrate the clear necessity of measuring the levels of ribosomal proteins directly, instead of relying on transcriptomic data to draw conclusion about the proteins they encode. Relying on transcriptomic data would lead to an overestimation of the importance of some eQTL-hotspots and to an underestimation of the importance of some pQTL-hotspots. In our detailed analysis of the transmission of effects on transcript levels to the proteome, we observed that overall there is a strong relationship between the inter-strain variation on these two molecular layers. We also observed this in fission yeast where the transcriptomic and proteomic changes in response to the *pka1*-allele replacement were highly correlated. While this relationship is not perfect, as seen from the example of ribosomal proteins, it holds true for most genes.

The vast majority of known causal variants of QTL-hotspots of the crosses considered in this work are polymorphisms that impact on the amino acid sequence of proteins.

This is true for both validated causal variants in the JB50xJB759xJB760 cross (*swc5* and *pka1*) and for a large number of hotspots in the BYxRM cross (*IRA2*, *HAP1*, *GPA1/STE20*, and *MKT1*). While an investigation bias (i.e., the motivation to undertake experiments to validate a candidate) might promote the analysis of loci with well annotated proteins that carry coding polymorphisms, only few large hotspots remain in either cross where regulatory variation (i.e., variation of abundance rather than seuquence) could play a major role.

Another common feature for both studies was the observation of QTL-hotspots that were caused by variation in members of the RAS/PKA-pathway (*pka1* in fission yeast and *IRA2* in budding yeast). As RAS-signaling affects a large number of targets in a coordinated manner, genetic variants within the genes of this pathways have the potential to confer large fitness benefits when adapting to a new environment. In both studies, the integration of data from different molecular layers added information that was not obvious from just the transcriptome. For instance, the effects of the hotspot p-chrIII:1 in fission yeast and chrXII:1 for budding yeast targeted the protein layer directly. The phosphorylation data in the budding yeast study added valuable information about the biology of previously identified hotspots. For both the *IRA2*-hotspot and the *GPA1*-hotspot the phosphorylation data would have aided in the identification of the causal genes, if they were unknown at this time.

With the maturation of existing proteomics technologies and the development of new approaches, the investigation of the proteome in many samples will become even more feasible, as throughput, precision, and coverage are expected to improve (Swaminathan et al., 2018; Collins and Aebersold, 2018; Schubert et al., 2017). Emerging technologies, such as LiP-proteomics, will further add to the protein states that can be analyzed (Schopper et al., 2017). Additional states of proteins that could be assessed are oxidation levels, protein localization, complex formation and different PTMs such as ubiquitination, methylation, or acetylation. pQTL studies might, in addition, be improved through the artificial development of crosses with lower linkage disequilibrium through the use of CRISPR-based approaches (Sadhu et al., 2016). These data would readily allow the identification of causal variants and could be used to study interactions between nearby polymorphisms. These approaches could

125

be combined to study a variety of protein states under a multitude of environmental perturbations.

Bibliography

M. Ackermann, W. Sikora-Wohlfeld, and A. Beyer. Impact of Natural Genetic Variation on Gene Expression Dynamics. *PLoS Genetics*, 2013. ISSN 15537390. doi: 10.1371/journal.pgen.1003514.

E. Ahrné, T. Glatter, C. Viganò, C. Von Schubert, E. A. Nigg, and A. Schmidt. Evaluation and improvement of quantification accuracy in isobaric mass tag-based protein quantification experiments. *Journal of Proteome Research*, 2016. ISSN 15353907. doi: 10.1021/acs.jproteome.6b00066.

F. W. Albert and L. Kruglyak. The role of regulatory variation in complex traits and disease. *Nature Reviews Genetics*, 2015. ISSN 14710064. doi: 10.1038/nrg3891.

F. W. Albert, D. Muzzey, J. S. Weissman, and L. Kruglyak. Genetic Influences on Translation in Yeast. *PLoS Genetics*, 2014a. ISSN 15537404. doi: 10.1371/journal.pgen.1004692.

F. W. Albert, S. Treusch, A. H. Shockley, J. S. Bloom, and L. Kruglyak. Genetics of single-cell protein abundance variation in large yeast populations. *Nature*, 2014b. ISSN 00280836. doi: 10.1038/nature12904.

F. W. Albert, J. S. Bloom, J. Siegel, L. Day, and L. Kruglyak. Genetics of trans-regulatory variation in gene expression. *eLife*, 2018. doi: 10.7554/elife.35471.

A. L. Alejandro-Osorio, D. J. Huebert, D. T. Porcaro, M. E. Sonntag, S. Nillasithanukroh, J. L. Will, and A. P. Gasch. The histone deacetylase Rpd3p is re-

quired for transient changes in genomic expression in response to stress. *Genome Biology*, 2009. ISSN 14747596. doi: 10.1186/gb-2009-10-5-r57.

A. Alexa. Gene set enrichment analysis with topGO. 2007. ISSN 20411723 (ISSN). doi: 10.1038/ncomms8832.

L. Almasy and J. Blangero. Human QTL linkage mapping. *Genetica*, 2009. ISSN 00166707. doi: 10.1007/s10709-008-9305-3.

F. Ardito, M. Giuliani, D. Perrone, G. Troiano, and L. L. Muzio. The crucial role of protein phosphorylation in cell signalingand its use as targeted therapy (Review). *International Journal of Molecular Medicine*, 2017. ISSN 1791244X. doi: 10.3892/ijmm.2017.3036.

B. M. Babior. Phagocytes and oxidative stress. *American Journal of Medicine*, 2000. ISSN 00029343. doi: 10.1016/S0002-9343(00)00481-2.

V. Bali and Z. Bebok. Decoding mechanisms by which silent codon changes influence protein biogenesis and function. *International Journal of Biochemistry and Cell Biology*, 2015. ISSN 18785875. doi: 10.1016/j.biocel.2015.03.011.

R. Ben-Levy, I. A. Leighton, Y. N. Doza, P. Attwood, N. Morrice, C. J. Marshall, and P. Cohen. Identification of novel phosphorylation sites required for activation of MAPKAP kinase-2. *The EMBO journal*, 1995. ISSN 0261-4189.

A. Bhattacharyya, R. Chattopadhyay, S. Mitra, and S. E. Crowe. Oxidative stress: an essential factor in the pathogenesis of gastrointestinal mucosal diseases. *Physiological reviews*, 2014. ISSN 1522-1210. doi: 10.1152/physrev.00040.2012.

J. S. Bloom, I. M. Ehrenreich, W. T. Loo, T. L. V. Lite, and L. Kruglyak. Finding the sources of missing heritability in a yeast cross. *Nature*, 2013. ISSN 00280836. doi: 10.1038/nature11867.

B. Bodenmiller and R. Aebersold. Quantitative analysis of protein phosphorylation on A system-wide scale by mass spectrometry-based proteomics. *Methods in Enzymology*, 2010. ISSN 00766879. doi: 10.1016/S0076-6879(10)70013-6.

B. Bodenmiller, S. Wanka, C. Kraft, J. Urban, D. Campbell, P. G. Pedrioli, B. Gerrits, P. Picotti, H. Lam, O. Vitek, M. Y. Brusniak, B. Roschitzki, C. Zhang, K. M. Shokat, R. Schlapbach, A. Colman-Lerner, G. P. Nolan, A. I. Nesvizhskii, M. Peter, R. Loewith, C. V. Mering, and R. Aebersold. Phosphoproteomic analysis reveals interconnected system-wide responses to perturbations of kinases and phosphatases in yeast. *Science Signaling*, 2010. ISSN 19450877. doi: 10.1126/scisignal.2001182.

A. M. Bolger, M. Lohse, and B. Usadel. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, 2014. ISSN 14602059. doi: 10.1093/bioinformatics/btu170.

L. Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.

R. B. Brem, G. Yvert, R. Clinton, and L. Kruglyak. Genetic dissection of transcriptional regulation in budding yeast. *Science*, 2002. ISSN 00368075. doi: 10.1126/science.1069516.

R. B. Brem, J. D. Storey, J. Whittle, and L. Kruglyak. Genetic interactions between polymorphisms that affect gene expression in yeast. *Nature*, 2005. ISSN 00280836. doi: 10.1038/nature03865.

M. Y. Brusniak, B. Bodenmiller, D. Campbell, K. Cooke, J. Eddes, A. Garbutt, H. Lau, S. Letarte, L. N. Mueller, V. Sharma, O. Vitek, N. Zhang, R. Aebersold, and J. D. Watts. Corra: Computational framework and tools for LC-MS discovery and targeted mass spectrometry-based proteomics. *BMC Bioinformatics*, 2008. ISSN 14712105. doi: 10.1186/1471-2105-9-542.

D. Chen. Global Transcriptional Responses of Fission Yeast to Environmental Stress. *Molecular Biology of the Cell*, 2003. ISSN 10591524. doi: 10.1091/mbc.e02-08-0499.

D. Chen, C. R. M. Wilkinson, S. Watt, C. J. Penkett, W. M. Toone, N. Jones, and J. Bähler. Multiple pathways differentially regulate global oxidative stress

responses in fission yeast. *Molecular biology of the cell*, 2008. ISSN 1939-4586. doi: 10.1091/mbc.e07-08-0735.

J. M. Chick, S. C. Munger, P. Simecek, E. L. Huttlin, K. Choi, D. M. Gatti, N. Raghupathy, K. L. Svenson, G. A. Churchill, and S. P. Gygi. Defining the consequences of genetic variation on a proteome-wide scale. *Nature*, 2016. ISSN 14764687. doi: 10.1038/nature18270.

S. Chiron, M. Gaisne, E. Guillou, P. Belenguer, G. D. Clark-Walker, and N. Bonnefoy. Studying Mitochondria in an Attractive Model: Schizosaccharomyces pombe. *Methods in Molecular Biology*, pages 91–105, 1 2008. doi: 10.1007/978-1-59745-365-3_7.

M. Choi, C. Y. Chang, T. Clough, D. Broudy, T. Killeen, B. MacLean, and O. Vitek. MSstats: An R package for statistical analysis of quantitative mass spectrometry-based proteomic experiments. *Bioinformatics*, 2014. ISSN 14602059. doi: 10.1093/bioinformatics/btu305.

R. Christiano, N. Nagaraj, F. Fröhlich, and T. C. Walther. Global Proteome Turnover Analyses of the Yeasts S.cerevisiae and S.pombe. *Cell Reports*, 2014. ISSN 22111247. doi: 10.1016/j.celrep.2014.10.065.

M. Civelek and A. J. Lusis. Systems genetics approaches to understand complex traits. *Nature Reviews Genetics*, 2014. ISSN 14710056. doi: 10.1038/nrg3575.

M. Clement-Ziza, F. X. Marsellach, S. Codlin, M. A. Papadakis, S. Reinhardt, M. Rodriguez-Lopez, S. Martin, S. Marguerat, A. Schmidt, E. Lee, C. T. Workman, J. Bahler, and A. Beyer. Natural genetic variation impacts expression levels of coding, non-coding, and antisense transcripts in fission yeast. *Molecular Systems Biology*, 2014. doi: 10.15252/msb.20145123.

B. C. Collins and R. Aebersold. Proteomics goes parallel. *Nature Biotechnology*, 2018. ISSN 1087-0156. doi: 10.1038/nbt.4288.

C. Dandine-Roulland, C. Bellenguez, S. Debette, P. Amouyel, E. Génin, and H. Perdry. Accuracy of heritability estimations in presence of hidden population stratification. *Scientific reports*, 2016. ISSN 20452322. doi: 10.1038/srep26471.

A. Dobin and T. R. Gingeras. Mapping RNA-seq Reads with STAR. *Current protocols in bioinformatics*, 2015. ISSN 1934340X. doi: 10.1002/0471250953.bi1114s51.

V. Emilsson, M. Ilkov, J. R. Lamb, N. Finkel, E. F. Gudmundsson, R. Pitts, H. Hoover, V. Gudmundsdottir, S. R. Horman, T. Aspelund, L. Shu, V. Trifonov, S. Sigurdsson, A. Manolescu, J. Zhu, . Olafsson, J. Jakobsdottir, S. A. Lesley, J. To, J. Zhang, T. B. Harris, L. J. Launer, B. Zhang, G. Eiriksdottir, X. Yang, A. P. Orth, L. L. Jennings, and V. Gudnason. Co-regulatory networks of human serum proteins link genetics to disease. *Science*, 2018. ISSN 0036-8075. doi: 10.1126/science.aaq1327.

C. Escher, L. Reiter, B. Maclean, R. Ossola, F. Herzog, J. Chilton, M. J. Maccoss, and O. Rinner. Using iRT, a normalized retention time for more targeted measurement of peptides. *Proteomics*, 2012. ISSN 16159853. doi: 10.1002/pmic.201100463.

E. J. Foss, D. Radulovic, S. A. Shaffer, D. M. Ruderfer, A. Bedalov, D. R. Goodlett, and L. Kruglyak. Genetic basis of proteome variation in yeast. *Nature Genetics*, 2007. ISSN 10614036. doi: 10.1038/ng.2007.22.

E. J. Foss, D. Radulovic, S. A. Shaffer, D. R. Goodlett, L. Kruglyak, and A. Bedalov. Genetic variation shapes protein networks mainly through non-transcriptional mechanisms. *PLoS Biology*, 2011. ISSN 15449173. doi: 10.1371/journal.pbio.1001144.

M. Garcia-Barrio. Association of GCN1-GCN20 regulatory complex with the N-terminus of eIF2alpha kinase GCN2 is required for GCN2 activation. *The EMBO Journal*, 2002. doi: 10.1093/emboj/19.8.1887.

A. P. Gasch, P. T. Spellman, C. M. Kao, O. Carmel-Harel, M. B. Eisen, G. Storz, D. Botstein, and P. O. Brown. Genomic expression programs in the response of

yeast cells to environmental changes. *Molecular biology of the cell*, 2000. ISSN 1059-1524. doi: 10.1091/mbc.11.12.4241.

H. H. Göring, J. E. Curran, M. P. Johnson, T. D. Dyer, J. Charlesworth, S. A. Cole, J. B. Jowett, L. J. Abraham, D. L. Rainwater, A. G. Comuzzie, M. C. Mahaney, L. Almasy, J. W. MacCluer, A. H. Kissebah, G. R. Collier, E. K. Moses, and J. Blangero. Discovery of expression QTLs using large-scale transcriptional profiling in human lymphocytes. *Nature Genetics*, 2007. ISSN 10614036. doi: 10.1038/ng2119.

E. Gross, D. Goldberg, and A. Levitzki. Phosphorylation of the S. cerevisiae Cdc25 in response to glucose results in its dissociation from Ras. *Nature*, 1992. ISSN 00280836. doi: 10.1038/360762a0.

D. R. Gupta, S. K. Paul, Y. Oowatari, Y. Matsuo, and M. Kawamukai. Multistep regulation of protein kinase A in its localization, phosphorylation and binding with a regulatory subunit in fission yeast. *Current Genetics*, 2011. ISSN 01728083. doi: 10.1007/s00294-011-0354-2.

J. Hayles and P. Nurse. Introduction to fission yeast as a model system. *Cold Spring Harbor Protocols*, 2018. ISSN 15596095. doi: 10.1101/pdb.top079749.

S. Henchoz, Y. Chi, B. Catarin, I. Herskowitz, R. J. Deshaies, and M. Peter. Phosphorylation- and ubiquitin-dependent degradation of the cyclin- dependent kinase inhibitor Far1p in budding yeast. *Genes and Development*, 1997. ISSN 08909369. doi: 10.1101/gad.11.22.3046.

L. M. Holdt, A. von Delft, A. Nicolaou, S. Baumann, M. Kostrzewa, J. Thiery, and D. Teupser. Quantitative trait loci mapping of the mouse plasma proteome (pQTL). *Genetics*, 2013. ISSN 00166731. doi: 10.1534/genetics.112.143354.

D. C. Jeffares, C. Rallis, A. Rieux, D. Speed, M. Převorovský, T. Mourier, F. X. Marsellach, Z. Iqbal, W. Lau, T. M. Cheng, R. Pracana, M. Mülleder, J. L.

Lawson, A. Chessel, S. Bala, G. Hellenthal, B. O'Fallon, T. Keane, J. T. Simpson, L. Bischof, B. Tomiczek, D. A. Bitton, T. Sideri, S. Codlin, J. E. Hellberg, L. Van Trigt, L. Jeffery, J. J. Li, S. Atkinson, M. Thodberg, M. Febrer, K. McLay, N. Drou, W. Brown, J. Hayles, R. E. Salas, M. Ralser, N. Maniatis, D. J. Balding, F. Balloux, R. Durbin, and J. Bähler. The genomic and phenotypic diversity of Schizosaccharomyces pombe. *Nature Genetics*, 2015. ISSN 15461718. doi: 10.1038/ng.3215.

W. E. Johnson, C. Li, and A. Rabinovic. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics*, 2007. ISSN 14654644. doi: 10.1093/biostatistics/kxj037.

C. Jüschke, I. Dohnal, P. Pichler, H. Harzer, R. Swart, G. Ammerer, K. Mechtler, and J. A. Knoblich. Transcriptome and proteome quantification of a tumor model provides novel insights into post-transcriptional gene regulation. *Genome Biology*, 2013. ISSN 1474760X. doi: 10.1186/gb-2013-14-11-r133.

N. F. Käufer and J. Potashkin. Analysis of the splicing machinery in fission yeast: a comparison with budding yeast and mammals. *Nucleic acids research*, 2000. ISSN 1362-4962. doi: 10.1093/nar/28.16.3003.

A. Keller, A. I. Nesvizhskii, E. Kolker, and R. Aebersold. Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Analytical Chemistry*, 2002. ISSN 00032700. doi: 10.1021/ac025747h.

Z. Khan, M. J. Ford, D. A. Cusanovich, A. Mitrano, J. K. Pritchard, and Y. Gilad. Primate transcript and protein expression levels evolve under compensatory selection pressures. *Science*, 2013. ISSN 10959203. doi: 10.1126/science.1242379.

D. Kim, G. Pertea, C. Trapnell, H. Pimentel, R. Kelley, and S. L. Salzberg. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome biology*, 2013. ISSN 1474-760X. doi: 10.1186/gb-2013-14-4-r36.

A. J. King, L. R. Montes, J. G. Clarke, J. Itzep, C. A. Perez, R. E. Jongschaap, R. G. Visser, E. N. Van Loo, and I. A. Graham. Identification of QTL markers contributing to plant growth, oil yield and fatty acid composition in the oilseed crop Jatropha curcas L. *Biotechnology for Biofuels*, 2015. ISSN 17546834. doi: 10.1186/s13068-015-0326-8.

A. Kobayashi, T. Kanaba, R. Satoh, T. Fujiwara, Y. Ito, R. Sugiura, and M. Mishima. Structure of the second RRM domain of Nrd1, a fission yeast MAPK target RNA binding protein, and implication for its RNA recognition and regulation. *Biochemical and Biophysical Research Communications*, 2013. ISSN 0006291X. doi: 10.1016/j.bbrc.2013.06.008.

P. Kumar, S. Henikoff, and P. C. Ng. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nature protocols*, 2009. ISSN 1750-2799. doi: 10.1038/nprot.2009.86.

P. Kunszt, L. Blum, B. Hullar, E. Schmid, A. Srebniak, W. Wolski, B. Rinn, F. J. Elmer, C. Ramakrishnan, A. Quandt, and L. Malmstrom. IPortal: The swiss grid proteomics portal: Requirements and new features based on experience and usability considerations. *Concurrency Computation*, 2015. ISSN 15320634. doi: 10.1002/cpe.3294.

B. Langmead and S. L. Salzberg. Bowtie2. *Nature methods*, 2013. ISSN 1548-7091. doi: 10.1038/nmeth.1923.Fast.

B. Langmead, C. Trapnell, M. Pop, and S. L. Salzberg. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology*, 2009. ISSN 14747596. doi: 10.1186/gb-2009-10-3-r25.

A. Liaw and M. Wiener. Classification and Regression with Random Forest. *R News*, 2002. ISSN 16093631.

C. R. Lickwar, B. Rao, A. A. Shabalin, A. B. Nobel, B. D. Strahl, and J. D. Lieb. The set2/Rpd3S pathway suppresses cryptic transcription without regard to gene

length or transcription frequency. *PLoS ONE*, 2009. ISSN 19326203. doi: 10. 1371/journal.pone.0004886.

I. Liguori, G. Russo, F. Curcio, G. Bulli, L. Aran, D. Della-Morte, G. Gargiulo, G. Testa, F. Cacciatore, D. Bonaduce, and P. Abete. Oxidative stress, aging, and diseases. *Clinical Interventions in Aging*, 2018. ISSN 11781998. doi: 10.2147/CIA. S158513.

Y. Liu, A. Beyer, and R. Aebersold. On the Dependency of Cellular Protein Levels on mRNA Abundance. *Cell*, 2016. ISSN 10974172. doi: 10.1016/j.cell.2016.03.014.

Y. Liu, C. Borel, L. Li, T. Müller, E. G. Williams, P. L. Germain, M. Buljan, T. Sajic, P. J. Boersema, W. Shao, M. Faini, G. Testa, A. Beyer, S. E. Antonarakis, and R. Aebersold. Systematic proteome and proteostasis profiling in human Trisomy 21 fibroblast cells. *Nature Communications*, 2017. ISSN 20411723. doi: 10.1038/ s41467-017-01422-6.

Y. Liu, Y. Mi, T. Mueller, S. Kreibich, E. G. Williams, A. Van Drogen, C. Borel, M. Frank, P. L. Germain, I. Bludau, M. Mehnert, M. Seifert, M. Emmenlauer, I. Sorg, F. Bezrukov, F. S. Bena, H. Zhou, C. Dehio, G. Testa, J. Saez-Rodriguez, S. E. Antonarakis, W. D. Hardt, and R. Aebersold. Multi-omic measurements of heterogeneity in HeLa cells across laboratories. *Nature Biotechnology*, 2019. ISSN 15461696. doi: 10.1038/s41587-019-0037-y.

M. Love, S. Anders, and W. Huber. Differential analysis of RNA-Seq data at the gene level using the DESeq package. *European Molecular Biology Laboratory*, 2013. ISSN 1465-6906. doi: 10.1186/gb-2010-11-10-r106.

R. Lyne, G. Burns, J. Mata, C. J. Penkett, G. Rustici, D. Chen, C. Langford, D. Vetrie, and J. Bähler. Whole-genome microarrays of fission yeast: Characteristics, accuracy, reproducibility, and processing of array data. *BMC Genomics*, 2003. ISSN 14712164. doi: 10.1186/1471-2164-4-27.

M. J. MacCoss, C. C. Wu, H. Liu, R. Sadygov, and J. R. Yates. A Correlation Algorithm for the Automated Quantitative Analysis of Shotgun Proteomics Data. *Analytical Chemistry*, 2003. ISSN 00032700. doi: 10.1021/ac034790h.

M. E. MacDonald, C. M. Ambrose, M. P. Duyao, R. H. Myers, C. Lin, L. Srinidhi, G. Barnes, S. A. Taylor, M. James, N. Groot, H. MacFarlane, B. Jenkins, M. A. Anderson, N. S. Wexler, J. F. Gusella, G. P. Bates, S. Baxendale, H. Hummerich, S. Kirby, M. North, S. Youngman, R. Mott, G. Zehetner, Z. Sedlacek, A. Poustka, A. M. Frischauf, H. Lehrach, A. J. Buckler, D. Church, L. Doucette-Stamm, M. C. O'Donovan, L. Riba-Ramirez, M. Shah, V. P. Stanton, S. A. Strobel, K. M. Draths, J. L. Wales, P. Dervan, D. E. Housman, M. Altherr, R. Shiang, L. Thompson, T. Fielder, J. J. Wasmuth, D. Tagle, J. Valdes, L. Elmer, M. Allard, L. Castilla, M. Swaroop, K. Blanchard, F. S. Collins, R. Snell, T. Holloway, K. Gillespie, N. Datson, D. Shaw, and P. S. Harper. A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. *Cell*, 1993. ISSN 00928674. doi: 10.1016/0092-8674(93)90585-E.

T. Maeda, Y. Watanabe, H. Kunitomo, and M. Yamamoto. Cloning of the pka1 gene encoding the catalytic subunit of the cAMP-dependent protein kinase in Schizosaccharomyces pombe. *Journal of Biological Chemistry*, 1994. ISSN 00219258.

M. Malecki and J. Bähler. Identifying genes required for respiratory growth of fission yeast. *Wellcome Open Research*, 2016. doi: 10.12688/wellcomeopenres.9992.1.

T. A. Manolio, F. S. Collins, N. J. Cox, D. B. Goldstein, L. A. Hindorff, D. J. Hunter, M. I. McCarthy, E. M. Ramos, L. R. Cardon, A. Chakravarti, J. H. Cho, A. E. Guttmacher, A. Kong, L. Kruglyak, E. Mardis, C. N. Rotimi, M. Slatkin, D. Valle, A. S. Whittemore, M. Boehnke, A. G. Clark, E. E. Eichler, G. Gibson, J. L. Haines, T. F. C. Mackay, S. A. McCarroll, and P. M. Visscher. Finding the missing heritability of complex diseases. *Nature*, 2009. ISSN 1476-4687. doi: 10.1038/nature08494.

S. Marguerat, K. Lawler, A. Brazma, and J. Bähler. Contributions of transcription

and mRNA decay to gene expression dynamics of fission yeast in response to oxidative stress. *RNA Biology*, 2014. ISSN 15558584. doi: 10.4161/rna.29196.

B. H. Meldal, O. Forner-Martinez, M. C. Costanzo, J. Dana, J. Demeter, M. Dumousseau, S. S. Dwight, A. Gaulton, L. Licata, A. N. Melidoni, S. Ricard-Blum, B. Roechert, M. S. Skyzypek, M. Tiwari, S. Velankar, E. D. Wong, H. Hermjakob, and S. Orchard. The complex portal - An encyclopaedia of macromolecular complexes. *Nucleic Acids Research*, 2015. ISSN 13624962. doi: 10.1093/nar/gku975.

J. J. Michaelson, R. Alberts, K. Schughart, and A. Beyer. Data-driven assessment of eQTL mapping methods. *BMC Genomics*, 2010. ISSN 14712164. doi: 10.1186/1471-2164-11-502.

M. E. Miller and F. R. Cross. Mechanisms Controlling Subcellular Localization of the G1 Cyclins Cln2p and Cln3p in Budding Yeast. *Molecular and Cellular Biology*, 2002. ISSN 0270-7306. doi: 10.1128/mcb.21.18.6292-6311.2001.

G. A. Nagana Gowda and D. Djukovic. Overview of mass spectrometry-based metabolomics: Opportunities and challenges. *Methods in Molecular Biology*, 2014. ISSN 10643745. doi: 10.1007/978-1-4939-1258-2_1.

A. I. Nesvizhskii, A. Keller, E. Kolker, and R. Aebersold. A statistical model for identifying proteins by tandem mass spectrometry. *Analytical Chemistry*, 2003. ISSN 00032700. doi: 10.1021/ac0341261.

E. Nicolas, T. Yamada, H. P. Cam, P. C. FitzGerald, R. Kobayashi, and S. I. Grewal. Distinct roles of HDAC complexes in promoter silencing, antisense suppression and DNA damage protection. *Nature Structural and Molecular Biology*, 2007. ISSN 15459993. doi: 10.1038/nsmb1239.

E. S. Okerberg, A. Hainley, H. Brown, A. Aban, S. Alemayehu, A. Shih, J. Wu, M. P. Patricelli, J. W. Kozarich, T. Nomanbhoy, and J. S. Rosenblum. Identification of a tumor specific, active-site mutation in casein kinase 1α by chemical proteomics. *PLoS ONE*, 2016. ISSN 19326203. doi: 10.1371/journal.pone.0152934.

H. A. Orr. Testing natural selection vs. genetic drift in phenotypic evolution using quantitative trait locus data. *Genetics*, 1998. ISSN 00166731.

G. Pesole, F. Mignone, C. Gissi, G. Grillo, F. Licciulli, and S. Liuni. Structural and functional features of eukaryotic mRNA untranslated regions. *Gene*, 2001. ISSN 03781119. doi: 10.1016/S0378-1119(01)00674-6.

P. Picotti, M. Clément-Ziza, H. Lam, D. S. Campbell, A. Schmidt, E. W. Deutsch, H. Röst, Z. Sun, O. Rinner, L. Reiter, Q. Shen, J. J. Michaelson, A. Frei, S. Alberti, U. Kusebauch, B. Wollscheid, R. L. Moritz, A. Beyer, and R. Aebersold. A complete mass-spectrometric map of the yeast proteome applied to quantitative trait analysis. *Nature*, 2013. ISSN 00280836. doi: 10.1038/nature11835.

E. Piskounova, M. Agathocleous, M. M. Murphy, Z. Hu, S. E. Huddlestun, Z. Zhao, A. M. Leitch, T. M. Johnson, R. J. DeBerardinis, and S. J. Morrison. Oxidative stress inhibits distant metastasis by human melanoma cells. *Nature*, 2015. ISSN 14764687. doi: 10.1038/nature15726.

G. S. Plastow, D. Carrión, M. Gil, J. A. García-Regueiro, M. Font i Furnols, M. Gispert, M. A. Oliver, A. Velarde, M. D. Guàrdia, M. Hortós, M. A. Rius, C. Sárraga, I. Diíaz, A. Valero, A. Sosnicki, R. Klont, S. Dornan, J. M. Wilkinson, G. Evans, C. Sargent, G. Davey, D. Connolly, B. Houeix, C. M. Maltin, H. E. Hayes, V. Anandavijayan, A. Foury, N. Geverink, M. Cairns, R. E. Tilley, P. Mormède, and S. C. Blott. Quality pork genes and meat production. 2005. doi: 10.1016/j.meatsci.2004.06.025.

T. J. Polderman, B. Benyamin, C. A. De Leeuw, P. F. Sullivan, A. Van Bochoven, P. M. Visscher, and D. Posthuma. Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nature Genetics*, 2015. ISSN 15461718. doi: 10.1038/ng.3285.

S. M. Purcell, N. R. Wray, J. L. Stone, P. M. Visscher, M. C. O'Donovan, P. F. Sullivan, D. M. Ruderfer, A. McQuillin, D. W. Morris, C. T. Odushlaine, A. Corvin, P. A. Holmans, M. C. Odonovan, S. MacGregor, H. Gurling, D. H. Blackwood,

N. J. Craddock, M. Gill, C. M. Hultman, G. K. Kirov, P. Lichtenstein, W. J. Muir, M. J. Owen, C. N. Pato, E. M. Scolnick, D. St Clair, N. M. Williams, L. Georgieva, I. Nikolov, N. Norton, H. Williams, D. Toncheva, V. Milanova, E. F. Thelander, C. T. O'Dushlaine, E. Kenny, E. M. Quinn, K. Choudhury, S. Datta, J. Pimm, S. Thirumalai, V. Puri, R. Krasucki, J. Lawrence, D. Quested, N. Bass, C. Crombie, G. Fraser, S. Leh Kuan, N. Walker, K. A. McGhee, B. Pickard, P. Malloy, A. W. MacLean, M. Van Beck, M. T. Pato, H. Medeiros, F. Middleton, C. Carvalho, C. Morley, A. Fanous, D. Conti, J. A. Knowles, C. Paz Ferreira, A. MacEdo, M. Helena Azevedo, A. N. Kirby, M. A. Ferreira, M. J. Daly, K. Chambert, F. Kuruvilla, S. B. Gabriel, K. Ardlie, J. L. Moran, and P. Sklar. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature*, 2009. ISSN 00280836. doi: 10.1038/nature08185.

A. R. Quinlan and I. M. Hall. BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics*, 2010. ISSN 13674803. doi: 10.1093/bioinformatics/btq033.

R Development Core Team. R: A Language and Environment for Statistical Computing. *R Foundation for Statistical Computing*, 2016. ISSN 16000706. doi: 10.1007/978-3-540-74686-7.

P. Rentzsch, D. Witten, G. M. Cooper, J. Shendure, and M. Kircher. CADD: Predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Research*, 2019. ISSN 13624962. doi: 10.1093/nar/gky1016.

N. Riedel, B. S. Khatri, M. Lässig, and J. Berg. Multiple-line inference of selection on quantitative traits. *Genetics*, 2015. ISSN 19432631. doi: 10.1534/genetics.115.178988.

M. A. Rodríguez-Gabriel, G. Burns, W. H. McDonald, V. Martín, J. R. Yates, J. Bähler, and P. Russell. RNA-binding protein Csx1 mediates global control of gene expression in response to oxidative stress. *EMBO Journal*, 2003. ISSN 02614189. doi: 10.1093/emboj/cdg597.

139

M. A. Rodriguez-Gabriel, S. Watt, J. Bahler, and P. Russell. Upf1, an RNA Helicase Required for Nonsense-Mediated mRNA Decay, Modulates the Transcriptional Response to Oxidative Stress in Fission Yeast. *Molecular and Cellular Biology*, 2006. ISSN 0270-7306. doi: 10.1128/mcb.00286-06.

M. Rodríguez-López, C. Cotobal, O. Fernández-Sánchez, N. Borbarán Bravo, R. Oktriani, H. Abendroth, D. Uka, M. Hoti, J. Wang, M. Zaratiegui, and J. Bähler. A CRISPR/Cas9-based method and primer design tool for seamless genome editing in fission yeast. *Wellcome Open Research*, 2017. doi: 10.12688/wellcomeopenres. 10038.2.

H. L. Röst, G. Rosenberger, P. Navarro, L. Gillet, S. M. Miladinoviä, O. T. Schubert, W. Wolski, B. C. Collins, J. Malmström, L. Malmström, and R. Aebersold. OpenSWATH enables automated, targeted analysis of data-independent acquisition MS data. *Nature Biotechnology*, 2014. ISSN 15461696. doi: 10.1038/nbt.2841.

H. L. Röst, R. Aebersold, and O. T. Schubert. Automated swath data analysis using targeted extraction of ion chromatograms. *Methods in Molecular Biology*, 2017. ISSN 10643745. doi: 10.1007/978-1-4939-6747-6_20.

J. Rutter, B. L. Probst, and S. L. McKnight. Coordinate regulation of sugar flux and translation by PAS kinase. *Cell*, 2002. ISSN 00928674. doi: 10.1016/ S0092-8674(02)00974-1.

M. J. Sadhu, J. S. Bloom, L. Day, and L. Kruglyak. CRISPR-directed mitotic recombination enables genetic mapping without crosses. *Science*, 2016. ISSN 10959203. doi: 10.1126/science.aaf5124.

R. Satoh, A. Tanaka, A. Kita, T. Morita, Y. Matsumura, N. Umeda, M. Takada, S. Hayashi, T. Tani, K. Shinmyozu, and R. Sugiura. Role of the RNA-binding protein Nrd1 in stress granule formation and its implication in the stress response in fission yeast. *PLoS ONE*, 2012. ISSN 19326203. doi: 10.1371/journal.pone. 0029683.

S. Schopper, A. Kahraman, P. Leuenberger, Y. Feng, I. Piazza, O. Müller, P. J. Boersema, and P. Picotti. Measuring protein structural changes on a proteome-wide scale using limited proteolysis-coupled mass spectrometry. *Nature Protocols*, 2017. ISSN 17502799. doi: 10.1038/nprot.2017.100.

O. T. Schubert, L. C. Gillet, B. C. Collins, P. Navarro, G. Rosenberger, W. E. Wolski, H. Lam, D. Amodei, P. Mallick, B. Maclean, and R. Aebersold. Building high-quality assay libraries for targeted analysis of SWATH MS data. *Nature Protocols*, 2015. ISSN 17502799. doi: 10.1038/nprot.2015.015.

O. T. Schubert, H. L. Röst, B. C. Collins, G. Rosenberger, and R. Aebersold. Quantitative proteomics: Challenges and opportunities in basic and applied research. *Nature Protocols*, 2017. ISSN 17502799. doi: 10.1038/nprot.2017.040.

N. Selevsek, C.-Y. Chang, L. C. Gillet, P. Navarro, O. M. Bernhardt, L. Reiter, L.-Y. Cheng, O. Vitek, and R. Aebersold. Reproducible and Consistent Quantification of the Saccharomyces cerevisiae Proteome by SWATH-mass spectrometry . *Molecular & Cellular Proteomics*, 2015. ISSN 1535-9476. doi: 10.1074/mcp.m113.035550.

E. N. Smith and L. Kruglyak. Gene-environment interaction in yeast gene expression. *PLoS Biology*, 2008. ISSN 15449173. doi: 10.1371/journal.pbio.0060083.

B. L. Snoek, M. G. Sterken, R. P. Bevers, R. J. Volkers, A. van't Hof, R. Brenchley, J. A. Riksen, A. Cossins, and J. E. Kammenga. Contribution of trans regulatory eQTL to cryptic genetic variation in C. elegans. *BMC Genomics*, 2017. ISSN 14712164. doi: 10.1186/s12864-017-3899-8.

P. S. Spencer and J. M. Barral. Genetic Code Redundancy and its Influence on the encoded Polypeptides. *Computational and Structural Biotechnology Journal*, 2012. ISSN 20010370. doi: 10.5936/csbj.201204006.

J. Swaminathan, A. A. Boulgakov, E. T. Hernandez, A. M. Bardo, J. L. Bachman, J. Marotta, A. M. Johnson, E. V. Anslyn, and E. M. Marcotte. Highly parallel single-molecule identification of proteins in zeptomole-scale mixtures. *Nature Biotechnology*, 2018. ISSN 15461696. doi: 10.1038/nbt.4278.

K. Tanaka, M. Nakafuku, F. Tamanoi, Y. Kaziro, K. Matsumoto, and A. Toh-e. IRA2, a second gene of Saccharomyces cerevisiae that encodes a protein with a domain homologous to mammalian ras GTPase-activating protein. *Molecular and cellular biology*, 1990. ISSN 0270-7306. doi: 10.1128/MCB.10.8.4303.

J. Teleman, H. L. Röst, G. Rosenberger, U. Schmitt, L. Malmström, J. Malmström, and F. Levander. DIANA-algorithmic improvements for analysis of data-independent acquisition MS data. *Bioinformatics*, 2015. ISSN 14602059. doi: 10.1093/bioinformatics/btu686.

G. Teo, S. Kim, C. C. Tsou, B. Collins, A. C. Gingras, A. I. Nesvizhskii, and H. Choi. MapDIA: Preprocessing and statistical analysis of quantitative proteomics data from data independent acquisition mass spectrometry. *Journal of Proteomics*, 2015. ISSN 18767737. doi: 10.1016/j.jprot.2015.09.013.

W. M. Toone, S. Kuge, M. Samuels, B. A. Morgan, T. Toda, and N. Jones. Regulation of the fission yeast transcription factor Pap1 by oxidative stress: Requirement for the nuclear export factor Crm1 (Exportin) and the stress-activated MAP kinase Sty1/Spc1. *Genes and Development*, 1998. ISSN 08909369. doi: 10.1101/gad.12.10.1453.

Y. F. Tsay, J. R. Thompson, M. O. Rotenberg, J. C. Larkin, and J. L. Woolford. Ribosomal protein synthesis is not regulated at the translational level in Saccharomyces cerevisiae: balanced accumulation of ribosomal proteins L16 and rp59 is mediated by turnover of excess protein. *Genes & development*, 1988. ISSN 08909369. doi: 10.1101/gad.2.6.664.

G. A. Van der Auwera, M. O. Carneiro, C. Hartl, R. Poplin, G. del Angel, A. Levy-Moonshine, T. Jordan, K. Shakir, D. Roazen, J. Thibault, E. Banks, K. V. Garimella, D. Altshuler, S. Gabriel, and M. A. DePristo. From fastQ data to high-confidence variant calls: The genome analysis toolkit best practices pipeline. *Current Protocols in Bioinformatics*, 2013. ISSN 1934340X. doi: 10.1002/0471250953.bi1110s43.

M. R. van der Sijde, A. Ng, and J. Fu. Systems genetics: From GWAS to disease pathways. *Biochimica et Biophysica Acta - Molecular Basis of Disease*, 1842(10): 1903–1909, 10 2014. ISSN 1879260X. doi: 10.1016/j.bbadis.2014.04.025.

J. Verdière, F. Creusot, and M. Guérineau. Regulation of the expression of iso 2-cytochrome c gene in S. cerevisiae: cloning of the positive regulatory gene CYP1 and identification of the region of its target sequence on the structural gene CYP3. *MGG Molecular & General Genetics*, 1985. ISSN 00263925. doi: 10.1007/BF00330769.

P. M. Visscher, W. G. Hill, and N. R. Wray. Heritability in the genomics era - Concepts and misconceptions. *Nature Reviews Genetics*, 2008. ISSN 14710056. doi: 10.1038/nrg2322.

P. M. Visscher, N. R. Wray, Q. Zhang, P. Sklar, M. I. McCarthy, M. A. Brown, and J. Yang. 10 Years of GWAS Discovery: Biology, Function, and Translation. *American Journal of Human Genetics*, 2017. ISSN 15376605. doi: 10.1016/j.ajhg. 2017.06.005.

C. Vogel and E. M. Marcotte. Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nature Reviews Genetics*, 2012. ISSN 14710056. doi: 10.1038/nrg3185.

M. G. Wilkinson, M. Samuels, T. Takeda, W. Mark Toone, J. C. Shieh, T. Toda, J. B. Millar, and N. Jones. The Atf1 transcription factor is a target for the Sty1 stress-activated MAP kinase pathway in fission yeast. *Genes and Development*, 1996. ISSN 08909369. doi: 10.1101/gad.10.18.2289.

V. Wood, R. Gwilliam, M. A. Rajandream, M. Lyne, R. Lyne, A. Stewart, J. Sgouros, N. Peat, J. Hayles, S. Baker, D. Basham, S. Bowman, K. Brooks, D. Brown, S. Brown, T. Chillingworth, C. Churcher, M. Collins, R. Connor, A. Cronin, P. Davis, T. Feltwell, A. Fraser, S. Gentles, A. Goble, N. Hamlin, D. Harris, J. Hidalgo, G. Hodgson, S. Holroyd, T. Hornsby, S. Howarth, E. J. Huckle, S. Hunt, K. Jagels, K. James, L. Jones, M. Jones, S. Leather, S. McDonald, J. McLean,

P. Mooney, S. Moule, K. Mungall, L. Murphy, D. Niblett, C. Odell, K. Oliver, S. O'Neil, D. Pearson, M. A. Quail, E. Rabbinowitsch, K. Rutherford, S. Rutter, D. Saunders, K. Seeger, S. Sharp, J. Skelton, M. Simmonds, R. Squares, S. Squares, K. Stevens, K. Taylor, R. G. Taylor, A. Tivey, S. Walsh, T. Warren, S. Whitehead, J. Woodward, G. Volckaert, R. Aert, J. Robben, B. Grymonprez, I. Weltjens, E. Vanstreels, M. Rieger, M. Schäfer, S. Müller-Auer, C. Gabel, M. Fuchs, C. Fritzc, E. Holzer, D. Moestl, H. Hilbert, K. Borzym, I. Langer, A. Beck, H. Lehrach, R. Reinhardt, T. M. Pohl, P. Eger, W. Zimmermann, H. Wedler, R. Wambutt, B. Purnelle, A. Goffeau, E. Cadieu, S. Dréano, S. Gloux, V. Lelaure, S. Mottier, F. Galibert, S. J. Aves, Z. Xiang, C. Hunt, K. Moore, S. M. Hurst, M. Lucas, M. Rochet, C. Gaillardin, V. A. Tallada, A. Garzon, G. Thode, R. R. Daga, L. Cruzado, J. Jimenez, M. Sánchez, F. del Rey, J. Benito, A. Domínguez, J. L. Revuelta, S. Moreno, J. Armstrong, S. L. Forsburg, L. Cerrutti, T. Lowe, W. R. McCombie, I. Paulsen, J. Potashkin, G. V. Shpakovski, D. Ussery, B. G. Barrell, and P. Nurse. The genome sequence of Schizosaccharomyces pombe. *Nature*, 2002. ISSN 00280836. doi: 10.1038/nature724.

V. Wood, M. A. Harris, M. D. McDowall, K. Rutherford, B. W. Vaughan, D. M. Staines, M. Aslett, A. Lock, J. Bähler, P. J. Kersey, and S. G. Oliver. PomBase: A comprehensive online resource for fission yeast. *Nucleic Acids Research*, 2012. ISSN 03051048. doi: 10.1093/nar/gkr853.

C. Wu, D. L. Delano, N. Mitro, S. V. Su, J. Janes, P. McClurg, S. Batalov, G. L. Welch, J. Zhang, A. P. Orth, J. R. Walker, R. J. Glynne, M. P. Cooke, J. S. Takahashi, K. Shimomura, A. Kohsaka, J. Bass, E. Saez, T. Wiltshire, and A. I. Su. Gene set enrichment in eQTL data identifies novel annotations and pathway regulators. *PLoS Genetics*, 2008. ISSN 15537390. doi: 10.1371/journal.pgen. 1000070.

L. Wu, S. I. Candille, Y. Choi, D. Xie, L. Jiang, J. Li-Pook-Than, H. Tang, and M. Snyder. Variation and genetic control of protein abundance in humans. *Nature*, 2013. ISSN 00280836. doi: 10.1038/nature12223.

L. J. Yan. Positive oxidative stress in aging and aging-related disease tolerance. *Redox Biology*, 2014. ISSN 22132317. doi: 10.1016/j.redox.2014.01.002.

V. J. Yohai. High Breakdown-Point and High Efficiency Robust Estimates for Regression. *The Annals of Statistics*, 2007. ISSN 0090-5364. doi: 10.1214/aos/1176350366.

G. Yvert, R. B. Brem, J. Whittle, J. M. Akey, E. Foss, E. N. Smith, R. Mackelprang, and L. Kruglyak. Trans-acting regulatory variation in Saccharomyces cerevisiae and the role of transcription factors. *Nature Genetics*, 2003. ISSN 10614036. doi: 10.1038/ng1222.

M. M. Zhang, P. Y. J. Wu, F. D. Kelly, P. Nurse, and H. C. Hang. Quantitative Control of Protein S-Palmitoylation Regulates Meiotic Entry in Fission Yeast. *PLoS Biology*, 2013. ISSN 15449173. doi: 10.1371/journal.pbio.1001597.

J. Zhu, B. Zhang, E. N. Smith, B. Drees, R. B. Brem, L. Kruglyak, R. E. Bumgarner, and E. E. Schadt. Integrating large-scale functional genomic data to dissect the complexity of yeast regulatory networks. *Nature Genetics*, 2008. ISSN 10614036. doi: 10.1038/ng.167.

Chapter 5

Supplements

5.1   Supplementary Tables

Table 5.1: Enrichments of stress-repressed transcripts in JB50.

Ontology	GO term	Description	p-value
BP	GO:0006364	rRNA processing	< 2e-16
BP	GO:0030490	maturation of SSU-rRNA	< 2e-16
BP	GO:0042273	ribosomal large subunit biogenesis	< 2e-16
BP	GO:0042254	ribosome biogenesis	1.4e-10
BP	GO:0006360	transcription from RNA polymerase I prom...	3.7e-08
BP	GO:0042274	ribosomal small subunit biogenesis	5.7e-08
BP	GO:0030488	tRNA methylation	2.5e-07
BP	GO:0006383	transcription from RNA polymerase III pr...	3.5e-07
BP	GO:0000054	ribosomal subunit export from nucleus	7.5e-07
BP	GO:0000466	maturation of 5.8S rRNA from tricistroni...	1.2e-06
MF	GO:0030515	snoRNA binding	< 2e-16
MF	GO:0004004	ATP-dependent RNA helicase activity	8.9e-11
MF	GO:0008757	S-adenosylmethionine-dependent methyltra...	2.1e-08
MF	GO:0003899	DNA-directed 5'-3' RNA polymerase activi...	2.9e-08
MF	GO:0003723	RNA binding	1.1e-07
MF	GO:0008175	tRNA methyltransferase activity	5.9e-07
MF	GO:0008173	RNA methyltransferase activity	0.00036
MF	GO:0140101	catalytic activity, acting on a tRNA	0.00041
MF	GO:0000049	tRNA binding	0.00044
MF	GO:0008170	N-methyltransferase activity	0.00129
CC	GO:0005730	nucleolus	< 2e-16
CC	GO:0032040	small-subunit processome	< 2e-16
CC	GO:0005736	DNA-directed RNA polymerase I complex	2.9e-09
CC	GO:0030686	90S preribosome	1.2e-08
CC	GO:0005666	DNA-directed RNA polymerase III complex	1.8e-08
CC	GO:0030687	preribosome, large subunit precursor	3.6e-07
CC	GO:0030688	preribosome, small subunit precursor	3.8e-07
CC	GO:0005732	small nucleolar ribonucleoprotein comple...	5.2e-06
CC	GO:0044452	nucleolar part	6.3e-06
CC	GO:0009277	fungal-type cell wall	2e-04

Table 5.2: Enrichments of stress-induced transcripts in JB50.

Ontology	GO term	Description	p-value
BP	GO:0006081	cellular aldehyde metabolic process	5.7e-07
BP	GO:0034599	cellular response to oxidative stress	7.9e-07
BP	GO:0098754	detoxification	9.3e-07
BP	GO:0044248	cellular catabolic process	2.1e-06
BP	GO:0016236	macroautophagy	2.1e-06
BP	GO:0044282	small molecule catabolic process	4.3e-06
BP	GO:0034614	cellular response to reactive oxygen spe...	4.8e-06
BP	GO:0032258	protein localization by the Cvt pathway	7.9e-06
BP	GO:0006111	regulation of gluconeogenesis	1.2e-05
BP	GO:0031327	negative regulation of cellular biosynth...	1.4e-05
MF	GO:0016616	oxidoreductase activity, acting on the C...	5.0e-10
MF	GO:0016491	oxidoreductase activity	2.2e-05
MF	GO:0004601	peroxidase activity	0.00018
MF	GO:0035091	phosphatidylinositol binding	0.00141
MF	GO:0016209	antioxidant activity	0.00171
MF	GO:0031072	heat shock protein binding	0.00268
MF	GO:0008238	exopeptidase activity	0.00583
CC	GO:0000324	fungal-type vacuole	5.4e-06
CC	GO:0034045	pre-autophagosomal structure membrane	7.8e-06
CC	GO:0000329	fungal-type vacuole membrane	0.0020
CC	GO:0005829	cytosol	0.0025
CC	GO:0005634	nucleus	0.0030
CC	GO:0044437	vacuolar part	0.0037
CC	GO:0016021	integral component of membrane	0.0052
CC	GO:0005783	endoplasmic reticulum	0.0076

Table 5.3: Enrichments of targets of e-chrI:1 among all measured transcripts.

Ontology	GO term	Description	p-value
BP	GO:0002181	cytoplasmic translation	< 2e-16
BP	GO:0042254	ribosome biogenesis	6.0e-09
BP	GO:0006364	rRNA processing	1.6e-07
BP	GO:0042273	ribosomal large subunit biogenesis	5.3e-06
BP	GO:0030490	maturation of SSU-rRNA	8.1e-06
BP	GO:0030488	tRNA methylation	9.1e-05
BP	GO:0000054	ribosomal subunit export from nucleus	0.00013
BP	GO:0002182	cytoplasmic translational elongation	0.00017
BP	GO:0002183	cytoplasmic translational initiation	0.00098
BP	GO:0006914	autophagy	0.00144
MF	GO:0003735	structural constituent of ribosome	2.5e-08
MF	GO:0030515	snoRNA binding	1.6e-07
MF	GO:0003743	translation initiation factor activity	6.9e-05
MF	GO:0019843	rRNA binding	0.00036
MF	GO:0008175	tRNA methyltransferase activity	0.00059
MF	GO:0008757	S-adenosylmethionine-dependent methyltra...	0.00230
MF	GO:0008536	Ran GTPase binding	0.00361
MF	GO:0004601	peroxidase activity	0.00648
MF	GO:0005524	ATP binding	0.00733
CC	GO:0022625	cytosolic large ribosomal subunit	< 2e-16
CC	GO:0005730	nucleolus	< 2e-16
CC	GO:0022627	cytosolic small ribosomal subunit	4.4e-12
CC	GO:0032040	small-subunit processome	1.2e-08
CC	GO:0005736	DNA-directed RNA polymerase I complex	0.00085
CC	GO:0005829	cytosol	0.00113
CC	GO:0030686	90S preribosome	0.00262
CC	GO:0016282	eukaryotic 43S preinitiation complex	0.00389
CC	GO:0030688	preribosome, small subunit precursor	0.00389
CC	GO:0005852	eukaryotic translation initiation factor...	0.00389

Table 5.4: Enrichments of targets of e-chrI:2 among all measured transcripts.

Ontology	GO term	Description	p-value
BP	GO:0002181	cytoplasmic translation	$< 2e{-}16$
BP	GO:0042254	ribosome biogenesis	1.4e-08
BP	GO:0006364	rRNA processing	6.4e-07
BP	GO:0042273	ribosomal large subunit biogenesis	1.7e-05
BP	GO:0030490	maturation of SSU-rRNA	3.2e-05
BP	GO:0000054	ribosomal subunit export from nucleus	4.2e-05
BP	GO:0030488	tRNA methylation	0.00023
BP	GO:0002182	cytoplasmic translational elongation	0.00052
BP	GO:0002183	cytoplasmic translational initiation	0.00280
BP	GO:0006914	autophagy	0.00284
MF	GO:0003735	structural constituent of ribosome	6.0e-08
MF	GO:0030515	snoRNA binding	6.2e-07
MF	GO:0003743	translation initiation factor activity	0.00022
MF	GO:0019843	rRNA binding	0.00112
MF	GO:0008175	tRNA methyltransferase activity	0.00127
MF	GO:0008536	Ran GTPase binding	0.00656
MF	GO:0008757	S-adenosylmethionine-dependent methyltra...	0.00766
CC	GO:0022625	cytosolic large ribosomal subunit	$< 2e{-}16$
CC	GO:0005730	nucleolus	$< 2e{-}16$
CC	GO:0022627	cytosolic small ribosomal subunit	6.3e-11
CC	GO:0032040	small-subunit processome	6.0e-08
CC	GO:0030687	preribosome, large subunit precursor	0.00099
CC	GO:0005736	DNA-directed RNA polymerase I complex	0.00158
CC	GO:0030686	90S preribosome	0.00534
CC	GO:0016282	eukaryotic 43S preinitiation complex	0.00630
CC	GO:0030688	preribosome, small subunit precursor	0.00630
CC	GO:0005852	eukaryotic translation initiation factor...	0.00630

Table 5.5: Enrichments of targets of e-chrII:1 among all measured transcripts.

Ontology	GO term	Description	p-value
BP	GO:1901607	alpha-amino acid biosynthetic process	7.1e-05
BP	GO:0043650	dicarboxylic acid biosynthetic process	7.2e-05
BP	GO:0018130	heterocycle biosynthetic process	0.00020
BP	GO:0061621	canonical glycolysis	0.00043
BP	GO:0006536	glutamate metabolic process	0.00048
BP	GO:0051604	protein maturation	0.00052
BP	GO:0006364	rRNA processing	0.00060
BP	GO:1901362	organic cyclic compound biosynthetic pro...	0.00062
BP	GO:0019751	polyol metabolic process	0.00096
BP	GO:0009082	branched-chain amino acid biosynthetic p...	0.00168
MF	GO:0005524	ATP binding	1.4e-06
MF	GO:0016836	hydro-lyase activity	7.8e-05
MF	GO:0016616	oxidoreductase activity, acting on the C...	8.8e-05
MF	GO:0008483	transaminase activity	0.00027
MF	GO:0030515	snoRNA binding	0.00053
MF	GO:0016620	oxidoreductase activity, acting on the a...	0.00085
MF	GO:0004004	ATP-dependent RNA helicase activity	0.00093
MF	GO:0004812	aminoacyl-tRNA ligase activity	0.00295
MF	GO:0019200	carbohydrate kinase activity	0.00645
MF	GO:0004298	threonine-type endopeptidase activity	0.00967
CC	GO:0032040	small-subunit processome	7.2e-06
CC	GO:0005730	nucleolus	0.00031
CC	GO:0005634	nucleus	0.00064
CC	GO:0005739	mitochondrion	0.00071
CC	GO:0005887	integral component of plasma membrane	0.00214
CC	GO:0005829	cytosol	0.00274
CC	GO:0005838	proteasome regulatory particle	0.00638
CC	GO:0044452	nucleolar part	0.00850

Table 5.6: Enrichments of targets of e-chrIII:1 among all measured transcripts.

Ontology	GO term	Description	p-value
BP	GO:0016573	histone acetylation	0.00038
BP	GO:0006338	chromatin remodeling	0.00073
BP	GO:0008360	regulation of cell shape	0.00151
BP	GO:0006378	mRNA polyadenylation	0.00165
BP	GO:0006400	tRNA modification	0.00360
BP	GO:0006379	mRNA cleavage	0.00425
BP	GO:0006357	regulation of transcription from RNA pol...	0.00426
BP	GO:0006284	base-excision repair	0.00659
BP	GO:1901615	organic hydroxy compound metabolic proce...	0.00662
BP	GO:0016579	protein deubiquitination	0.00933
MF	GO:1990837	sequence-specific double-stranded DNA bi...	0.0017
MF	GO:0140101	catalytic activity, acting on a tRNA	0.0043
MF	GO:0003756	protein disulfide isomerase activity	0.0066
MF	GO:0004843	thiol-dependent ubiquitin-specific prote...	0.0083
CC	GO:0005847	mRNA cleavage and polyadenylation specif...	0.0045
CC	GO:0000124	SAGA complex	0.0070
CC	GO:0033202	DNA helicase complex	0.0082

Table 5.7: Enrichments of targets of p-chrI:1 at FDR<25% among all measured proteins.

Ontology	GO term	Description	p-value
MF	GO:0003779	actin binding	0.0035
MF	GO:0005525	GTP binding	0.0080
CC	GO:0030479	actin cortical patch	0.0068

Table 5.8: Enrichments of targets of p-chrI:2 at FDR<25% among all measured proteins.

Ontology	GO term	Description	p-value
BP	GO:0007035	vacuolar acidification	0.00032
BP	GO:0015991	ATP hydrolysis coupled proton transport	0.00058
MF	GO:0046961	proton-transporting ATPase activity, rot...	0.0042
MF	GO:0000287	magnesium ion binding	0.0081
CC	GO:0016471	vacuolar proton-transporting V-type ATPa...	0.00032
CC	GO:0033178	proton-transporting two-sector ATPase co...	0.00074

Table 5.9: Enrichments of targets of p-chrI:4 at FDR<25% among all measured proteins.

Ontology	GO term	Description	p-value
BP	GO:0006099	tricarboxylic acid cycle	7.5e-05
BP	GO:0042775	mitochondrial ATP synthesis coupled elec...	0.0014
BP	GO:0043648	dicarboxylic acid metabolic process	0.0046
BP	GO:0072593	reactive oxygen species metabolic proces...	0.0066
BP	GO:1902600	hydrogen ion transmembrane transport	0.0068
BP	GO:0006090	pyruvate metabolic process	0.0092
CC	GO:0005759	mitochondrial matrix	1e-06

Table 5.10: Enrichments of targets of p-chrII:1 at FDR<25% among all measured proteins.

Ontology	GO term	Description	p-value
BP	GO:0006098	pentose-phosphate shunt	7.0e-07
BP	GO:0006099	tricarboxylic acid cycle	1.4e-05
BP	GO:0042775	mitochondrial ATP synthesis coupled elec...	4.0e-05
BP	GO:0006536	glutamate metabolic process	0.00051
BP	GO:0042776	mitochondrial ATP synthesis coupled prot...	0.00051
BP	GO:0046365	monosaccharide catabolic process	0.00123
BP	GO:0044262	cellular carbohydrate metabolic process	0.00254
BP	GO:0006091	generation of precursor metabolites and ...	0.00275
BP	GO:0032787	monocarboxylic acid metabolic process	0.00534
BP	GO:0006081	cellular aldehyde metabolic process	0.00553
MF	GO:0016616	oxidoreductase activity, acting on the C...	2.3e-06
MF	GO:0046933	proton-transporting ATP synthase activit...	0.00078
MF	GO:0046872	metal ion binding	0.00160
MF	GO:0016773	phosphotransferase activity, alcohol gro...	0.00651
MF	GO:0016787	hydrolase activity	0.00655
MF	GO:0000287	magnesium ion binding	0.00762
MF	GO:0046961	proton-transporting ATPase activity, rot...	0.00898
MF	GO:0003779	actin binding	0.00926
MF	GO:0016651	oxidoreductase activity, acting on NAD(P...	0.00991
MF	GO:0004721	phosphoprotein phosphatase activity	0.00991
CC	GO:0005746	mitochondrial respiratory chain	9.2e-05
CC	GO:0098803	respiratory chain complex	0.00063
CC	GO:0070069	cytochrome complex	0.00126
CC	GO:1990204	oxidoreductase complex	0.00395
CC	GO:0005759	mitochondrial matrix	0.00442
CC	GO:0005739	mitochondrion	0.00547
CC	GO:0098800	inner mitochondrial membrane protein com...	0.00761
CC	GO:0005753	mitochondrial proton-transporting ATP sy...	0.00886

Table 5.11: Enrichments of targets of p-chrIII:1 at FDR<25% among all measured proteins.

Ontology	GO term	Description	p-value
BP	GO:0002181	cytoplasmic translation	3.3e-10
BP	GO:0061077	chaperone-mediated protein folding	0.00032
BP	GO:0006457	protein folding	0.00056
BP	GO:0061621	canonical glycolysis	0.00203
BP	GO:0009082	branched-chain amino acid biosynthetic p...	0.00550
BP	GO:0045842	positive regulation of mitotic metaphase...	0.00597
BP	GO:0002183	cytoplasmic translational initiation	0.01000
MF	GO:0003735	structural constituent of ribosome	3.2e-06
MF	GO:0051082	unfolded protein binding	2.1e-05
MF	GO:0019843	rRNA binding	2.8e-05
MF	GO:0016887	ATPase activity	6.9e-05
MF	GO:0005524	ATP binding	0.00043
MF	GO:0003743	translation initiation factor activity	0.00595
MF	GO:0004175	endopeptidase activity	0.00685
MF	GO:0003755	peptidyl-prolyl cis-trans isomerase acti...	0.00777
CC	GO:0022625	cytosolic large ribosomal subunit	3.0e-06
CC	GO:0005838	proteasome regulatory particle	1.9e-05
CC	GO:0044445	cytosolic part	0.00024
CC	GO:0000785	chromatin	0.00037
CC	GO:0005829	cytosol	0.00195
CC	GO:0022627	cytosolic small ribosomal subunit	0.00384

Table 5.12: Enrichments of targets of p-chrIII:2 at FDR<25% among all measured proteins.

Ontology	GO term	Description	p-value
BP	GO:0006536	glutamate metabolic process	0.0020
BP	GO:0009225	nucleotide-sugar metabolic process	0.0046
BP	GO:0002181	cytoplasmic translation	0.0050
BP	GO:0007021	tubulin complex assembly	0.0050
BP	GO:0035967	cellular response to topologically incor...	0.0067
MF	GO:0051082	unfolded protein binding	0.00099
MF	GO:0005524	ATP binding	0.00132
MF	GO:0005525	GTP binding	0.00451

Table 5.13: Enrichment analysis of targets of e-chrI:2 in the stress-condition. All quantified transcripts were chosen as the background.

Ontology	GO term	Description	p-value
BP	GO:0002181	cytoplasmic translation	< 2e-16
BP	GO:0006364	rRNA processing	< 2e-16
BP	GO:0042254	ribosome biogenesis	7.1e-16
BP	GO:0002183	cytoplasmic translational initiation	8.4e-09
BP	GO:0002182	cytoplasmic translational elongation	6.3e-08
BP	GO:0030490	maturation of SSU-rRNA	4.7e-07
BP	GO:0042273	ribosomal large subunit biogenesis	3.2e-06
BP	GO:0006457	protein folding	6.3e-06
BP	GO:0009168	purine ribonucleoside monophosphate bios...	0.00019
BP	GO:0009152	purine ribonucleotide biosynthetic proce...	0.00022
MF	GO:0003735	structural constituent of ribosome	< 2e-16
MF	GO:0019843	rRNA binding	4.2e-09
MF	GO:0003743	translation initiation factor activity	1.6e-08
MF	GO:0003723	RNA binding	6.2e-08
MF	GO:0051082	unfolded protein binding	1.8e-06
MF	GO:0030515	snoRNA binding	3.4e-06
MF	GO:0003899	DNA-directed 5'-3' RNA polymerase activi...	6.5e-05
MF	GO:0043021	ribonucleoprotein complex binding	0.0007
MF	GO:0008175	tRNA methyltransferase activity	0.0014
MF	GO:0016879	ligase activity, forming carbon-nitrogen...	0.0025
CC	GO:0005730	nucleolus	< 2e-16
CC	GO:0022625	cytosolic large ribosomal subunit	< 2e-16
CC	GO:0022627	cytosolic small ribosomal subunit	< 2e-16
CC	GO:0032040	small-subunit processome	4.6e-07
CC	GO:0044452	nucleolar part	3.0e-05
CC	GO:0030686	90S preribosome	4.9e-05
CC	GO:0005732	small nucleolar ribonucleoprotein comple...	4.9e-05
CC	GO:0005736	DNA-directed RNA polymerase I complex	0.00018
CC	GO:0005829	cytosol	0.00043
CC	GO:0005852	eukaryotic translation initiation factor...	0.00070

Table 5.14: Enrichment analysis of targets of e-chrI:2 in the stress-condition that were up-regulated in the strains with JB759-allele. All significant targets of this hotspot in the stress condition were chosen as the background.

Ontology	GO term	Description	p-value
BP	GO:0006897	endocytosis	3.8e-09
BP	GO:0035556	intracellular signal transduction	1.5e-06
BP	GO:0034613	cellular protein localization	4.2e-06
BP	GO:0007165	signal transduction	6.3e-06
BP	GO:0030437	ascospore formation	9.1e-06
BP	GO:0006886	intracellular protein transport	1.3e-05
BP	GO:0006875	cellular metal ion homeostasis	3.4e-05
BP	GO:0061024	membrane organization	4.6e-05
BP	GO:0010927	cellular component assembly involved in ...	7.1e-05
BP	GO:0006887	exocytosis	0.00014
MF	GO:0004674	protein serine/threonine kinase activity	5.8e-06
MF	GO:0000978	RNA polymerase II core promoter proximal...	1.4e-05
MF	GO:0035091	phosphatidylinositol binding	0.00014
MF	GO:0008017	microtubule binding	0.00065
MF	GO:0004003	ATP-dependent DNA helicase activity	0.00126
MF	GO:0061630	ubiquitin protein ligase activity	0.00226
MF	GO:0008270	zinc ion binding	0.00348
MF	GO:0001077	transcriptional activator activity, RNA ...	0.00406
MF	GO:0005515	protein binding	0.00505
MF	GO:0000982	transcription factor activity, RNA polym...	0.00730
CC	GO:0016021	integral component of membrane	6.8e-09
CC	GO:0032153	cell division site	2.2e-08
CC	GO:0051286	cell tip	9.5e-08
CC	GO:0005794	Golgi apparatus	1.3e-07
CC	GO:0000329	fungal-type vacuole membrane	2.0e-07
CC	GO:0044732	mitotic spindle pole body	7.8e-07
CC	GO:0000790	nuclear chromatin	1.7e-06
CC	GO:0005886	plasma membrane	5.6e-06
CC	GO:0000778	condensed nuclear chromosome kinetochore	1.9e-05
CC	GO:0000324	fungal-type vacuole	0.00011

Table 5.15: Enrichment analysis of targets of e-chrI:2 in the stress-condition that were down-regulated in the strains with JB759-allele. All significant targets of this hotspot in the stress condition were chosen as the background.

Ontology	GO term	Description	p-value
BP	GO:0002181	cytoplasmic translation	< 2e-16
BP	GO:0006364	rRNA processing	< 2e-16
BP	GO:0042254	ribosome biogenesis	7.1e-16
BP	GO:0002183	cytoplasmic translational initiation	8.4e-09
BP	GO:0002182	cytoplasmic translational elongation	6.3e-08
BP	GO:0030490	maturation of SSU-rRNA	4.7e-07
BP	GO:0042273	ribosomal large subunit biogenesis	3.2e-06
BP	GO:0006457	protein folding	6.3e-06
BP	GO:0009168	purine ribonucleoside monophosphate bios...	0.00019
BP	GO:0009152	purine ribonucleotide biosynthetic proce...	0.00022
MF	GO:0003735	structural constituent of ribosome	< 2e-16
MF	GO:0019843	rRNA binding	4.2e-09
MF	GO:0003743	translation initiation factor activity	1.6e-08
MF	GO:0003723	RNA binding	6.2e-08
MF	GO:0051082	unfolded protein binding	1.8e-06
MF	GO:0030515	snoRNA binding	3.4e-06
MF	GO:0003899	DNA-directed 5'-3' RNA polymerase activi...	6.5e-05
MF	GO:0043021	ribonucleoprotein complex binding	0.0007
MF	GO:0008175	tRNA methyltransferase activity	0.0014
MF	GO:0016879	ligase activity, forming carbon-nitrogen...	0.0025
CC	GO:0005730	nucleolus	< 2e-16
CC	GO:0022625	cytosolic large ribosomal subunit	< 2e-16
CC	GO:0022627	cytosolic small ribosomal subunit	< 2e-16
CC	GO:0032040	small-subunit processome	4.6e-07
CC	GO:0044452	nucleolar part	3.0e-05
CC	GO:0030686	90S preribosome	4.9e-05
CC	GO:0005732	small nucleolar ribonucleoprotein comple...	4.9e-05
CC	GO:0005736	DNA-directed RNA polymerase I complex	0.00018
CC	GO:0005829	cytosol	0.00043
CC	GO:0005852	eukaryotic translation initiation factor...	0.00070

Table 5.16: Enrichment analysis of targets of e-chrI:1 that were affected by conditional eQTL that were specific to the stress condition, and up-regulated in strains with the JB50 allele at this locus, and showed a stronger stress response in the strains with the JB50-allele. All transcripts that were targeted by a conditional eQTL that was specific to the stress-condition were chosen as the background.

Ontology	GO term	Description	p-value
BP	GO:0006914	autophagy	7.2e-05
BP	GO:0033554	cellular response to stress	0.00015
BP	GO:0016197	endosomal transport	0.00021
BP	GO:0010033	response to organic substance	0.00050
BP	GO:0072665	protein localization to vacuole	0.00050
BP	GO:0006511	ubiquitin-dependent protein catabolic pr...	0.00050
BP	GO:0016192	vesicle-mediated transport	0.00054
BP	GO:0016482	cytosolic transport	0.00141
BP	GO:0016567	protein ubiquitination	0.00197
BP	GO:0044262	cellular carbohydrate metabolic process	0.00232
MF	GO:0070011	peptidase activity, acting on L-amino ac...	0.0002
MF	GO:0061630	ubiquitin protein ligase activity	0.0030
MF	GO:0008270	zinc ion binding	0.0085
CC	GO:0016021	integral component of membrane	2.7e-06
CC	GO:0098805	whole membrane	4.1e-05
CC	GO:0000329	fungal-type vacuole membrane	0.00016
CC	GO:0098588	bounding membrane of organelle	0.00143
CC	GO:0031966	mitochondrial membrane	0.00345
CC	GO:0032153	cell division site	0.00460
CC	GO:0000324	fungal-type vacuole	0.00825
CC	GO:0030864	cortical actin cytoskeleton	0.00955
CC	GO:0030176	integral component of endoplasmic reticu...	0.00955

Table 5.17: Enrichment analysis of targets of e-chrI:1 that were affected by conditional eQTL that were specific to the stress condition, and down-regulated in strains with the JB50 allele at this locus, and showed a stronger stress response in the strains with the JB50-allele. All transcripts that were targeted by a conditional eQTL that was specific to the stress-condition were chosen as the background.

Ontology	GO term	Description	p-value
BP	GO:0006364	rRNA processing	< 2e-16
BP	GO:0042273	ribosomal large subunit biogenesis	3.6e-11
BP	GO:0042254	ribosome biogenesis	9.6e-11
BP	GO:0030490	maturation of SSU-rRNA	2.0e-10
BP	GO:0006360	transcription from RNA polymerase I prom...	4.7e-06
BP	GO:0002181	cytoplasmic translation	1.2e-05
BP	GO:0030488	tRNA methylation	5.7e-05
BP	GO:0006400	tRNA modification	7.3e-05
BP	GO:0000054	ribosomal subunit export from nucleus	9.6e-05
BP	GO:0000466	maturation of 5.8S rRNA from tricistroni...	0.00013
MF	GO:0003723	RNA binding	1.4e-14
MF	GO:0008757	S-adenosylmethionine-dependent methyltra...	1.0e-09
MF	GO:0030515	snoRNA binding	1.0e-09
MF	GO:0004004	ATP-dependent RNA helicase activity	1.6e-07
MF	GO:0003899	DNA-directed 5'-3' RNA polymerase activi...	9.3e-06
MF	GO:0008170	N-methyltransferase activity	0.00013
MF	GO:0008175	tRNA methyltransferase activity	0.00029
CC	GO:0005730	nucleolus	2e-16
CC	GO:0032040	small-subunit processome	5.9e-12
CC	GO:0030686	90S preribosome	2.0e-06
CC	GO:0044452	nucleolar part	6.5e-06
CC	GO:0005634	nucleus	2.4e-05
CC	GO:0005736	DNA-directed RNA polymerase I complex	5.5e-05
CC	GO:0005666	DNA-directed RNA polymerase III complex	0.00029
CC	GO:0030687	preribosome, large subunit precursor	0.00029
CC	GO:0030684	preribosome	0.00070
CC	GO:0005732	small nucleolar ribonucleoprotein comple...	0.00090

Table 5.18: Enrichment analysis of targets of e-chrI:1 that were affected by conditional eQTL that were specific to the stress condition, and up-regulated in strains with the JB50 allele at this locus, and showed a stronger stress response in the strains with the JB759-allele. All transcripts that were targeted by a conditional eQTL that was specific to the stress-condition were chosen as the background.

Ontology	GO term	Description	p-value
BP	GO:1901607	alpha-amino acid biosynthetic process	0.00027
BP	GO:0010557	positive regulation of macromolecule bio...	0.00156
BP	GO:0031328	positive regulation of cellular biosynth...	0.00264
BP	GO:0030001	metal ion transport	0.00416
BP	GO:0051640	organelle localization	0.00496
BP	GO:0098662	inorganic cation transmembrane transport	0.00624
BP	GO:0043604	amide biosynthetic process	0.00842
CC	GO:0016021	integral component of membrane	0.0069
CC	GO:0005783	endoplasmic reticulum	0.0099

Table 5.19: Enrichment analysis of targets of e-chrI:1 that were affected by conditional eQTL that were specific to the stress condition, and down-regulated in strains with the JB50 allele at this locus, and showed a stronger stress response in the strains with the JB759-allele. All transcripts that were targeted by a conditional eQTL that was specific to the stress-condition were chosen as the background.

Ontology	GO term	Description	p-value
BP	GO:0034645	cellular macromolecule biosynthetic proc...	0.00042
BP	GO:1901990	regulation of mitotic cell cycle phase t...	0.00211
BP	GO:0000122	negative regulation of transcription fro...	0.00284
BP	GO:0033044	regulation of chromosome organization	0.00384
BP	GO:0000070	mitotic sister chromatid segregation	0.00887
CC	GO:0043234	protein complex	0.0023

Table 5.20: Enrichment analysis of targets of e-chrII:1 in the normal condition. All coding transcripts were chosen as the background.

Ontology	GO term	Description	p-value
BP	GO:0006364	rRNA processing	4.4e-12
BP	GO:0030490	maturation of SSU-rRNA	0.00030
BP	GO:0006222	UMP biosynthetic process	0.00031
BP	GO:0001522	pseudouridine synthesis	0.00153
BP	GO:0030488	tRNA methylation	0.00175
BP	GO:0002098	tRNA wobble uridine modification	0.00233
BP	GO:0015718	monocarboxylic acid transport	0.00284
BP	GO:1901566	organonitrogen compound biosynthetic pro...	0.00374
BP	GO:0000466	maturation of 5.8S rRNA from tricistroni...	0.00504
BP	GO:0006479	protein methylation	0.00510
MF	GO:0004004	ATP-dependent RNA helicase activity	8.8e-08
MF	GO:0030515	snoRNA binding	2.3e-05
MF	GO:0008757	S-adenosylmethionine-dependent methyltra...	0.00036
MF	GO:0008175	tRNA methyltransferase activity	0.00246
MF	GO:0016614	oxidoreductase activity, acting on CH-OH...	0.00374
MF	GO:0008276	protein methyltransferase activity	0.00655
MF	GO:0008170	N-methyltransferase activity	0.00936
CC	GO:0005730	nucleolus	2.4e-11
CC	GO:0032040	small-subunit processome	7.2e-07
CC	GO:0044452	nucleolar part	6.0e-05
CC	GO:0030687	preribosome, large subunit precursor	0.00033
CC	GO:0030686	90S preribosome	0.00064
CC	GO:0005753	mitochondrial proton-transporting ATP sy...	0.00131
CC	GO:0005736	DNA-directed RNA polymerase I complex	0.00960

Table 5.21: Enrichment analysis of targets of e-chrII:1 in the normal condition that are up-regulated in the strains with the JB760-allele. All coding transcripts that were affected by conditional eQTL specific to the normal condition, were included in the background.

Ontology	GO term	Description	p-value
BP	GO:0044248	cellular catabolic process	0.00078
BP	GO:0006839	mitochondrial transport	0.00445
BP	GO:0044282	small molecule catabolic process	0.00574
BP	GO:0019953	sexual reproduction	0.00804
MF	GO:0016788	hydrolase activity, acting on ester bond...	0.00095
MF	GO:0016614	oxidoreductase activity, acting on CH-OH...	0.00445
MF	GO:0043168	anion binding	0.00537
CC	GO:0005739	mitochondrion	0.00049

Table 5.22: Enrichment analysis of targets of e-chrII:1 in the normal condition that are down-regulated in the strains with the JB760-allele. All coding transcripts that were affected by conditional eQTL specific to the normal condition, were included in the background.

Ontology	GO term	Description	p-value
BP	GO:0006364	rRNA processing	1.7e-10
BP	GO:0006400	tRNA modification	6.7e-05
BP	GO:0042254	ribosome biogenesis	0.0005
BP	GO:0030490	maturation of SSU-rRNA	0.0024
BP	GO:0001510	RNA methylation	0.0041
BP	GO:0042273	ribosomal large subunit biogenesis	0.0041
BP	GO:0071426	ribonucleoprotein complex export from nu...	0.0067
MF	GO:0003723	RNA binding	1.3e-06
MF	GO:0004004	ATP-dependent RNA helicase activity	0.00019
MF	GO:0008757	S-adenosylmethionine-dependent methyltra...	0.00147
MF	GO:0005525	GTP binding	0.00244
MF	GO:0030515	snoRNA binding	0.00244
MF	GO:0140101	catalytic activity, acting on a tRNA	0.00244
MF	GO:0140098	catalytic activity, acting on RNA	0.00527
MF	GO:0008173	RNA methyltransferase activity	0.00673
MF	GO:0005524	ATP binding	0.00692
MF	GO:0003677	DNA binding	0.00889
CC	GO:0005730	nucleolus	4.6e-14
CC	GO:0043234	protein complex	1.0e-05
CC	GO:0044452	nucleolar part	7.0e-05
CC	GO:0032040	small-subunit processome	0.00055
CC	GO:0030684	preribosome	0.00068
CC	GO:0030529	intracellular ribonucleoprotein complex	0.00548
CC	GO:0000790	nuclear chromatin	0.00550

Table 5.23: Enrichment analysis of genes affected by eQTL classified as 'similar'.

ontology	GO.ID	Term	pvalue
BP	GO:0030150	protein import into mitochondrial matrix	0.0036
BP	GO:0097502	mannosylation	0.0059
BP	GO:0031505	fungal-type cell wall organization	0.0064
BP	GO:0006696	ergosterol biosynthetic process	0.0073
BP	GO:0016485	protein processing	0.0091
MF	GO:0005525	GTP binding	0.0017
MF	GO:0016887	ATPase activity	0.0099

Table 5.24: Enrichment analysis of genes affected by eQTL classified as 'buffered'.

ontology	GO.ID	Term	pvalue
BP	GO:0000463	maturation of LSU-rRNA from tricistronic...	3.6e-09
BP	GO:0006364	rRNA processing	4e-06
BP	GO:0002181	cytoplasmic translation	3.9e-05
BP	GO:0000447	endonucleolytic cleavage in ITS1 to sepa...	0.00011
BP	GO:0000466	maturation of 5.8S rRNA from tricistroni...	0.00036
BP	GO:0000472	endonucleolytic cleavage to generate mat...	0.00101
BP	GO:0000027	ribosomal large subunit assembly	0.00108
BP	GO:0042797	tRNA transcription from RNA polymerase I...	0.00137
BP	GO:0000480	endonucleolytic cleavage in 5'-ETS of tr...	0.00163
BP	GO:0006259	DNA metabolic process	0.00189
BP	GO:0006526	arginine biosynthetic process	0.00334
BP	GO:0006360	transcription from RNA polymerase I prom...	0.00517
BP	GO:0000055	ribosomal large subunit export from nucl...	0.0077
BP	GO:0042254	ribosome biogenesis	0.00774
BP	GO:0031668	cellular response to extracellular stimu...	0.00824
MF	GO:0001056	RNA polymerase III activity	0.00026
MF	GO:0003723	RNA binding	0.00048
MF	GO:0001054	RNA polymerase I activity	0.00216
MF	GO:0042134	rRNA primary transcript binding	0.00718
MF	GO:0019200	carbohydrate kinase activity	0.00837
CC	GO:0005730	nucleolus	2.7e-18
CC	GO:0030687	preribosome, large subunit precursor	6.5e-11
CC	GO:0022625	cytosolic large ribosomal subunit	6.1e-05
CC	GO:0005666	DNA-directed RNA polymerase III complex	0.00026
CC	GO:0005654	nucleoplasm	0.00054
CC	GO:0032040	small-subunit processome	0.00106
CC	GO:0005736	DNA-directed RNA polymerase I complex	0.00216
CC	GO:0070545	PeBoW complex	0.00845

Table 5.25: Enrichment analysis of genes affected by eQTL classified as 'enhanced'.

ontology	GO.ID	Term	pvalue
BP	GO:0032543	mitochondrial translation	1.9e-20
BP	GO:0019419	sulfate reduction	0.00068
BP	GO:0019344	cysteine biosynthetic process	0.0031
BP	GO:0006839	mitochondrial transport	0.00548
BP	GO:0070814	hydrogen sulfide biosynthetic process	0.00592
BP	GO:0000947	amino acid catabolic process to alcohol ...	0.00592
BP	GO:0005975	carbohydrate metabolic process	0.006
MF	GO:0003735	structural constituent of ribosome	4.6e-13
MF	GO:0016744	transferase activity, transferring aldeh...	0.0025
MF	GO:0004022	alcohol dehydrogenase (NAD) activity	0.0025
CC	GO:0005762	mitochondrial large ribosomal subunit	2.5e-14
CC	GO:0005763	mitochondrial small ribosomal subunit	3.9e-09
CC	GO:0005737	cytoplasm	6e-04
CC	GO:0005759	mitochondrial matrix	0.0011
CC	GO:0005739	mitochondrion	0.0048

Table 5.26: Enrichment analysis of genes affected by ptQTL.

ontology	GO.ID	Term	pvalue
BP	GO:0002181	cytoplasmic translation	4.7e-05
BP	GO:0055114	oxidation-reduction process	0.0076
BP	GO:0006099	tricarboxylic acid cycle	0.0085
CC	GO:0022625	cytosolic large ribosomal subunit	0.0015
CC	GO:0042645	mitochondrial nucleoid	0.0095

Table 5.27: QTL affecting growth under various conditions were taken from (Bloom, 2013). Listed below are the growth QTL affected by hotspots that were mentioned in the text. Hotspots that affected only expression typically affected fewer growth traits than hotspots that affected many phospho-traits.

chrV:2	chrVIII:1	chrXII:1	chrXII:2	chrXV:1
Menadione	Caffeine	Cobalt_Chloride	4-Hydroxybenzaldehyde	4NQO
	Congo_red		4NQO	5-Fluorocytosine
	Cycloheximide		Cadmium_Chloride	5-Fluorouracil
	Diamide		Calcium_Chloride	6-Azauracil
	E6_Berbamine		Cisplatin	Calcium_Chloride
	Formamide		Cobalt_Chloride	Cisplatin
	Indoleacetic_Acid		Copper	Cobalt_Chloride
	Menadione		Cycloheximide	Congo_red
	SDS		E6_Berbamine	Copper
	YNB		Ethanol	Cycloheximide
	YNB:ph3		Formamide	E6_Berbamine
	YNB:ph8		Lithium_Chloride	Ethanol
	YPD		Manganese_Sulfate	Formamide
	YPD:15C		Mannose	Galactose
	Zeocin		Neomycin	Hydrogen_Peroxide
	5-Fluorocytosine		Tunicamycin	Lactate
	5-Fluorouracil		YNB	Lactose
			YPD	Lithium_Chloride
			Zeocin	Magnesium_Chloride
			Hotspot XIV:1	Maltose
			4-Hydroxybenzaldehyde	Manganese_Sulfate
			4NQO	Mannose
			5-Fluorocytosine	Neomycin
			5-Fluorouracil	Paraquat
			Caffeine	Raffinose
			Calcium_Chloride	SDS
			Cisplatin	Sorbitol
			Copper	Trehalose
			Cycloheximide	Xylose
			E6_Berbamine	YNB
			Ethanol	YNB:ph3
			Formamide	YPD
			Hydrogen_Peroxide	YPD:15C
			Indoleacetic_Acid	YPD:37C
			Magnesium_Chloride	YPD:4C
			Raffinose	Zeocin
			Trehalose	
			Tunicamycin	
			Xylose	
			YNB	
			YNB:ph3	
			YPD	
			YPD:37C	
			YPD:4C	

Table 5.28: Targets of chrVIII:1 with a reported link to *STE20/GPA1*.

name	references
GIP4	PMID: 16537909
FUS1	PMID: 17604854, PMID: 1903837
KAR4	PMID: 17604854, PMID: 16522208
LRG1	PMID: 27704052, PMID: 11591390
RGA2	PMID: 12455995
MFA1	PMID: 2659433, PMID: 17604854
APA2	PMID: 16522208, PMID: 12732146
RGD2	PMID: 24062589, PMID: 11591390
CRH1	PMID: 16522208, PMID: 19234305, PMID: 18184748
GPA1	PMID: 2536595, PMID: 9832519, PMID: 10712512, PMID: 12556475
PRM5	PMID: 10535956, PMID: 11062271
EXO70	PMID: 19955214
FAR1	PMID: 12029138, PMID: 18261907, PMID: 8500168, PMID: 15690603
STE18	PMID: 2536595, PMID: 7834739
HYM1	PMID: 16522208
SST2	PMID: 9537998
PRP39	PMID: 16522208
KAR5	PMID: 12052881, PMID:16522208
FUS2	PMID: 16522208, PMID: 11166190
PRM1	PMID: 11062271, PMID: 11166190, PMID: 16522208
AGA1	PMID: 11166190, PMID: 15690603, PMID: 2072914
VHS3	PMID: 25904326
STE13	PMID: 2685554
DIG1	PMID: 17604854, PMID: 8918885, PMID: 9094309
PRM4	PMID: 11062271, PMID: 11166190, PMID: 16522208
RGC1	PMID: 24298058, PMID: 12142009, PMID: 10970855
TEC1	PMID: 17118154
PST1	PMID: 10535956, PMID: 11062271
AFR1	PMID: 16522208, PMID: 11743162, PMID: 19841731, PMID: 18552279, PMID: 20489023, PMID: 20093466
CDC20	PMID: 21329885, PMID: 12642613
SPO11	PMID: 9456310
RIM101	PMID: 20333241
HAL1	PMID: 16522208, PMID: 11525741, PMID: 12040128
FUS3	PMID: 11525741
ERG11	PMID: 17895367
SNL1	PMID: 24121774
SKT5	PMID: 16522208, PMID: 12732146
AKL1	PMID: 19269370, PMID: 22875988
REG1	PMID: 24003253
GIC2	PMID: 14734533, PMID: 9367979
DIG2	PMID: 15690603, PMID: 17604854, PMID: 9841672
BOI2	PMID: 23785492
AVT1	PMID: 16522208
MSN4	PMID: 12142009, PMID: 10970855
MNR2	PMID: 16319894
SRP40	PMID: 17101777
SSK1	PMID: 10970855, PMID: 8808622
RCK2	PMID: 10805732
NUP2	PMID: 23645671
BDF1	PMID: 16118188, PMID: 16319894
RSC9	PMID: 19153600
MLF3	PMID: 16816427
NBA1	PMID: 25416945, PMID: 19841731, PMID: 28751498
RTS1	PMID: 8846889
MSC3	PMID: 21329885, PMID: 15665377

Table 5.29: Targets of chrVIII:1 with a plausible link to *STE20/GPA1*.

name	references
MUM2	PMID: 14585977, PMID: 22875988
SHU1	PMID: 19234305, PMID: 22438580, PMID: 17604854
YAP3	PMID: 21127252, PMID: 17417638
CNA1	PMID: 1651503
SPT21	PMID: 21057056
PSK1	PMID: 19269370
TAF12	PMID: 16118188, PMID: 29079657
ZRG8	PMID: 24062589, PMID: 15972461
VPS27	PMID: 20093466, PMID: 22282571, PMID: 27708008
GDS1	PMID: 27708008, PMID: 20093466
HEK2	PMID: 18805955
SEG1	PMID: 26359496
LSP1	PMID: 26359496
EIS1	PMID: 26359496

Table 5.30: Functional enrichment analysis of eQTL targets of chrVIII:1.

ontology	GO.ID	Term	pvalue
BP	GO:0000755	cytogamy	4.2e-06
BP	GO:0000742	karyogamy involved in conjugation with c...	1.4e-05
BP	GO:0000750	pheromone-dependent signal transduction ...	1.9e-05
BP	GO:0000753	cell morphogenesis involved in conjugati...	0.00059
BP	GO:2000220	regulation of pseudohyphal growth	0.00128
BP	GO:0022604	regulation of cell morphogenesis	0.00212
BP	GO:0000754	adaptation of signaling pathway by respo...	0.003
BP	GO:0046020	negative regulation of transcription fro...	0.00765
BP	GO:1900429	negative regulation of filamentous growt...	0.00765
BP	GO:0043547	positive regulation of GTPase activity	0.00909
BP	GO:0120031	plasma membrane bounded cell projection ...	0.0091
MF	GO:0016836	hydro-lyase activity	0.0047
MF	GO:0005096	GTPase activator activity	0.0048
CC	GO:0043332	mating projection tip	5e-04
CC	GO:0000131	incipient cellular bud site	0.0077
CC	GO:0098797	plasma membrane protein complex	0.0091

Table 5.31: Targets of chrVIII:1 without a known link to *STE20/GPA1*.

YBR225W	LOT5	YHL017W	APN1	SFB3
RAD59	VPS36	MCO14	YIL108W	MOB1
YGL194C-A	YMR226C	OPI1	ACF2	SRS2
YGR122C-A	RFC3	NPR3	PRD1	RSF2
SNM1	CPR8	SNF6	SAS4	DPS1
DUR3	PPG1	MMO1	SSD1	NOG2
TCD1	YOL106W	MFT1	NPL3	
HTD2	DCI1	PUT4	CHD1	
ORC6	BUD26	YPR127W	YGL114W	
APL2	YBP2	TNA1	CAB4	

5.2 *Curriculum vitae*

Jan Großbach

Employment history

Since April 2015:	PhD student in the lab of Andreas Beyer, University of Cologne, Cologne, Germany
March 2014 - February 2015:	Research assistant in the lab of Andreas Beyer, University of Cologne, Cologne, Germany
February 2013 - February 2014:	Research assistant in the lab of Marcel Bucher, University of Cologne, Cologne, Germany
September 2008 - May 2009:	Mandatory civil service at the Weyertal-hospital in Cologne, Germany

Education

Since April 2015:	PhD in Computational Biology at the University of Cologne Supervisor: Prof. Andreas Beyer Title: **Identification and Validation of Regulatory Genetic Variation in Yeasts**
April 2013 - March 2015:	Msc. in Biological Sciences with a focus on computational biology at the University of Cologne, Cologne, Germany Thesis Supervisor: Prof. Andreas Beyer Title: **Detection of Epistasis and eQTL Mapping with Random Forest**
October 2009 - March 2013:	Bsc. in Biology Thesis Supervisor: Prof. Marcel Bucher Title: ***In silico* prediction and experimental validation of transcription factor binding sites of phosphate- and mycorrhiza-regulated genes in *Zea mays* and *Lotus japonicus***
June 2008:	General qualification for university entrance (Abitur)

5.3 Erklärung zu dieser Arbeit

Ich versichere, dass ich die von mir vorgelegte Dissertation selbständig angefertigt, die benutzten Quellen und Hilfsmittel vollständig angegeben und die Stellen der Arbeit - einschließlich Tabellen, Karten und Abbildungen -, die anderen Werken im Wortlaut oder dem Sinn nach entnommen sind, in jedem Einzelfall als Entlehnung kenntlich gemacht habe; dass diese Dissertation noch keiner anderen Fakultät oder Universität zur Prüfung vorgelegen hat; dass sie - abgesehen von unten angegebenen Teilpublikationen - noch nicht veröffentlicht worden ist, sowie, dass ich eine solche Veröffentlichung vor Abschluss des Promotionsverfahrens nicht vornehmen werde. Die Bestimmungen der Promotionsordnung sind mir bekannt. Die von mir vorgelegte Dissertation ist von Prof. Dr. Andreas Beyer betreut worden.