Imperial College of Science, Technology and Medicine
Institute of Clinical Sciences

Coordination between cell size and global gene expression

Amalia Martínez Segura

Submitted for the degree of Doctor of Philosophy in Imperial College London
September 2017

Abstract

All cells change size during the cell cycle, as they have to double their mass in order to produce two equally sized daughter cells. For biochemical reactions to work as intended in a changing volume, the reactants concentration should be stable. Data shows that the total mass of RNA and protein per cell are proportional to cell size. Expression data also shows a coordinated increase of the majority of the transcripts when there is an increase in average cell length [173], the mechanism behind it is still poorly understood.

In this study I took advantage of a previously described genetic model in fission yeast to investigate this question. This particular strain has a mutation in the *cdc2* gene that makes the gene product sensitive to a nucleotide analogue. When the drug is added, cells arrest in G2 and start growing in size without replicating their genome or dividing. Using transcriptomics and proteomics, I characterised how all transcripts and the majority of proteins respond to an increase in cell size. As previously described, most transcripts and proteins concentration is proportional to cell size. However, there is a subset of molecules that scape the global regulation that the rest of the genome is subjected to. Examining the features that makes this molecule circumvent the coordination of size and gene expression could reveal what is the mechanism behind it. I also applied a mathematical model to study how the cell allocates its limited resources at a maximal size.

Acknowledgements

I would like to thank, first of all, my supervisor Sam Marguerat. I am immensely grateful for all his support (and patience) during this PhD.

Thanks also to Vahid, our collaborator in the Mathematics department of Imperial College. Thanks for all the endless hours discussing the modelling. I would like to thanks Julius for starting the modelling project during his Masters.

I am also very grateful to Alex Schmidt for letting me spend a week in his lab learning everything there is to know about proteomics.

A huge thanks to all the members of my lab, past and present. I am so grateful for the open and friendly environment we have in the office, where everything can be discussed and there is always somebody ready to help.

I would like to also thank Manolo del Pino, my high school biology teacher. He was the first one to show me how fascinating biology can be.

And last but not least, I would also like to thank all my friends (specially Anne, Steven and Ángel) for putting up with all my crises with nothing but a smile and a cup of tea. Not forgetting to mention my family, an endless source of support and encouragement. They always believed I could do it even when I didn't.

Dedication

The author produced the work presented in this thesis, any work carried out by others has been acknowledged and appropriately referenced in the text.

'I have a friend who's an artist and has sometimes taken a view which I don't agree with very well. He'll hold up a flower and say "look how beautiful it is," and I'll agree. Then he says "I as an artist can see how beautiful this is but you as a scientist take this all apart and it becomes a dull thing," and I think that he's kind of nutty. First of all, the beauty that he sees is available to other people and to me too, I believe. Although I may not be quite as refined aesthetically as he is ... I can appreciate the beauty of a flower. At the same time, I see much more about the flower than he sees. I could imagine the cells in there, the complicated actions inside, which also have a beauty. I mean it's not just beauty at this dimension, at one centimeter; there's also beauty at smaller dimensions, the inner structure, also the processes. The fact that the colors in the flower evolved in order to attract insects to pollinate it is interesting; it means that insects can see the color. It adds a question: does this aesthetic sense also exist in the lower forms? Why is it aesthetic? All kinds of interesting questions which the science knowledge only adds to the excitement, the mystery and the awe of a flower. It only adds. I don't understand how it subtracts.'

Richard Feynman

Contents

List of Tables

List of Figures

5.6 Correlation between the posterior distributions of all possible parameter pairs. There does not seem to be any correlation between parameters, except the transcription rates of ribosomes and metabolic enzymes. Larger ribosomal pools require more precursors to be efficient, thus the correlation between the two. 3.12127

Chapter 1

Introduction

1.1 Overview

Even within a single multicellular organism, cell size extends orders of magnitudes, from micrometres-long lymphocytes to neuronal axons that spread across meters. In individual cells, size increases during the cell cycle until cells divide and give birth to smaller daughter cells. Cell size at division can vary, but is constrained in order to maintain size homeostasis in dividing cell populations. During development, cells change their shape and size to best adapt to their new function and form complex tissues. In unicellular organisms, cell size changes depending on environmental conditions. In conclusion, size is a dynamic feature that is always changing throughout the life of a proliferating cell.

Cell size increases during the cell cycle at the same time as biomolecules accumulate with cellular growth. The average size of a population is a function of the size at division and the growth rate of individual cells. Both vary depending on the external nutrient availability and other environmental factors, such as stress inducing conditions. Several molecular mechanism have been described to connect cellular growth and the environment. To ensure proper partitioning of cellular components during cell division, the regulation of cell size is also connected to cell cycle progression. All this information is integrated by the cells and produces as a result a constant size distribution in the population, despite individual variability. Two different

strategies have been described to maintain cell size homeostasis: adder and sizer. An adder would add a constant amount of mass each cell cycle, regardless of the initial birth size. In a sizer, cells have a mechanism that measures their size and all divide at the same volume. Different organism follow different strategies, sometimes even the strategy changes depending on the external conditions [141].

A changing cell size implies a variable cell volume that would affect concentration of molecules inside the cell. Changes in concentration could affect the equilibrium and rates of biochemical reactions, notably those related to gene expression. Strikingly, experimental data obtained for cells in different sizes shows a direct proportionality between cell size and the total amount of protein and RNA per cell [173]. Other experimental evidence also suggests that the majority of transcripts scale with cell size, accompanied by an increase in transcription rates. How, and if, these changes are translated into the proteome is yet undescribed. DNA amount is also related to cell size, as a higher ploidy goes together with larger cell sizes in yeasts and plants. Other cell types, such as cells in the fruit fly salivary glands, increase their ploidy to sustain higher synthesis capacities [46]. These findings suggest a relationship between cell size and the global regulation of gene expression.

Despite being such a fundamental question, the molecular mechanism between the coordination of cell size and gene expression has not been characterised yet. Gene expression is a intricate process, with a lot of steps that are highly regulated and could potentially exert a global control of RNA and protein amounts. Chromatin structure, transcription rates, splicing and protein/transcript degradation could all have an influence in the global regulation of gene expression.

During my PhD I have tackled these questions by characterising changes in proteome and transcriptome composition in *S. pombe* as cells increase in size up to a point when synthesis capacity becomes limiting. I have defined genes and proteins that escape coordination with cell size and used them to pinpoint potential regulatory paradigms of gene expression scaling. Finally, I have used coarse-grained modelling of cell physiology to identify processes that connect the cell synthetic capacity to cell size.

1.2 Cell cycle regulation in fission yeast

The driver of cell cycle progression in fission yeast is the cyclin-dependent kinase Cdc2, homologous to the metazoan CDK1. During the progression of the cell cycle Cdc2 associates with several cyclins, depending on the cell cycle stage [118]. It is the cyclin expression levels that change with the cycle phase while Cdc2 levels remain constant. At the start of G1, the activity of Cdc2 is inhibited by the protein Rum1 that accumulates during the previous mitosis [111]. However, when Cdc2 forms a complex with the cyclins Cig1 or Puc1, it becomes immune to the Rum1 inhibition and phosphorylates it [100, 43]. The phosphorylated version of Rum1 is quickly degraded, permitting the activity of the kinase-cyclin complex Cdc2-Cig2 [110]. Consecutively, the activation of this complex triggers the entry into S phase. During this phase, the genome is duplicated by the DNA polymerase. The Origin Recognition Complex (ORC) binds to the origins of replication and guides the activity of the polymerase through the replication of the genome. Cig2 is rapidly degraded after this event, to give way to Cdc2 to associate with the cyclin Cdc13. Cdc13 shows higher protein levels in G2-M compared to G1-S [23] and has been shown to be degraded after mitosis [168]. At the beginning of G2 phase the complex has a low activity due to a phosphorylation in tyrosine 15 by the kinases Mik1 and Wee1. Cdc25 is a phosphatase that removes this modification and activates the Cdc2-cdc13 complex [28]. Cdc25 accumulates as the cell gets larger, until its activity is enough to dephosphorylate Cdc2 and trigger mitosis [67]. The activation of Cdc2 increases the activity of Cdc25, forming a positive feedback loop. At the end of M phase, the daughter cells quickly degrade Cdc13 and the presence of Rum1 inhibits Cdc2, readying the system for another round of the cell cycle [16, 168, 111].

As described before, a complex array of proteins are needed for advance the cell cycle progression. Strikingly, a fusion of Cdc2 and Cdc13 has been shown to be sufficient to drive the cell cycle by Coudreuse and colleagues in 2010 [32]. The endogenous copy of *cdc2* was removed to impair the activity of the rest of the cyclins. A difference in activity of the fusion protein is what distinguishes the G1/S and G2/M transitions, providing the directionality of the cell cycle progression in the absence of the other cyclins [32].

1.3 Cell size homeostasis

Even though it is such a plastic trait, the size distribution in a population is constant in time. Without any mechanism setting size at which the cells divide, the distribution would change over time. Three mechanisms have been proposed to describe how the cell size distribution is maintained: timer, sizer and adder. In a timer mechanism, cells divide after a certain amount of time, regardless of their size at birth. If the increase in size is exponential, a timer can not lead to size homeostasis. A linear size increase, however, is compatible with cell size homeostasis in a timer scenario. If cells all divide at the same division size, then it is considered a sizer. The last model is an adder, in which cells add a constant amount of mass before dividing. All three mechanism are able to produce a size distribution that is stable in time [127]

Most bacterial cells have been shown to behave as adders. In other words, their size at division depends on their birth size, because they add they same amount of mass before dividing. This mechanism allows for cell size homeostasis passively, without an active molecular mechanism measuring size or the time it takes them to divide. A cell born with a size smaller than average will keep adding mass in every division, getting closer to the mean value each time [127]. When looking at single cell data in other organisms, however, sizer-like behaviour were found too, for example in fission yeast. When a sizer mechanism is involved, perturbations from the average cell size are corrected much faster than in an adder scenario. However, it requires a dedicated mechanism able to measure the size of the cell. Mammalian cells seem to follow a combination of the two models, with a sizer acting in G1 that is then followed by an adder after replication of the genome [150].

Tanouchi and colleagues in 2015 [143] proposed the noisy linear map model to describe single cell size data in a variety of organisms with different cell size homeostasis strategies. One of the advantages of this model is that it does not rely on assuming an adder or sizer behaviour, being able to represent both paradigms and intermediate behaviours. The model describes the dependence of the division size (L_F) on the initial size (L_I) of the cell as follows, with ν representing noise at cell size division:

$L_F = (aL_I + b + \nu)/2$

The parameter a represents the strength of cell size control and it has been named also memory term. When it is zero, the new born cells size does not depend on the initial one. This suggests a very tight molecular regulation of the division size. Thus, values of a close to zero represent sizer-like behaviours. If $a = 1$, the cells always add a constant amount b to the initial cell size therefore behaving like an adder. Applying this model to *S. pombe* single cell size data reveals an average a of approximately 0.6. *E. coli* and other prokaryotes present an a closer to 1, suggesting an adder behaviour compared to the more sizer-like one found in fission yeast [143]. Single cell data obtained from budding yeast also points to an adder behaviour, similarly to bacterial cells [135].

1.4 Cell size control mechanisms

The distribution of cell sizes in a population is a function of the individual growth rate and the rate at which the cells divide. Both parameters are very variable, changing depending on the environment. At the molecular level there are several signalling pathways taking information about the environment and translating it into changes in growth rate or proliferation. The pathways reviewed here are either conserved in fission yeast or specific to the organism.

1.4.1 TOR and Hippo pathways

The TOR (Target Of Rapamycin) signalling pathway was first discovered when identifying the molecular targets of the anti-proliferation effects of rapamycin. In mammalian cells the pathway is controlled by two different complexes, mTORC1 and mTORC2, that share the serine/threonine kinase subunit, TOR. The two complexes have slightly different roles, with the first one having been more studied in literature. The activity of mTORC1 is regulated by external growth factors, amino acid concentration, oxygen pressure and the energy status of the cell. All these inputs are integrated into the activity of mTORC1, that in turn phosphorylates

different effectors that change the behaviour of several components of the cell. Protein synthesis is directly activated through the addition of a phosphate in a translation initiation factor (eIF4E) and in the S6 kinase 1, that activates transcript and protein synthesis. mTORC1 also activates lipid synthesis and ATP production through glycolisis, as well as inhibiting autophagy. On the other hand, the second complex (mTORC2) only responds to growth factors controlling the organisation of the cytoskeleton and cell survival. In fission yeast, instead of one TOR subunit that is shared between the two complexes, there are two different kinases: *Tor2* (part of TORC1) and *Tor1* (part of TORC2). *S. pombe* cells change their average division size in response to their environment, growing larger in response to richer environments. Opposite to TORC1, TORC2 plays a role in the activation of quiescence when nutrients are low [124, 48].

Another pathway involved in the control of cell proliferation and death is the Hippo pathway, first described in *D. melanogaster*. The core of the pathway is the kinase MST1/2 (Hippo in fruit flies), that initiates a phosphorylation cascade leading to the degradation of the transcription factors YAP and TAZ. The activity of MST1/2 is affected by cell polarity, mechanical stress and growth factors. When activated, the pathway promotes cell death and differentiation in both fruit flies and mammalian cells. In turn, its inhibition activates cell proliferation. By regulating both proliferation and cell death, this pathway is key in setting the number of cells in a particular tissue. Evidence points to a connection between the Hippo and the TOR pathways through the activation of PTEN, an inhibitor of TOR, via the transcription factor YAP [51]. The core members phosphorylation cascade in Hippo are conserved in fission yeast, albeit regulating the exit of mitosis through the Septation initiation Network (SIN). SIN controls cytokinesis and septum formation, but its role in cell size control has not been described [53].

1.4.2 Fission yeast cell size control: Pom1 and its interaction with the cell cycle

Fission yeast is rod-shaped and grows elongating from the tips. Cellular growth is connected to cell cycle progression via two size-checkpoints: one in G1/S and another in G2. The G1/S checkpoint is considered cryptic when cells are growing exponentially, as newborn cells are

already above the cell size threshold. That is the reason why proliferating cells spend only 10% of their time in G1 in rich media. After G1, cells need to achieve a minimal cell size in order to proceed with the G2/M transition. The kinase Pom1 is transported to the tips of the cells, creating a concentration gradient towards the centre of the cell. Another kinase, Cdr2, is concentrated in the centre of the cell and is inhibited by Pom1. When cells are small, the local concentration of Pom1 in the middle of the cell is enough to inhibit the activity of Cdr2. As they grow in length, Pom1 concentration in the middle of the cell will keep decreasing until reaches a concentration in which Cdr2 is active. The activity of Cdr2 inhibits Wee1, a kinase that inhibits the entry into mitosis by phosphorylating the cyclin-dependent kinase Cdc2 [99, 113]. Pom1 was proposed to be the main regulator of cell size in fission yeast, however newer pieces of evidence challenged this model. Experiments by Pan and colleagues [120] disputed the notion of a Pom1 gradient in the cell, suggesting instead that the amount of Cdr2 in the centre of the cell is proportional to the surface area of the cell. In addition, Pom1 deletion mutants still show size homeostasis [164]. Cells with only one cyclin-kinase complex also show cell size homeostasis, without the influence of Wee1 and Mik2, and by extension Cdr2 [32, 164, 115]. Taking all this evidence together, it is likely that Cdr2 is only one part in a more complicated network controlling cell size in fission yeast [165, 113, 164].

Nutritional information relayed through TORC1 also influences cell cycle progression. Quality of external nitrogen sources is reflected in the activity of TORC1. A media rich in nitrogen leads to an increase in TORC1 activity. In turn, TORC1 activity lowers when the media is poor in nitrogen. TORC1 inhibits Ppk18, activating the PP2A-B55 pathway in high nitrogen conditions. PP2A-B55 activity prevents the activation Cdc2 through the activation of Wee1 and inhibition of Cdc25. In richer environments, fission yeast spends more time growing in G2 due to the activation of the PP2A-B55 pathway, delaying the entry into mitosis. On the other hand, when PP2A-B55 activity is low, the transition to mitosis is sped up producing smaller cells in low nitrogen conditions [26].

1.5 Overview of gene expression regulation

The first step in gene expression is transcription, in which genes are copied by an RNA polymerase into transcripts. Coding genes and long non-coding RNAs are transcribed by RNA polymerase II, whereas other non-coding transcripts (tRNA, rRNA, snoRNA) are produced by RNA polymerases I and III.

DNA is wrapped around histone proteins, forming chromatin. Histones can undergo several post-translational modifications in their tails, that influence transcript levels as well as how tight their binding to DNA is. Moreover, in multicellullar organisms nucleotides can be methylated adding another layer of transcriptional regulation. DNA methylation has not been shown to exist in fission or budding yeast, but histone methylation and acetylation do occur and have a proven regulatory role. Acetylation of histone in lysines 14 and 16 changes the net charge of the nucleosome and makes its union with DNA looser. More mobile nucleosomes correlate with higher transcription rates. Methylation of lysine 4 is a proxy for gene activation, whereas in lysine 9,27 and 36 inhibits gene expression [3, 75, 86].

Another modification with a very important role is H3K56 acetylation. It has been shown to play a role in regulating transcription globally during S phase. When the genome is copied, not all genes are duplicated at the same time. As the process goes along, early replicated genes would have twice the amount of template than the ones that are copied late. A higher availability of templates could result in higher expression levels for early replicated genes. However, experimental data does not show an increase, suggesting the existence a buffering mechanism. Transcription rates do increase when deleting Rtt109 from the genome, the enzyme responsible for H3K56 acetylation. Histones incorporated to the newly synthesised DNA have this modification inhibits transcription in early replicated genes. When S phase has been completed, the acetylation marks are removed by the activity of deacetylases [152, 151]. Transcriptional activation also depends on the positioning of the nucleosomes around the transcription start site. If a nucleosome sits in the transcription start site, it can inhibit the formation of the pre initiation complex and, therefore, the initiation of transcription. Most promoters lay in what has been named Nucleosome Free Regions (NFR), stretches of DNA just upstream the

transcription start site that are actively kept free from nucleosomes to permit the initiation of transcription. Several remodelling complexes have been shown to play a role in the positioning of nucleosomes and their eviction and the start of transcription, such as the SWI/SNF (Switch/Sucrose-Non Fermentable), ISWI (imitation switch), INO80 (Inositol-requiring) and CHD (Chromo-helicase/ATPase-DNA-Binding). Around this region, there are two nucleosomes, named -1 and +1, whose position is very stable [58]. Strikingly, some studies suggest that fission yeast does not show a well positioned -1 nucleosome, when compared to data obtained in budding yeast [78]. Higher resolution studies using sequencing do reveal the presence of a nucleosome structure upstream the NFR [136].

In coding genes PolII binds to the promoter region, together with general transcription factors (TFIIB, TFIID, TFIIE,TFIIF, TFIIH), to form the pre-initiation complex (PIC) [126]. These transcription factors are considered general, as they play a role in the transcription initiation of all the genes. Regulatory transcription factors control the transcription of a subset of genes, by binding to specific sequences in the promoter and affect the transcription rates of their targets. Transcription factors can have activating or repressing properties. Most transcription factors are formed by DNA-binding domains that recognise the binding sequence in the promoter and an activation/inhibition domain that affects the initiation of transcription. In activating transcription factor those domains are thought to interact with general transcription factors and enhance transcription. Repressors either impede the binding of the initiation complex, or competes with them for binding to the promoter reducing the rate of initiation. Example of specific transcription factors in fission yeast are *cdc10*, that controls the G1/S transition [160], and *ste11*, the master regulator of the stress response [102].

An important feature of transcription in eukaryotes is what has been termed PolII pausing. After the complex has been formed at the promoter, PolII complexes accumulate at the region ready to start transcribing. Although it might look like an inhibitory mechanism that reduces the number of elongating complexes, data has shown that highly transcribed genes have higher levels of PolII pausing. Compared to mammalian cells, where 41% of genes present higher levels of pausing, only 23% of loci show this feature in fission yeast. Ribosomal genes are enriched in this feature, corroborating the relationship between PolII pausing and expression

levels [18]. Another layer of regulation is the existence of post-translational modifications in the polymerase. Rpb1, the largest subunit of PolII, has tandem repeats in its C-terminal region or CTD (C-terminal domain). The CTD is subjected to post-translational modifications and acts as a platform for other proteins to bind to. Phosporylation of Ser5 indicate initiation, that changes from Ser5 to Ser2 when the complex starts elongating [19].

Paused complexes are released by Spt4-Spt5 and the recruitment of elongation factors jump-starts the elongation process and the gene is transcribed into an pre-mRNA molecule [60].

When PolII start transcribing the polyA signal, a set of termination factors bind to the polyadenylation signal present in the 3' end of the transcript, detaching the transcript from the genome. In yeast, two complexes are involved in the termination of transcription, CPF (Cleavage and Polyadenilation Factor) and CFIA and CFIB (Cleavage Factor). Pcf11 in the CFIA complex interacts with the Ser2 phosphorylated Rpb1, and together with the polyadenylation sequence finish recruiting the rest of cleavage complexes. Ysh1 cleaves the transcripts at the polyA site leaving place for Pap1 to create the polyA tail by adding adenine nucleotides. PolyA binding proteins (Pab1 and Nab2) coat the 3' end of the transcripts to protect it from degradation and enhance export to the cytoplasm. Pab1 have also been proposed to play a role in the regulation of translation once the transcripts are exported. After the transcripts has been release, PolII keeps elongating for a short stretch to then be detached [121]. There are two competing models at the moment explaining how the polymerase is released from the genome. Some evidence suggest that the union of the cleaving complexes produces conformational changes in the elongating complex that lead to its destabilisation and detachment [172]. Other publications propose what has been termed the torpedo model. When the transcript is cleaved, an exonuclease Rat1 starts degrading the nascent transcript until it collides with PolII, dissociating the complex from the genome [70, 159].

At the same time as elongation occurs, the transcripts goes through the splicing process in which introns are eliminated from the sequence. The process is carried out by the spliceosome, an RNA-protein complex that recognises specific sequences at the edges of introns and removes them from the transcript sequence [104, 41]. First, an adenosine in the intron, the branch

point, performs a nucleophilic attack on the first nucleotide of the intron. This forms the lariat intermediate that is still bound to the downstream exon. In the second step, the 3' end of the upstream exon attacks the last nucleotide of the intron. This releases the lariat, containing the intron sequence, and binds the two exons together [153]. The spliceosome is then recycled in another round of splicing [7, 98]. The splicing machinery is very conserved in fission yeast, with 47% of the genes containing at least one intron. Strikingly, in budding yeast only 5% of the genes have a single intron. Alternative splicing is very common in metazoans, affecting up to 90% of the transcriptome. It does occur in fission yeast, but those species are quickly degraded [15, 14].

Once the mRNA is in its mature form after removing all introns, it is transported to the cytoplasm for translation.

During translation, the ribosome will produce a polypeptide according to the sequence of the transcripts. Ribosomes are a ribonucleoprotein complex, formed by a small (40S) and large (60S) subunits. Initiation of translation is a highly regulated process, with 12 initiation factors involved. A methionine tRNA together with eIF2 (eukaryotic initiation factor 2) and GTP form the ternary complex. Subsequently, the ternary complex binds the small subunit and other initation factors (eIFs 1,1A,3 and 5) to assemble the pre-initiation complex. The 5' of the transcript is marked by the presence of the methylguanosine cap, where the pre-initiation complex binds accompanied by other initiation factors. Among them, cap and polyA binding proteins and RNA helicases. The complex scans the transcripts to locate the start codon. When the start codon is reached, eIF1 leaves the complex and triggers a tighter union of the pre-initiation complex and the transcript. After that, the large subunit binds to form a full complex and start translating [1]. The elongation factor eEF1A feeds tRNAs to the complex as it moves along the transcript. During the elongation, tRNAs interact with the transcript-ribosome complex to bring the necessary amino acids for the elongating peptide [87]. When the ribosome reaches the stop codon, a termination factor takes the tRNAs place. Binding of this factor releases the newly synthesised protein and the two subunits of the ribosome, ready to start another round of translation.

Another layer of gene expression regulation is molecule degradation, both for proteins and transcripts. The half-lives of mRNAs have been shown to be regulated and play an important role in the regulation of cellular processes, such as the cell cycle or stress responses [145]. For instance, transcripts involved in cell cycle progression are specifically degraded in certain phases of the cell cycle, ensuring that they are only present in the correct stage [39] . The first step in degradation is the removal of the 5' cap of the transcript in a process known as decapping. This is carried out by a complex formed by two proteins Dcp1 and Dcp2 [13, 38]. The length of the polyA tail is inversely correlated with the decapping rates, suggesting a role of Pab1 in inhibiting the activity of the Dcp1-Dcp2 complex [21]. Usually, decapping only occurs when the polyA tail has been shortened to a few nucleotides. A nuclease complex named PAN, consisting of Pan2 and Pan3 is responsible for starting the degradation of the 3' end of the transcript [17]. Ccr4 and Pop2 also form a nuclease complex that degrades the polyA tail [148]. Once the polyA tail and the 5' end cap have been removed, two complexes at each end finish the degradation of the transcript. Xrn1 has nuclease activity in the 5' to 3' direction [55], while at the same time the exosome degrades in the other direction [2]. Experiments in budding yeast have shown a global coordination between transcription and degradation rates, mediated by Xrn1 [139, 138].

Proteins are also subjected to degradation in the cytoplasm, carried out by the proteasome complex. The complex is formed by 33 distinct subunits, that form two complexes: the regulatory particle and the core particle. The regulatory particle selects ubiquitininated proteins for proteolysis by the core particle [42].

All these processes form an intricate network that control de abundance of each transcript and protein at a given point in time.

1.6 Coordination of gene expression and cell size

It has been described in fission yeast, bacteria and mouse fibroblast that the amount of RNA and protein per cell correlate with cell size [130, 44, 69]. However, the effect of size on individual transcripts was not described in these experiments. A handful of transcripts could

be dominating this increase in molecule numbers that matches cell size. In Zhurinsky et al. [173] microarrays were used to probe the whole transcriptome in different cell cycle mutants of fission yeast. The strains used were *cdc25-22* and *wee1-50*, thermosensitive mutants that at 32°C are larger or smaller than the wild type respectively. Surprisingly, looking at the whole transcriptome revealed a global regulation of the majority of mRNAs. Most transcripts adjust their number of molecules to the size of the cell, leading to the maintenance of homeostatic concentrations as the cell grows in size. To check if this was due to transcription, the same study also performed polymerase II ChIP-seq in both strains. Again, the occupancy of polymerase II correlated with cell size for the majority of genes, indicating a possible correlation of transcription rates too. Similar results are obtained when checking individual transcripts in single mammalian cells using smFISH (single molecule Fluorescence In Situ Hibridisation). Even multicellular organisms, such as *C. elegans*, follow this trend [119].

The study in Zhurinsky et. al [173] showed how the output of the genome is tuned globally to match the size of the cell. They also asked the question of what would happen to gene expression when cells reach extreme cell size . At this point it would be useful to introduce the concept of DPR (DNA to Protein Ratio) to measure the synthesis capacity of a genome. The lower the DPR, the more proteins a single copy of the genome has produced. Two thermosensitive strains, *cdc2-33* and *cdc10-M17*, can grow without replicating their genome when exposed to the restrictive temperatures. The proteins affected by the thermosensitive mutation are key for the G2-M and the G1-S transitions respectively. When they are inactive, for instance at the restrictive temperature, the cells arrest at G1 and G2. When arrested, the quantity of RNA and protein per cell increase at the beginning but after some time it reaches a plateau. In the plateau the DPR is at its lowest, raising the possibility that the cell has reached the limit of its synthesis capacity. It is interesting to note that tissues with a high synthesis capacity sometimes suffer endoreplication and have a higher ploidy that the organism to support this increase in molecule production. This phenomenon occurs, for example, in salivary glands in *D. melanogaster* [81].

In yeast and plants, diploid cells are twice as big as haploid cells [157, 137] The hypothesis at the moment is that more copies of the genome translates into more protein, increasing the

size of the cell. Despite having ample experimental evidence for this behaviour, the mechanism it has proven to be elusive as there are not many differentially expressed genes when ploidy changes [46, 167].

Organelles have been shown to be responsive to the size of the cell. A positive relationship between the size of the nucleus and the cell's length has been reported in fission yeast [116], resulting in a constant nuclear to cell size ratio. Moreover, when this ratio is perturbed using genetic modification or centrifugation, the difference is rapidly corrected. This suggests the existence of an active mechanism linking nuclear and cell size volume. Kume and colleagues [77] performed a genetic screening looking for mutants whose nuclear to cell size ratio is altered compared to wild type. Deletion of genes involved in transcript export and membrane synthesis were shown to produce an enlarged nucleus, suggesting a role of these processes in nuclear size homeostasis. The molecular mechanism, however, still remains elusive. An inverse correlation between cell size and nucleoli [155].

The relationship between the size of the nucleolus has been studied in *C. elegans* wild type embryos [155]. Mutants whose embryo size is different from wild type exhibit an inverse relationship: the size of the nucleoli decreases in large cells and increases in small ones. All embryos are born with the same number of nucleolar components inherited from the mother. Weber and colleagues hypothesises that the formation of the nucleolus depends on a phase transition. In other words, the separate components will only form nucleoli above a threshold concentration. As the number of starting components is the same, smaller nuclei will produce larger nucleoli due to an increase in concentration. An increase in nuclear volume would, in turn, reduce the concentration resulting in smaller or absent nucleoli.

The number of mitochondria are proportional to the size of the cell. Strinkingly, the activity of those mitochondria decreases as the cell grows in size [107]. The relationship between mitochondrial activity and size depends on the degree of mitochondrial connectivity. For instance, inhibiting mitochondrial fission results in an increase of membrane potential in large cells. The authors hypothesise that in large cells, the transport of metabolites and energy is limited by their size leading to lower membrane potentials. However, when connectivity is increased by

inhibiting mitochondrial fission, all the mitochondria act as a network that can overcome the transport limitations set by a larger size.

1.6.1 Possible factors involved in the coordination of cell size and gene expression

The coordination between cell size and gene expression affects the majority of transcripts, as well as their transcription rates. A potential mechanism behind this behaviour would be able to control the number of transcripts or proteins globally in order to adjust them to the size of the cell. Potentially, any step involved globally in gene expression could be responsible for this phenomenon. For instance, the number of PolII molecules in the nucleus could regulate transcription rates globally, the more polymerase the more transcripts would be produced. PolII subunits are produced in the cytoplasm and then imported by Iwr1 into the nucleus [33]. Iwr1 is a transporting molecule specific only for polII, providing a mechanism to control the local concentration of mRNA transcribing complexes and potentially affecting transcription rates globally. For transcription to advance, the polII complex need to undergo several post-translational modifications that are carried out by certain factors. Concentration of these factors could also affect transcription rates globally.

Chromatin modifications have also been shown to be able to affect expression levels. Acetylation of histones, for instance, influences global transcription rates in budding yeast [142]. The activity of the acetylating complex picNuA4 has been hypothesised to be untargeted, maintaining global levels of acetylation throughout the chromatin. [45]. Lysine acetylation of histone tails neutralises the positive charges that mediate the union between the DNA and the histones. Loss of the positive charges make nucleosomes more mobile, inducing transcriptional activity. Interestingly, the only acetylase that is essential (*mst1* in fission yeast, the catalytic subunit of the picNuA4 complex) is the one hypothesised to be responsible for most of the untargeted acetylation, highlighting the importance of this mechanism [45]. These enzymes use acetyl-CoA as a substrate, a metabolite produced in the cytoplasm or the mitochondria. The absence of enzymes that synthesise this metabolite has been shown to affect global acetylation levels and,

ultimately, transcription rates. Adding Acetyl-Coa to the medium seems to promote growth in *S. cerevisae* by acetylating the promoters of growth genes and increasing their expression levels [20]. Other study in budding yeast shows the importance of the enzymes that produce Acetyl-CoA in the nucleus. Without them, transcription is severely impaired globally [142]. Therefore, Acetyl-CoA concentration could provide a link between the metabolic status of the cell and transcription rates, being able to affect them through acetylation.

In addition, transcription rates correlate with mitochondrial activity in single cells, through ATP concentration [34]. Taken together with the evidence that mitochondrial activity is dependent on cell size [106], it could suggest the existence of another mechanism able to connect cell size with transcription rates. Molecular mechanisms able to control translation globally have been described in literature, with TOR being an example (see subsection 1.4.1). The availability of translation initiation factors could also affect translation rates, as these proteins have been shown to be one of the most abundant in the cell. Abundance of initiation factors set growth rate in bacteria by limiting translation [72].

1.7 Non-scaling proteins

The proportionality between the majority of transcripts and cell size has been well established empirically. Nevertheless, there is a subset of transcripts and proteins whose concentration does not scale with cell size. This property makes them suitable to act as concentration dependent triggers for cell cycle progression or differentiation. There are several examples in literature of molecules whose concentration is cell size dependent and that is fundamental for their regulatory properties.

The first example described was the role that histones [5] and replication proteins [29] play in the regulation of the midblastula transition (MBT) in *Xenopus laevis* eggs. Egg fertilization is followed by rapid synchronised cell divisions without cell growth, reducing the size of the cells. These divisions can be maintained during growth thanks to the maternally supplied proteins. After several divisions without growth, the cell cycle starts elongating and the cell divisions

stop being synchronised between cells. Over-expressing 4 proteins involved in replication in *Xenopus* eggs leads to a delay in the MBT, suggesting that a limitation in the replication machinery triggers the transition. As the cells divide rapidly without producing biomass, the number of these proteins per cell decreases [29]. Histones H3 and H4 are shown to inhibit transcription after fertilization, getting diluted as the cells divide. As their concentration decreases, the chromatin is more accessible to transcription factors activating MBT [4]. Budding yeast presents a similar dilution mechanism regulating cell size. In this case it is based on the ratio between two proteins: Whi5, a transcriptional inhibitor, and Cln3, a cyclin that controls G1 progression. As the cell increases in size during the cell cycle, the concentration of Cln3 stays constant. At the same time, Whi5 is produced in a cell size independent manner. Smaller cells start a higher concentration of Whi5, that gets diluted with growth. The decrease in inhibitor concentration initiates the progression through G1 to S and the apparition of a new bud in the cell [128]. Through this mechanism, cells that are born at a smaller size have to grow in size more to dilute Whi5 enough. In turn, larger cells already have a lower concentration of Whi5 and need to increase in size less to reach the trigger concentration.

Fission yeast regulates its size using the opposite mechanism. In this case, it is based on the accumulation of the protein Cdc25. Cdc25 is an activator of the G2/M progression, in opposition to the inhibitor Wee1. As the cell grows in size, the concentration of Cdc25 increases while Wee1 stays constant. Mitosis is triggered when the concentration of Cdc25 is above Wee1, providing a size sensing mechanism that depends on the concentration of Cdc25 [67].

Despite the examples described in literature, how these genes are able to avoid the global scaling of the rest of the genome is not yet understood.

1.8 Resource allocation in cells

The global adaptation of the number of transcripts and proteins per cell in response to changes in cell size, raises the question of how does the cell allocate its resources to allow this plasticity. The cell operates with a limited number of proteins and precursors that need to be used

efficiently when a change in cell size arises. Theoretical work by the Hwa laboratory has laid the foundations of a description of resource allocation in prokaryotes. Initially these studies started by trying to understand how cells adapt to their environment. *E. coli* cells are able to proliferate in several substrates, adapting their growth rate to the quality and quantity of the nutrients in the media. Growth rate and division size have an inverse relationship in this organism, the faster they duplicate, the larger the cells are. Not only the cells are larger, but the richer the media, the more resources are devoted to the production of ribosomes. When size grows exponentially, protein mass increases exponentially too so the number of ribosomes in cell must be sufficient to sustain this regime. Higher rates of protein production have a higher demand for amino acids. Protein synthesis rate is thus limited by the amino acid influx. At a certain growth condition, the cell must balance the amino acid flux with the ribosomal fraction to maximise the growth rate. In a poor environment, the supply of amino acids is low so the need for metabolic enzymes increases, leaving less resources available for producing ribosomes. In richer media, the cell can devote more resources to make ribosomes as it does not need as many metabolic enzymes, leading to an increase in division rate. This model is enough to explain the relationship between growth rate and ribosomal fraction, as well as the effect of the inhibition of ribosomal activity using antibiotics [131, 71, 72, 56, 130].

Studies in budding yeast, however, reveal different principles in the allocation of resources when duplication rate changes. The fraction of ribosomes in the proteome still correlates with division rate, but data suggest that there is a fraction of ribosomes that remain inactive [105]. These are hypothesis to be a mechanism to quickly respond to changing conditions in the environment. For instance, a sudden increase in nutrients would increase growth rate more rapidly if the ribosomes just need to be activated instead of produced from scratch. Manipulating the expression levels of a protein useless to the cell, such as GFP, has also been used as a tool to understand what processes are limiting in gene expression. Interestingly, the quality and amount of external nutrients has a huge influence in this. A environment low in nitrogen produces a decrease in the amount of amino acids that produces a limitation in translation elongation. In turn, transcription is limited by a decrease in phosphate concentration in the media [65]. All these results taken together show the intricacy of factors that influence how resources are allocated

in the cell, and how strategies may be different between eukaryotes and prokaryotes.

1.9 Fission yeast

Fission yeast, or *Schizosaccharomyces pombe*, is a unicellular eukaryote that belongs to the ascomycete group in fungi. The majority of laboratory strains come from a single isolate obtained by Urs Leupold from Delft, Netherlands [54]. However, wild strains have been isolated all over the world, generally from fermented beverages [59].

Fission yeast cells can exist both in diploid or haploid form. In haploid cells it is easy to assess if a mutation produces a phenotype, as there is only one copy of the genome. In turn, diploid cells can be used to test if mutant alleles are dominant or recessive.

Targeted genetic modification of yeasts is very simple, due to their capacity to integrate foreign pieces of DNA into their genome through homologous recombination. The variety of environments that fission yeast can thrive in allows for the exploration on how the external conditions affects cellular physiology and gene expression.

Fission yeast has been instrumental in dissecting several basic physiological mechanisms. It has been key in discovering the cell cycle and the underlying genetic mechanisms. After division, these rod shaped cells accumulate mass by elongating from the tips, which facilitates the assesment of the cells' age. Thanks to this cell cycle mutants were easily identified for further testing. The large size of *S. pombe* chromosomes has also prompted a great deal of research in their structure and dynamics. In addition, the structure of telomeres and centromeres is similar to mammalian cells, allowing discoveries that can be translated to metazoans. The relationship between chromatin modifications and gene expression has also been extensively studied. For instance, dissecting how heterochromatin is formed in the centromeres, the histone modifications associated and their effects in transcription [54].

Chapter 2

Material and Methods

2.1 Cell culture and strains

2.1.1 Cell culture

All strains are grown in a shaker at 32°C at 170 rpm except for thermosensitive strains, that are cultivated at 25°C. For each experiment, a single colony is taken from an agar plate onto liquid media (see 2.1 for specific media conditions for each strain). The cells are kept in exponential phase for 24 hours in liquid media before starting the experiment. For the time course experiment with the thermosensitive strain (*cdc2-33),* the cells were cultivated in Yeast Extract (YE) media at 36.5°C for 12 hours in a water bath with agitation at 170 rpm. For the *cdc2-asM17* experiments, cells were grown in Edinburgh Minimal Media (EMM) containing 2 µM of 1NM-PP1 (Santa Cruz Biotechnology) at 32°C for 12 hours. Cultures were harvested before the exposure to the restrictive condition and after at every hour. Several cell pellets were snap-frozen using liquid nitrogen for posterior transcriptomics, proteomics or western blot analysis. Additional cell pellets were fixed with 4% formaldehyde for 15 minutes for posterior ImageStream and single molecule FISH analysis, or with 70% cold ethanol for measuring DNA content.

Full genotype	Abbreviated genotype	Culture conditions	Origin
h- 972	*wt*	32°C	[85]
h- cdc2-33	*cdc2-33*	25°C	[173]
h- cdc2-M17as::bsdR	*cdc2-M17as*	32°C	[6]

Table 2.1: Full genotype of the strains used, as well as the name used to refer to them and the culture conditions.

2.1.2 Media recipes

YE media recipe is as follows: 0.5% Difco Yeast Extract and 3% Glucose dissolved in water and autoclaved. YES consists of the same recipe, with the addition of adenine, uracil, leucine, histidine and lysine at 250 mg/l.

The recipe of EMM and the different stock solutions used for its preparation can be found in table 2.2. When necessary, G418 is used at 250 mg/l. In the experiments performed with different nitrogen sources, the corresponding amino acid was added at a final concentration of 20 mM.

2.1.3 Transformation

The strain to transform is grown to 0.9 OD and 50 ml of the culture are centrifuged (3000 rpm, 3 minutes) to pellet the cells. The cells are washed once with sterile water and once with 0.1M lithium acetate solution in TE (LiAc-TE). The cells are finally resuspended in 250 µl of LiAc-TE. 2 µl of boiled salmon sperm DNA (10 mg/ml) and 1 µgr of the PCR fragment are added to 100 µl of the previous solution and mixed gently. After incubating at room temperature for 10 minutes, 260 µl of a 40% PEG solution in LiAc-TE are added. The tube is then incubated at 30°C for an hour. After that, 43 µl of pre-warmed DMSO are added and the sample is heat-shocked at 42°C for 5 minutes. After the heat shock, the cells are spun down again and washed

EMM2 medium	Amount per litre	Final concentration
Potassium hydrogen phtalate	3 g	14.7 mM
Na_2HPO_4	2.2 g	15.5 mM
NH_4Cl	5 g	93.5 mM
Glucose	20 g	111 mM
Salts stock (50x)	20 ml	
Vitamins stock (1000x)	1 ml	
Minerals stock (10000x)	0.1 ml	
50x Salt solution	**Amount per litre**	**Final concentration**
$MgCl_26H_2O$	52.5 g	0.26M
$CaCl_22H_2O$	0.735 mg	4.99 mM
KCl	50 g	0.67 M
Na_2SO_4	2 g	14.1 mM
1000x Vitamin solution	**Amount per litre**	**Final concentration**
Pantothenic acid	1 g	4.20 mM
Nicotinic acid	10 g	81.2 mM
Inositol	10 g	55.5 mM
Biotin	10 mg	40.8 µM
10000x Mineral solution	**Amount per litre**	**Final concentration**
Boric acid	5 g	80.9 mM
$MnSO_4$	4 g	23.7 mM
$ZnSO_47H_2O$	4 g	13.9 mM
$FeCl_26H_2O$	2 g	7.4 mM
Molybdic acid	0.4 g	2.47 mM
KI	1 g	6.02 mM
$CuSO_45H_20$	0.4 g	1.60 mM
Citric acid	10 g	47.6 mM

Table 2.2: Composition of the supplement solutions for preparing EMM media.

once with sterile water. The cells are then spread in non-selective plates and incubated until a lawn grows. Using replica plating, the lawn is transferred to plates with the corresponding selective conditions. Cells in which the integration of the fragment was successful will produce colonies in selective media. Integration locus of the PCR fragment is checked using colony PCR to confirm positive colonies.

2.1.4 Mating and spore analysis

The two different strains need to be of opposing mating types. The two strains are patched on a malt extract plate (3% Difco malt extract, 2% Difco agar) and incubated at 25°C for three days. At this stage, the two strains have formed curved asci that contain four spores produced by meiosis. Spores can be analysed in two different ways: tetrad dissection or random spore analysis (RSA). In the case of tetrad dissection, asci are dissected in a Singer MSM 400 Dissection Microscope to separate the spores and analyse the phenotypes independently. For RSA, the sample is treated to remove vegetative cells, incubating the cells at 50°C for 30 minutes. Spores can resist this conditions, and are then plated in selective conditions for the phenotype of interest.

2.2 RNA-seq

Total RNA was extracted with the hot-phenol technique to produce total RNA [93]. Briefly, RNA is extracted using phenol-chloroform at 65°C and undergoes two purifications isoamyl alcohol and chloroform, to be resuspended in RNAse free water. RNA samples were first DNAse treated (TURBO DNA-free Kit, Ambion) and depleted of ribosomal RNA using Ribo-Zero Gold rRNA Removal Kit for Yeast (Epicentre, Illumina) following the manufacturer instructions. Sequencing libraries were prepared using the TruSeq Stranded mRNA LT Kit (Illumina). This protocol takes the mRNA after the depletion and produces a library with adaptors ready for sequencing. First, the mRNA is fragmented at a high temperature in the presence of divalent cations. Then, using reverse transcriptase and random primers the first strand of the cDNA is synthesised. In the second strand synthesis step, dUTP is used instead of dTTP and is incorporated in the second strand. The cDNA fragments have a single A at both ends that allows the posterior ligation of the adapters. These adapters are used to PCR amplify the cDNA library in the next step. In the amplification, strands with dUTP cannot be amplified using DNA polymerase. That way, only the strands produced in the first-strand synthesis step are amplified, producing a stranded library. Library are run on a Bioanalyzer 2100 for

quantification and to check that libraries are the correct size.

A pool of all the samples was sequenced in one lane in a HiSeq2500 (Illumina) platform.

2.2.1 Raw reads processing

The raw data files are reverse complemented using the software FASTX-toolkit (version 0.0.13, http://hannonlab.cshl.edu/fastx_toolkit/). The full genome sequence was downloaded from pombase.org [166], using assembly ASM294v2.22. The fasta file containing the genome was transformed into an indexed file usable by mapping software with the samtools index command [88]. Then the reads were aligned to the genome usig TopHat2 (version 2.0.8) with standard parameters except library-type fr-secondstrand [146]. This parameter is used to indicate that they library was prepared with a stranded protocol that produces reads that are reverse complementary to the mRNA sequence.

Read counting was performed using an in house R script. In summary, the script counts how many reads have been mapped to each of the *S. pombe* genes. When a read maps to overlapping genes, it is assigned randomly to one of them. The output of the script is a table with the read count for every gene in every sample for the sequencing. Gene coordinates come from the GFF3 annotation file (version ASM294v2.26) downloaded from pombase.org.

Total RNA extracted from the cells was first depleted of ribosomal RNA using a commercial kit. Samples were then used to construct stranded RNA-seq libraries. Sequencing data was processed to obtain how many reads map to each gene, as described in Chapter 2. In order to compare different libraries, these counts need to be normalised as total library size affects the number of counts per transcript. There are several normalisation methods that deal differently with this issue, traditionally the most used is converting the reads to RPKM (Reads Per Kilobase Per Million). This methods corrects for differences in library size and gene length, using the following formula $RPKM = \frac{10^9 C}{NL}$, where C is the number of reads mapping to a gene, N the total number of reads in the library and L the gene length in base pairs. Other RNA-seq processing pipelines offer their own solutions, such as DESeq2 and edgeR. The one

included in the DESeq2 package [92] has been described as the best dealing with differences in sequencing outputs [36], so it is the one used for the all the analysis and plotting of the data. It is based on the assumption that most of the genes are not changing between samples and calculates a scaling factor that brings all the libraries to the same average size.

2.2.2 Differential expression

The main experiment in this study produced a time course dataset that cannot be analysed using the conventional methods (DESeq2 [92], edgeR [103]) as the samples are not independent. First of all the data needs to be normalised, to remove the effect of differences in library size. I used two different methods made for time course expression data, MaSigPro [117, 30] and BETR [8], and collated a gene list with the intersection between the results of the two algorithms.

2.2.3 Splicing efficiency calculations

The pipeline used is based on the one outlined in [122]. Briefly, the junction reads produced by TopHat are annotated using the regtools (https://github.com/griffithlab/regtools) command junction-annotate, producing a BED file with a quantification of all the transreads per gene. At the same time, intronic reads are quantified with the bedtools multicov command (http://bedtools.readthedocs.io/en/latest/). Splicing efficiency is then calculated dividing the number of transreads by the sum of transreads and intronic reads. This will give a number between 1, all the reads found correspond to spliced mRNA, and 0, where no splicing is detected.

2.3 Proteomics

2.3.1 Sample preparation

Approximately 10^7 cells per time point were processed to obtain whole cell extract. Pellets were resuspended in lysis buffer (1% sodium deoxycholate, 1% ammoniumbicarbonate) and

disrupted in a FastPrep®FP120 Cell Disrupter (Thermo Fisher Scientific) for 8 cycles of 30 seconds at a speed of 6 m/s. Samples were nuclease treated using a mix consisting of micrococcal nuclease (S7 Nuclease) and Serratia marcescens (Pierce/Thermo) in a ratio 1:100 and incubated at 37°C for 30 minutes. Disulphide bonds in proteins are reduced using 5 mM TCEP for 15 minutes at room temperature and alkylated with iodoacetamide 10 mM for 30 minutes at 25°C in the dark. The reaction is quenched with N-acetyl-cysteine 12 mM for 10 minutes. Samples are quantified using the BCA Protein Assay Reducing Agent Compatible (Thermo Fisher Scientific). Before digestion with trypsin, a panel of standard peptides (see table 2.3) was spiked in the extract at a concentration of 50 femptomol of each peptide per microgram of cellular extract. These standards have the same sequence as some selected S. pombe peptides, but with lysines substituted by their heavy isotope version. Extracts undergo a double digestion: first with Lys-C (Wako chemicals) at a ratio 1:200 for 4 hours at 37 C and then overnight with porcine trypsin at 1:100 at 37°C. Samples are spun down at 14000 rpm at 4°C after addition of 1% TFA to precipitate the detergent. Peptide solutions are then dried prior to mass spectrometry analysis.

2.3.2 Mass spectrometry

50 micrograms of digested samples were resolubilised in 50 l of freshly prepared 0.1% trifluoroacetic acid (TFA) by mixing at room temperature for 10 minutes. A Thermo Scientific Ultimate 3000 Nano liquid chromatography system was used to separate peptides prior to mass spectrometry 7 analysis. An injection volume of 6 µl was removed from the insert vial and loaded onto a trap column. (Thermo Scientific Acclaim Pepmap 100, 100 µm internal diameter and 2 cm length, C18 reverse phase material with 5 µm diameter beads & 100 pore size) at 8 µL/min in 98% water, 2% acetonitrile, 0.05% TFA. The trap column was connected in-line to a Thermo Scientific Easy spray source which contained an analytical column (Thermo Scientific Acclaim Pepmap RSLC, 75 µm internal diameter and 25 cm length, C18 reverse phase material with 2 µm diameter beads & 100 pore size). Peptides were eluted using a ramped gradient Peptides were eluted using a ramped gradient with conditions: initial 5 minutes with 4% sol-

Gene ID	Common name	Peptide Sequence
SPAC23A1.10	tef102	STTTGHLIYK
SPBC1815.01	eno101	AADEFLLK
SPBP22H7.08	rps1002	AIHQALFQQGVLVAK
SPAC1F8.07c	SPAC1F8.07c	ATDIAAELISK
SPAC1F7.13c	rpl801	GVVGIVAGGGR
SPAC3A12.10	rpl2001	ATGEIVAINEISEPK
SPCC191.02c	SPCC191.02c	IIEGNDVEGVLAIR
SPCC622.09	htb1	LILPGELAK
SPAC26F1.03	pda1	IDVPSTEIEVTK
SPCC576.11	rpl15	FNNSPQHATWLR
SPCC1827.03c	SPCC1827.03c	ALVAPSLNAELSFSELR
SPBC11G11.03	mrt4	VLTLAQTEK
SPBC16H5.08c	SPBC16H5.08c	STFLESVAAR
SPCC584.04	sup35	AAPFIPSFQR
SPAC22F8.06	pam1	ITTDAFTSAGER
SPAC1834.01	sup45	FFDEISLDSGK
SPAC16E8.15	tif45	SETIEFSAHEDSSK
SPAC20H4.03c	tfs1	NEVVATEELLK
SPBC28F2.12	rpb1	GEDDLTHK
SPAC1F7.01c	spt6	VAEGSYQHIDVLELEK
SPAC27D7.14c	tpr1	SGGNILGFLGK
SPAC26A3.12c	dhp1	SVETQSTEVVTSSK
SPAC2G11.14	taf1	GLSINNLEELAK
SPBC216.07c	tor2	SFLPDLFK
SPCC132.02	hst2	LLGWSDELEK
SPBC3D6.03c	trz3	LCAGEAVLSK
SPAC6B12.19	rsa3	LQLLIESLR
SPBP23A10.04	apc2	AAELLDQPK

Table 2.3: Sequence and gene names for the peptide standards used in the proteomics experiment.

vent B (96% solvent A), then a 90 minute gradient 4-55% B, then a 10 minute isocratic at 90% B, then 5 minute isocratic at 4% B to equilibrate the column for the next sample. Solvent A is composed of 98% water, 2% acetonitrile and 0.1% formic acid and solvent B composition is 20% water, 80% acetonitrile and 0.1% formic acid. Mass spectrometry analysis was carried out in a Q Exactive (Thermo Scientific) equipped with an electrospray source using a Top12 HCD approach.

2.3.3 Peptide identification and quantification

Peptides then underwent mass-spectrometry (MS) analysis to produce a quantification of 33% of the total proteome. Raw files were loaded into Progenesis QI software (Waters Corp., v2.5). Features were mapped in each spectrum, then aligned between multiple raw files. Normalisation using the Progenesis algorithm was performed.This software aligns all the MS runs to correct for differences in the liquid chromatography (LC) separation prior to the analysis (see Methods). It then creates an aggregate dataset that contains all the peaks found in all the samples and compares all samples to this standard. This produces a dataset with no missing values, as every peak is analysed through all the samples. The data also needs to be normalised in order to correct for differences in the runs. In this case the software chooses a standard and applies different size factors so that the averages of all the runs are similar to the standard.

Features were exported in .mgf format for peptide and protein identification. Peptides and proteins were identified using Mascot Daemon v2.5 to search the *S. pombe* FASTA database downloaded from Uniprot on 2014.09.03. Mass tolerance was set to 5 parts per million and 25 millimass units. Carbamidomethylation in cysteine and deamidation were set as a fixed modification and oxidation in methionine as variable, allowing for 2 missed cleavages. For the dataset with acetylated and methylated peptides, those modifications in lysine and arginine were set as variables as well. Resultant XML files were exported from the Mascot server and loaded back into Progenesis for relative quantification with an FDR of 1%. Posterior data analysis was carried out using R 3.3.1 [123].

2.3.4 Absolute quantification of the proteome

The data obtained with conventional proteomics methods is relative, representing the proportion of each protein in the proteome but not the absolute amount per cell. When cell size increases, the majority of the proteome is hypothesis to increase with it coordinately, which is impossible to observe using relative data. This caveat can be circumvented using peptides standards at a known concentration. These peptides are the same as some selected *S. pombe*

peptides but some of their atoms are substituted by their heavy isotopes, allowing for absolute quantification of its light counterpart. The standards were chose so that they would span the range of protein expression described in [96], in order to use them to estimate the absolute quantities of the rest of the proteome. Samples with heavy spikes are quantified using PRM (Parallel Reaction Monitoring), a targeted mass spectrometry approach. Only the peptides of interest, both light and heavy, are measured using this technique. The ratio between heavy and light peptides is then used to calculate the amount of light peptide found in the sample. Combining this amount with quantification of the total protein per cell, provides an estimation of the copies per cell of these proteins. The absolute amount of these proteins is proportional to the intensity measured in the proteomics experiments. A calibration curve can then be extrapolated from their relationship and used to estimate the absolute amount of the rest of the proteome. There is a good agreement between the dataset produced in this study and the one in [96], with a Pearson correlation of 0.63 . As expected, the amount of protein per cell also increases with size when comparing initial and final time points and averaging the three biological replicates.

2.4 ImageStream Analysis

Samples fixed with 4% formaldehyde and stained with calcofluor were analysed using ImageStream (Amnis). The instrument was configured using the 405 nm laser, collecting the data in channel 1 (fluorescence) and using channel 3 for the bright-field images. 30000 cells per sample were acquired. Resulting images were analysed using the IDEAS v6.0 software. First, cells out of focus are excluded using the Gradient RMS parameter. In order to remove curved cells that might affect length measurements, images are also filtered using the Aspect Ratio Intensity parameter.

2.5 Flow cytometry analysis for nuclear content

Samples fixed with ethanol are resuspended in 50 mM sodium citrate and treated with 0.1 mg/ml of RNAse A at 37 C for two hours. DNA is then stained using 1 M SytoxGreen (Life Technologies). Stained cells were analysed using a BD LSR II, acquiring 10000 events per sample. The data was analysed using the R package flowCore [50].

2.6 Western blot

Whole cell extracts were prepared resuspending 10ˆ7 cells in Cell Lysis Buffer (New England Biolabs) buffer with protease inhibitors. The cell wall was disrupted using a FastPrep®FP120 Cell Disrupter (Thermo Fisher Scientific) at 6.0 m/s for 3 pulses of 30 seconds. Antibodies against acetylated histone 3 on both lysine 9 and 14 (Upstate-Millipore, 06-599) and against total histone 3 (Abcam, ab1791) were used. ImageJ (W. Rasband; National Institutes of Health, Bethesda, MD) was used to analyse the images.

2.7 Single molecule RNA-FISH (smFISH)

The protocol is based on the ones presented in [147], as well as the Stellaris yeast protocol. 10^8cells are fixed with formaldehyde for 15 minutes, and washed in buffer B (10 mM $KHPO_4$pH 7.5, 1.2M sorbitol) three times. The cell wall is then digested at 37C in spheroplasting buffer with Zymolase T100 (Zymoresearch) until 90% of the sample is digested. The spheroplasting buffer composition is as follows: 10 mM $KHPO_4$ pH 7.5, 1.2M sorbitol, 20 mM vanadyl ribonuclease complex (New England Biolabs) , 14.3 mM beta-mercaptoethanol, 1 mg/ml Zymolase T100. The digestion makes hole in the cell wall without reaching the spheroplast stage, allowing the probes to get to the cytoplasm. The pellet is then washed in buffer B and resuspended in 70% ethanol. After incubating at 4°C for 4 hours, the pellets are washed 3 times in buffer B and once in wash buffer 1 (2x SSC, 10% formamide). The pellet is finally resuspended in

50 µl of hybridisation buffer (4x SSC. 0.2 mg/ml BSA, 20 mM vanadyl ribonuclease complex). Simultaneously, 1 µl of each probe are mixed with 2 µl of yeast tRNA (10 mg/ml) and 2 µl of salmon sperm DNA (10 mg/ml). This mixture is dried in a speed-vac concentrator (Thermo Scientific Savant SPD121P) and dissolved in 50 µl of buffer F (1 mM $NaHPO_4$ pH 7.0, 20% formamide), incubated 3 minutes at 95°C. The probes and the cells are mixed in the same tube and left at 37°C overnight. The cells are then washed once with wash buffer 1, incubated for 15 minutes at 37°C, washed with wash buffer 2 (2x SSC, 0.1% Triton) and resuspended in PBS. The suspension is mounted on a slide with ProLong® Gold Antifade Mountant with DAPI (Thermo Fisher Scientific) and left to cured at room temperature for two days before imaging. Image acquisition is performed in a Leica SP8.

The resulting images are then analysed using FISH-QUANT [114], a Matlab package that allows automatic counting of the spots per cell.

2.8 Data analysis

The majority of the data analysis was carried out in R v.3.3.1 [123] , using Bioconductor packages [47]. The packages used, with their version number and reference paper can be found in table 2.4

2.9 Mathematical model

The ode15s solver from Matlab R2015a was used to simulate the system of ordinary differential equations that the model is based on. For the parameter fitting, I used the kindly provided SMC-ABC algorithm from Anthony Bowman and Vahid Sharezaei. The algorithm was implemented in Julia. The matlab library from Julia was used to call the simulations from the parameter estimation algorithm.

Package name	Version	Reference
ggplot2	2.2.0	[161]
gplots	3.0.1	[154]
pheatmap	1.0.8	[74]
DESeq2	1.12.4	[92]
BETR	1.28.0	[8]
MaSigPro	1.44.0	[30, 117]
plyr	1.8.4	[162]
dplyr	0.5.0	[163]
corrplot	0.77	[156]
zoo	1.7	[170]
rtracklayer	1.32.2	[79]
GenomicRanges	1.24.3	[80]
Rsamtools	1.24.0	[112]
GenomicAlignments	1.8.4	[80]

Table 2.4: R packages used for data analysis, with their version number and reference paper.

Chapter 3

Analysis of fission yeast transcriptome and proteome during growth and DPR limitation

3.1 Introduction

Cells double their mass to divide into two daughter cells. At a changing volume, the cell needs to maintain homoeostatic concentration of biomolecules for reactions to work correctly. This has been an object of study since the 1970s, specially for molecules involved in gene expression. It has been observed that the total RNA and protein per cell is proportional to the size of the cell [173, 129]. Moreover, when looking at the behaviour of individual transcripts using genome-wide expression techniques, the majority of transcripts are proportional to the population average length size in fission yeast. Thus, the homeostatic concentration of the majority of molecules is maintained. A similar trend can be observed in single cells using RNA-FISH techniques that allow the quantification of individual transcripts in mammalian cells [119] and fission yeast (data not published). Polymerase II occupancy also correlates with the average cell size, suggesting a scaling of transcription rates [173]. Despite being a pervasive phenomena that affects most transcripts, the biological mechanisms behind this phenomenon

are still unknown.

To tackle this problem, I will take advantage of a series of fission yeast strains that arrest and keep growing without replicating their genomes. Zhurinsky et al. 2010 [173] showed that the total mass of proteins and RNA molecules per cell increases as these strains grow in size. However, after some time growing, the amount of biomolecules per cell reaches a plateau. At that stage, the cells have reached the limit of what a single genome can produce. In other words, their DPR has reached a lower limit. Having cells growing in size without cycling, allows the study of the effects of cell size without the interference of cell cycle progression. In addition, we can see the effects of a low DPR on gene expression and biomolecule synthesis. Molecules that change in concentration in response to an increase in cell size are potential candidates to be involved in the coordination of cell size and gene expression. In order to describe the effects of cell size in transcripts and proteins in an unbiased approach, I chose to study these strains as they grow using transcriptomics and proteomics.

This chapter aims at describing the behaviour of both transcripts and proteins as the cell grows in size and when the genome becomes limiting, together with the dynamics of the cell size increase at both the population and the single cell levels.

3.2 Results

3.2.1 Characterisation of the strains used in this study

I started by characterising the changes in cell size in the strains used in Zhurinsky et al. 2010 [173], *cdc233* and *cdc10-M17*. The *cdc2-33* strain has a point mutation in the gene *cdc2*, which renders the protein temperature sensitive. At 25°C the protein retain its activity and the cells are similar in size to wild type strains. When the temperature is increased to 36.5°C, the protein loses its activity. Cdc2 is a kinase that triggers entry into mitosis by phosphorylating a series of substrates [140]. Inhibiting its activity stops cell cycle progression, arresting the cells in G2 before entering mitosis. The second strain used in the study, *cdc10-M17*, has a thermosensitive

allele of the gene *cdc10*. This gene produces a transcription factor that regulates the transition from G1 to S phase. Cells stop dividing and arrest in G1 when this protein is not active at higher temperatures. Despite the arrest, the cells keep growing until they reach their lowest possible DPR [173] in both strains.

Both strains were grown in the restrictive conditions for 12 hours, collecting samples before and 12 hours after the induction. From those samples, cell size was quantified using bright-field images from ImageStream. After the arrest, the *cdc2-33* strain reaches larger size (26.3 +/- 3.89 µm) compared to the *cdc10-M17* (21 +/- 2.91 µm) . The change in maximal cell size could be due to a difference in genome content. The *cdc10-M17* strain is arrested in G1, it has 1C genome content compared to the *cdc2-33* strain that arrests after S phase. This supports the hypothesis that the synthesis capacity of the genome is ultimately responsible for the maximal size a cell can attain. Two copies of the genome have twice as many templates for transcription, increasing the potential output in terms of molecules. The difference in maximal size suggests that the limitation in size is not due to physical constraints. For example, we could argue that it is physically impossible for the cell physiology to sustain sizes larger than wild type. If that were the case, both strains would reach the similar limiting size regardless of their genome content.

I hypothesised that a larger size would elicit larger effects in gene expression. Larger changes in transcript and protein number would be more detectable. Therefore, I chose the *cdc2-33* strain to perform further experiments. A proteomics and transcriptomics dataset was produced for the *cdc2-33* strain growing at the restrictive temperature for 12 hours. As the cells are exposed for a long time to a higher temperature than the physiological one, I checked the overlap between differentially expressed transcripts and the stress response described in Chen et al. 2003 [25]. The overlap between the two lists is represented in figure 3.1 as a Venn diagram. There is an statistically significant overlap (p-value $<2.2 \times 10^{-16}$) between differentially expressed transcripts and the stress response. The significant intersection between the response to stress and to a decrease in DPR obscures the difference between the two effects. I therefore carried out the same experiments using an analogue sensitive strain, described and optimised in [6, 37]. The strain contains a mutant allele of *cdc2* that makes the protein sensitive to a nucleotide

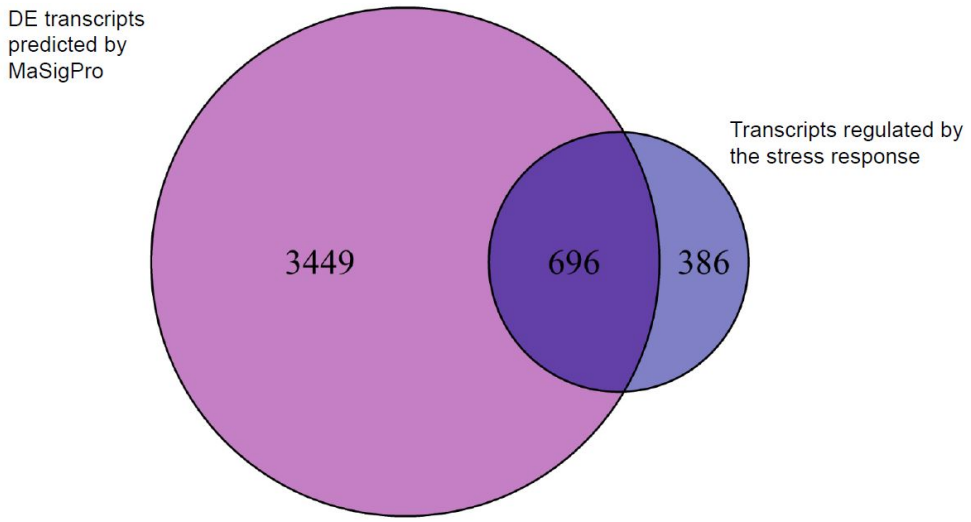


Figure 3.1: Differentially expressed transcripts from the time course data in the strain *cdc2-33* were identified using MaSigPro [30]. The cells were grown at 36.5°C, a temperature known to provoke a stress response in the transcriptome. To identify the extent of this response in the experimental system studied, I calculated the overlap between DE transcripts and the stress response described in [25] . Using the exact Fisher's test, the overlap between the two lists is significant (p-value $<2.2 \times 10^{-16}$). The increase in temperature that leads to the arrest has an effect in the transcriptome that is combined with the effect of a increased cell size.

analogue (1NM-PP1). The arrest is triggered by the addition of a drug (1NM-PP1) that binds to a mutated version of *cdc2* and inhibits its activity selectively, thus avoiding the stress due to the increased temperature. As figure 3.2 shows, the strain shows a similar behaviour as the temperature sensitive mutant, reaching a limiting size at 9 hours. The analogue sensitive strain also reaches a higher limiting size compared to the temperature-sensitive mutant (figure 3.2B). An increase in protein degradation caused by the restrictive temperature could be partly responsible for the difference in maximal size between the two strains. The stress response is characterised by an upregulation of the proteasome, a protein degradation complex. At constant synthesis rates, a higher degradation rate would lead to a lower protein amount in equilibrium. Because of the absence of heat shock and the larger minimal size, the *cdc2-M17as* strain was used for the rest of the genome-wide experiments.

To confirm the data obtained using ImageStream, time lapse imaging of living cells was also performed. Traces for 11 single cells are plotted against time in figure 3.3. Single cell data shows that length increases exponentially with time, up to a point where the cells stop growing in length. Although growth in length seems to stop abruptly, cell volume shows a more moderate

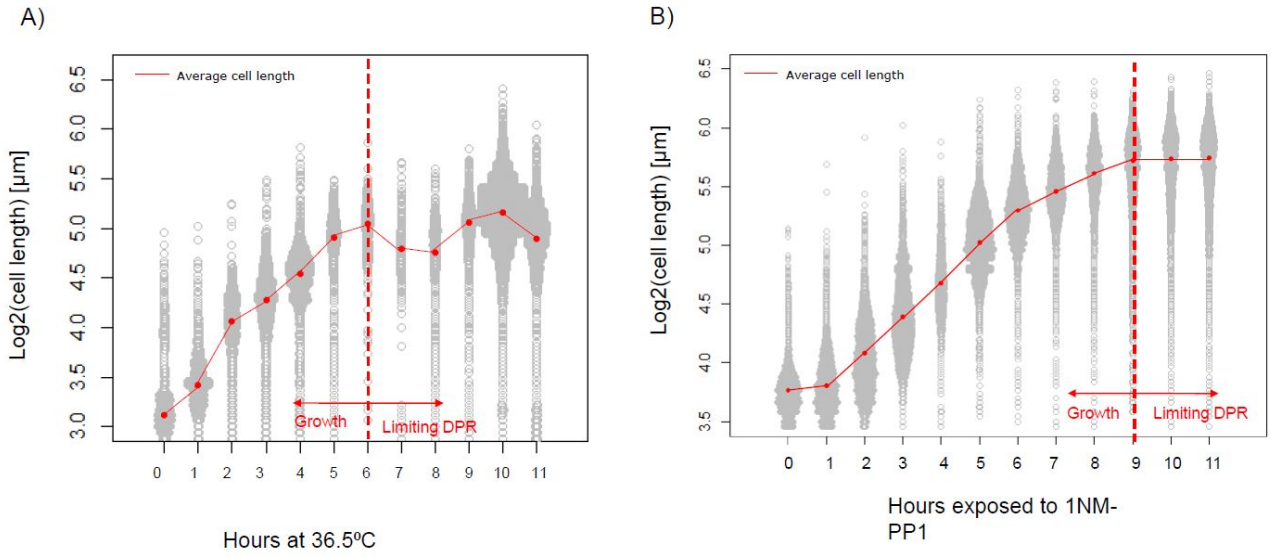


Figure 3.2: A) Progression of cell length in the *cdc2-33* strain when exposed to the restrictive temperature (36.5°C) for 11 hours. Cell size was measured using bright-field images in ImageStream. Each grey dot represents a single cell, with the red dots being the median of each sample. B) Single cell size data during the time course for the *cdc2-M17as* strain when exposed to the nucleotide analogue for 11 hours. Similarly to panel A, size data was obtained using ImageStream bright-field images. Each grey dot is the length of a single cell. Red ones are the median of each time point.

slowdown, pointing to a increase in width when the cell ceases to elongate. The potential inability of the cell to elongate could suggest the existence of a physical limit on the length of the cell. This could be due to, for example, a collapse of the microtubule structure that transports polarity factors to the tips. The arrest in growth seems to be more abrupt than the one observed in the population average data. Although it might seem contradictory, as the cells seem to slow down growth before stopping when measuring the total population. However, as the cells stop growing at different sizes, when averaged it would produce the trend observed in ImageStream.

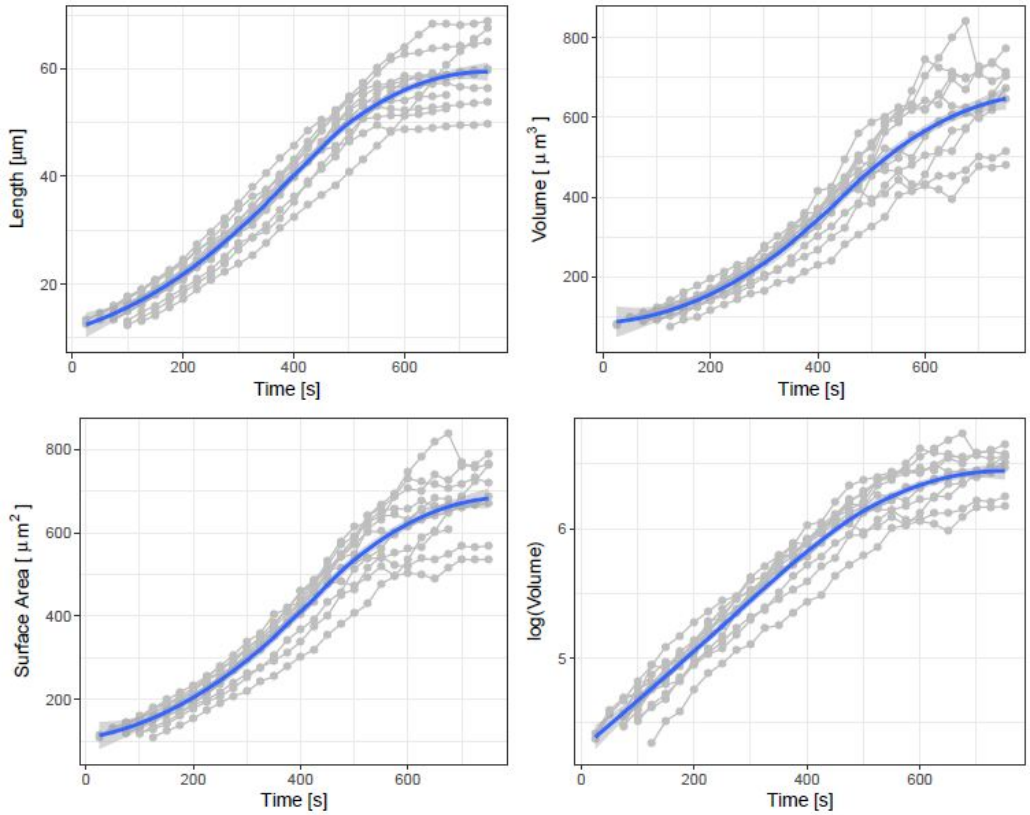


Figure 3.3: Single cell traces for the *cdc2-M17as* strain growing in the presence of 1NM-PP1, the nucleotide analogue that inhibits *cdc2* activity and arrests the cells in G2. Each panel represents the cell length, volume, surface area and the logarithm of the volume, respectively. Brigh-field images were processed using a custom MATLAB script to quantify the aforementioned size features of the cell. Grey lines are traces for individual cells, whereas the blue line represents a LOESS smoothing summarising the whole dataset (Data obtained by François Bertaux)

3.2.2 Transcriptomics and proteomics analysis of *cdc2-M17as* time courses

Once *cdc2-M17as* was selected as the experimental system, RNA-seq and proteomics were conducted in three biological replicates. The cells were grown in presence of 1NM-PP1 for 12 hours, taking samples for posterior analysis every hour. Those samples were treated to produce RNA and protein for further RNA-seq and mass-spectrometry analysis, respectively.

RNA-seq data was normalised according to the method described in chapter 2. Then, fold changes can be calculated using time point zero as the standard. One consideration for interpreting this type of data is that it represents relative expression levels, in other words, its proportion of the total transcriptome (figure 3.4). The majority of transcripts show this scaling with cell size, with similar results as Zhurinsky et al. 2010 [173]. A constant proportion implies a constant concentration as the cell size increases. However, some transcript do not seem to follow this trend. These transcripts will be explored in the next section.

In a similar fashion, fold changes for proteins are calculated normalising to time point zero (figure 3.5). This dataset is also relative, representing proportions instead of absolute quantities. The majority of proteins maintain their concentration as the cell grows in size, an effect that has not been described previously in literature. Proteins whose concentration changes as cell size increases are analysed in section 3.2.6.

Analysing transcripts and proteins that are differentially expressed has two goals. First, characterising the response of the cell to growth and a limitation in DPR. And second, understanding how these molecules are able to escape the global coordination between cell size and gene expression could shed light on the mechanism behind it. These molecules are also potential cell size sensors, as their concentration is not proportional to cell size.

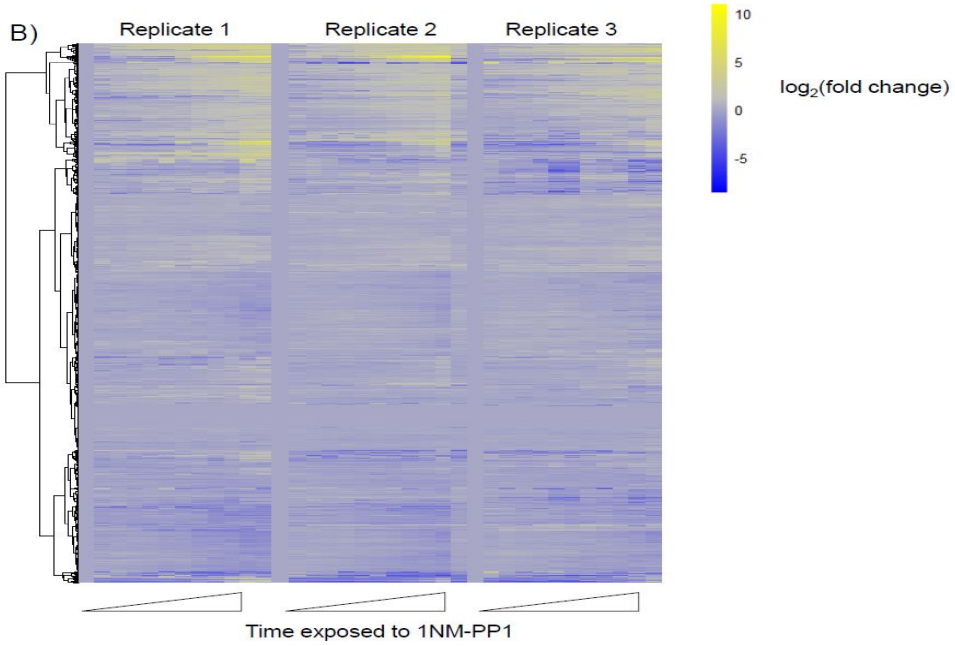


Figure 3.4: *cdc2-M17as* cells were grown with 1NM-PP1 for 11 hours, taking samples at each hour for RNA extraction. All samples were pooled and sequenced, obtaining a full transcriptomics dataset for all samples. This heatmap summarises that dataset. Raw counts were normalised using DESeq2 normalised counts, and then divided by the counts at time point zero to obtain fold changes that are represented in a log2 scale. It is important to note that at time point zero cells are proliferating normally, so it includes cells in all cell cycle stages in different proportions. The data was normalised to this time point. The proportion of the transcripts in gray does not change during the time course. The three biological replicates are represented separately

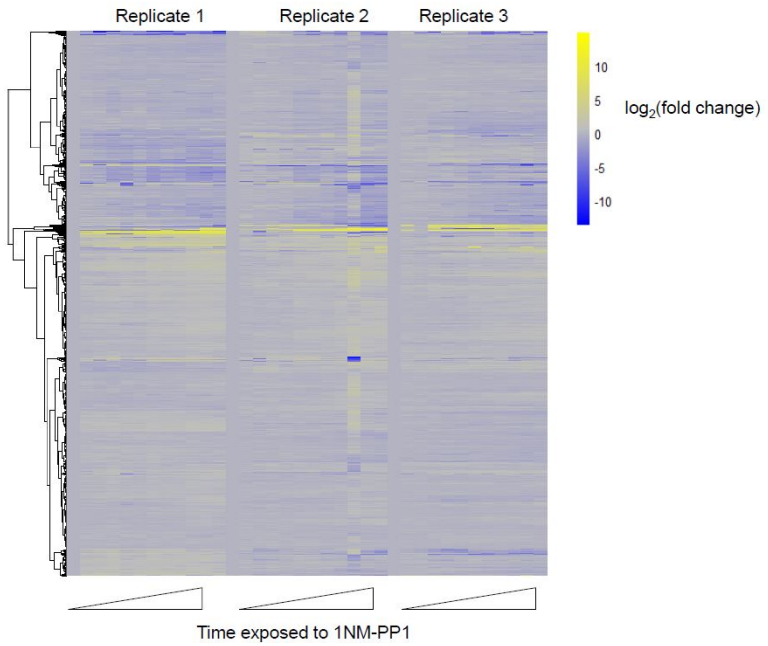


Figure 3.5: Proteins were extracted from *cdc2-M17as* cells grown in the presence of 1NM-PP1 for 11 hours. Samples taken every hour were processed using mass-spectrometry to quantifify 37% of total proteins. Heatmap representation of the proteomics dataset. Raw mass-spectrometry data is normalised using the quantification software Progenesis. Fold changes are calculated dividing by the quantification at time point zero. The figure shows the behaviour of the quantified proteins quantified in the three replicates separately.

3.2.3 Differentially expressed transcripts in *cdc2-M17as* time courses

The transcriptomics dataset were used to identify differentially expressed (DE) genes. This subset of genes are the ones that do not follow the global trend of scaling with cell size. There is a variety of statistical methods developed to identify these genes in RNA-seq data, such as the most commonly used DESeq2 and edgeR. However, these methods rely on statistical independence of the samples, so they cannot be applied in time series data. To ensure accuracy of the results, I used two different pipelines and consider only genes that appear in both algorithms' outputs. The first method used is MaSigPro (Microarray Significant Profiles) [117, 30], using the version specific for RNA-seq count data. First, DE genes are selected using a regression fit applied to all genes and next, variable selection is applied to identify profiles that are significantly different in between experimental groups. The second approach is called BETR (Bayesian Estimation of Temporal Regulation) [8] and consists in fitting two models to the data, one that assumes that expression is constant in time and one that permits changes

in the levels of gene expression. Then, the model that fits the data with a higher probability is chosen.

The number of differentially expressed transcripts for each method and the overlap between the two lists are represented in figure 3.6A. Only genes contained in both lists will be considered DE for the rest of the analysis (figure 3.6B). The data is grouped using a hierarchical clustering algorithm, producing two very differentiated sets of genes: the ones whose proportion goes up with time (Cluster 1) and the ones which proportion goes down during the time course (Cluster 2). Functional enrichment analysis of these gene lists was performed using AnGeLi [15], a tool developed for fission yeast to systematically perform enrichment analysis through a variety of available datasets. Cluster 1 is enriched in genes related to protein catabolism and nucleotide synthesis, whereas cluster 2 contains chromatin remodellers and ribosomal proteins. These gene subsets and their characteristics that are studied in depth in chapter 4. The decrease of ribosomal proteins observed in the transcriptomics dataset is striking, as the increase in size would need a higher number of ribosomes to support it. In addition, the fraction of ribosomal proteins have been described to correlate with cell size and growth rate in prokaryotes and eukaryotes [131, 130, 71, 73, 72, 94]. Besides, an exponential growth in size has been hypothesised to rely in the fact that ribosomes produce themselves. Taken together, the experimental evidence points to a potential limitation in ribosomes being responsible for the pattern of cell size increase. A mathematical modelling approach to study this phenomena is described in chapter 5.

Using the analogue sensitive strain, there is still an statistically significant overlap with the CESR (figure 3.7) [25]. However, as the temperature is the same one used for growing wild type cells, it is certain that the stress response is induced by the increase in cell size and not by environmental conditions .

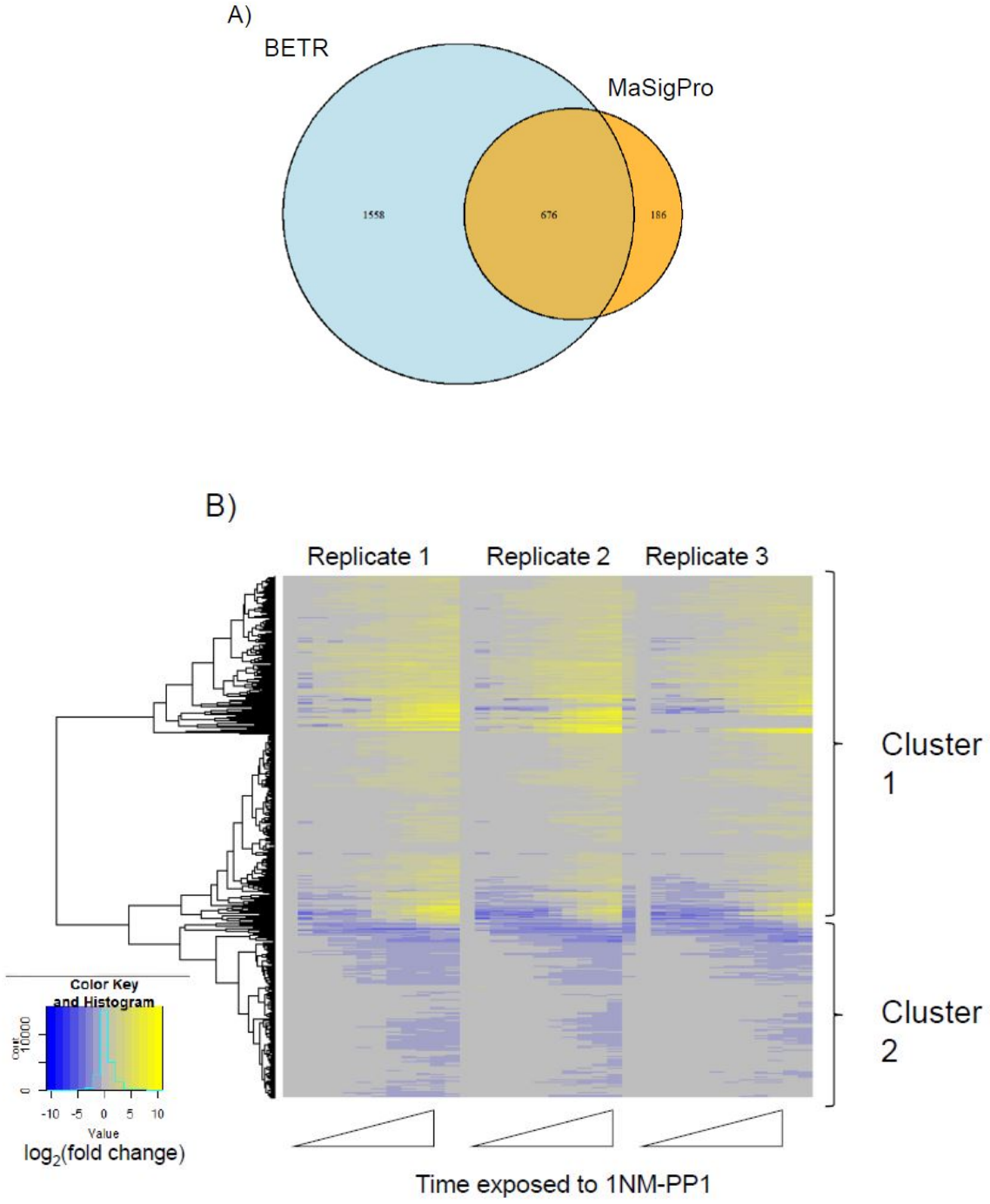


Figure 3.6: Summary of the overlap between the two methods used for calling statistically significant differentially expressed genes. A) Venn diagram showing the output of the two methods used for calling DE transcripts and their overlap. Transcripts considered regulated by the two methods are the ones represented in B) as a heatmap. Applying hierarchical clustering on the data produces two clusters of upregulated and downregulated genes.

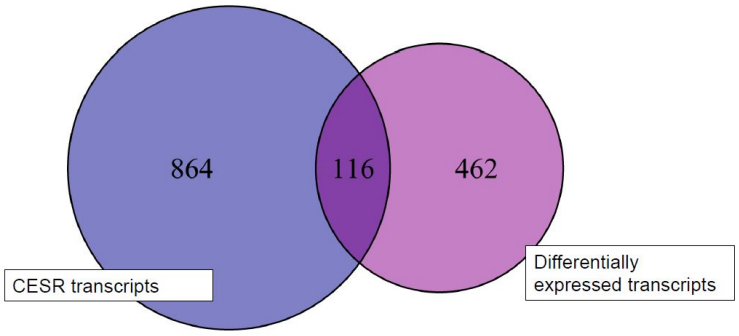


Figure 3.7: Overlap between the stress response described in [25] and differentially expressed transcripts in the *cdc2-M17as* time course (Fisher exact test p-value = 0.009). The signifcant overlap suggest an activation of the stress response when cell size increases, as the external conditions have not been described to induce such transcripts.

3.2.4 Splicing efficiency correlation with cell size

One of the factors that could potentially limit gene expression is splicing, During this process, introns are removed from newly synthesise transcripts to produce the sequence that will be translated. A shortage of spliceosome subunits, for example, would provoke a reduction of transcripts that, in turn, would decrease the number of proteins produced. Sequencing data can be used to quantify splicing rates, due to the availability of reads from spliced and unspliced forms of the transcript. Splicing efficiency is calculated as the relative amount between unspliced and spliced RNA [122]. Spliced mRNA species produce reads that span two exons, called junction reads or transreads. Splicing efficiency can then be estimated using the relative amount of transreads compared to read mapping to introns.

I did not observe a change in global splicing rates. To explore if there might be changes only affecting a subset of genes, splicing efficiencies for every gene are represented in a heatmap in figure 3.8.

In the data we can observe that splicing efficiency changes along the time course for some groups of genes. To study these groups further, the data was divided in 4 clusters using k-means and subjected to enrichment using the Angeli tool [15]. In clusters 1 and 2, splicing efficiency increases over time. These clusters are enriched in genes related to nitrogen metabolism, mem-

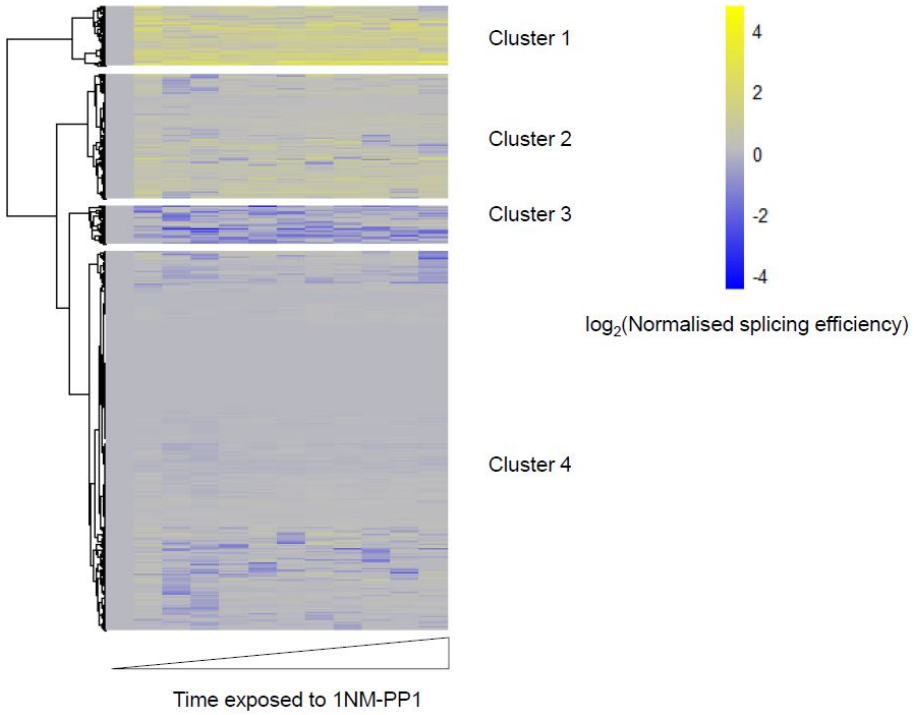


Figure 3.8: Splicing efficiencies for all genes with introns were calculated using the ration between reads that span two exons and reads that map to introns. Replicates were averaged due to the low number of these reads, leading to missing data for some genes in different samples. Splicing efficiencies during the time course, normalised to time point zero to show fold changes. Due to the scarcity of transreads and intronic reads, the three biological replicates were averaged to get a more complete dataset. The data was then clustered using k-means to produce 4 different clusters.

brane transporters and genes related to cell polarity. This could reflect the increased need for this transcripts as the cell grows larger, as transcripts of this kind are also up regulated. Cluster 3 contains genes whose splicing efficiency decreases as the cell grows. Interestingly, this cluster contains some polymerase subunits and proteins involved in the post-translational modification of histones. Cluster 4 is formed by transcripts whose splicing efficiency shows a slight decrease during the time course. It also includes genes related to metabolism, specially the nucleotide related pathways.

There is no change in global splicing rates, excluding the possibility of a shortage of splicing machinery as the cell increases in size. However, there is a change in efficiency in certain subsets of genes, suggesting a role for splicing efficiency in the regulation of these transcripts.

3.2.5 Non-coding transcription and cell size

Transcriptomics also allow the detection of a whole array of non-coding RNAs (ncRNA), whose role in cell physiology regulation is still unclear. I decided to then investigate if they could have a role during growth or DPR limitation. These unstable RNA species increase in amount when certain RNA degradation pathways are impaired, for instance in mutant strains. Using these criteria they can be classified in three types: 1) XUTs (Xrn1-dependent Unstable Transcripts), 2) CUTs (Cryptic Unstable Transcript) and 3) DUT (Dicer-dependent Unstable Transcripts) according to the pathway involved in their degradation [9]. Non-coding transcripts have also been shown to change when chromatin remodellers are compromised [35]. A change in their proportion could, therefore, point to a alteration in chromatin remodellers or in histone post-translational modifications. Larger sizes are accompanied by an increase in transcription rates [173]. One possible mechanism behind this could be a global change in the status of the chromatin to allow more accessibility of the transcribing complex. For instance, an overall increase in activating marks in the epigenome, would make transcription more pervasive and, consequently, increase the amount of these transcripts.

Approximately 7% of long non-coding RNAs are differentially regulated during the time course. In figure 3.9, only non-coding RNAs levels are represented, together with their classification according to the aforementioned criteria. Different ncRNA types do not cluster together according to expression levels, which rules out the deregulation of one specific RNA degradation pathway. It has been shown in [84], that in stress conditions the expression of ncRNAs is associated with reductions in protein levels from overlapping coding genes. To check if there is a possible influence of non-coding transcription on transcript levels in the context of growth and DPR limitation, I computed the correlation between ncRNAs and their adjacent transcripts levels (figure 3.10). Then, I compared those (n = 51) with the rest of gene pairs that can be found in the genome. There is a subset of transcripts whose expression levels seem to be anticorrelated to the ones of their ncRNA pairs. Some genes key for transcription show this behaviour, such as some components of the SAGA complex (*spt20* and *ngg1*) as well as genes involved in chromatin modification (*set6* and *rtt106*). The data suggests a possible regulatory

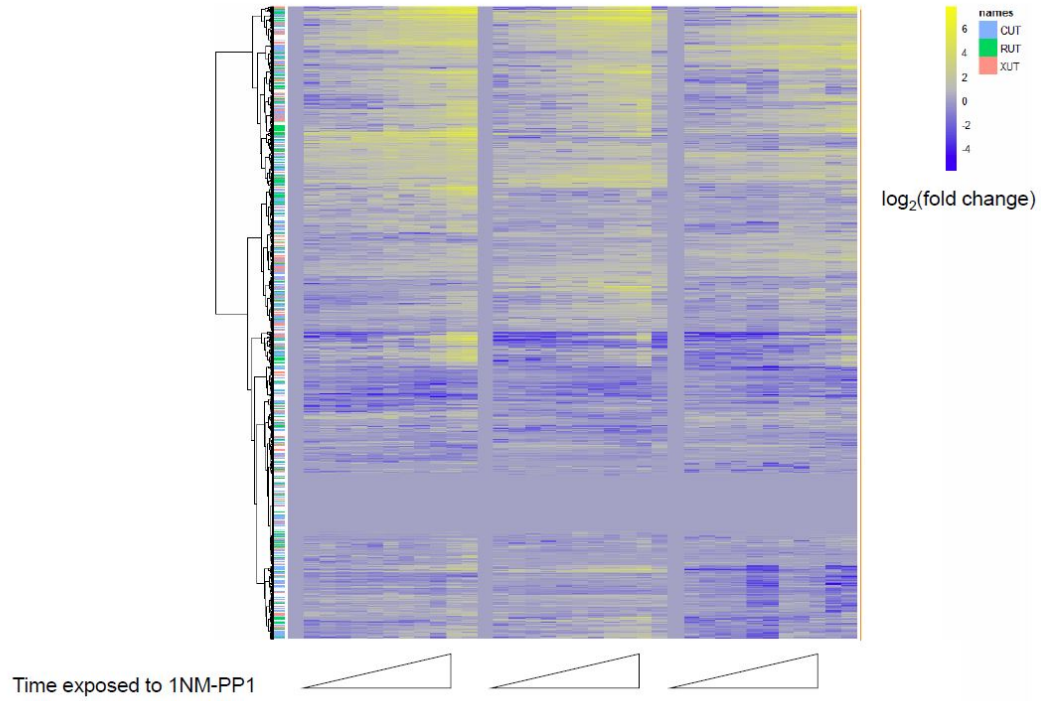


Figure 3.9: RNA-seq data for long non-coding transcripts during the time course using the *cdc2-M17* strain. DESeq2 normalised counts were divided by counts at time point zero to obtain fold changes. Coloured labels on the left of the heatmap indicate different types of non-coding transcripts, according to Atkison [9]. Long non-codings RNAs were classified according to their behaviour in different RNA-degradation mutant strains. CUTs are induced when the exosome is defective, whereas XUTs increase in concentration when *xrn1* is deleted. Last, the knock out of Dicer induces the transcription of DUTs.

role of ncRNAs in the expression of these genes as a function of cell size.

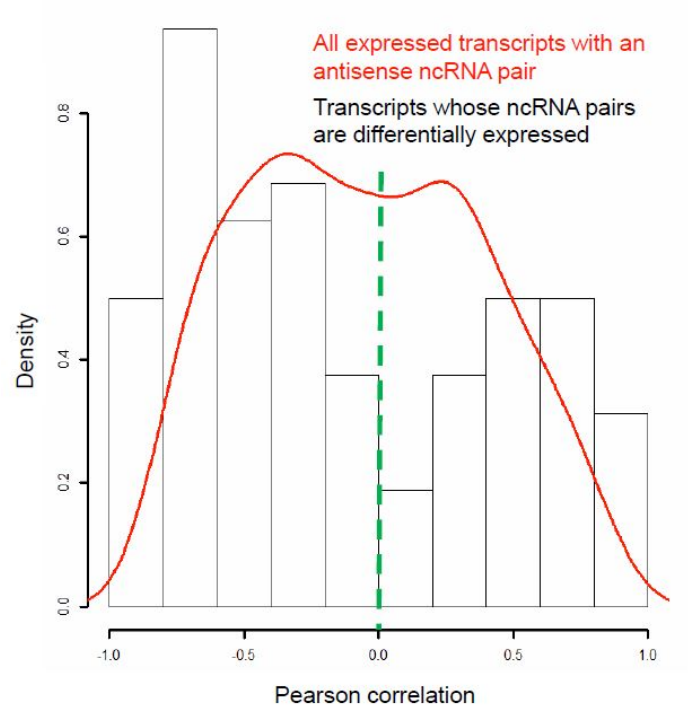


Figure 3.10: The histogram represents the distribution of the Pearson correlation between the expression levels of differentially expressed non-coding transcripts and their nearest coding neighbour. If non-coding RNAs have any influence in the transcriptional regulation of their coding neighbours, the expression of the pair would shown a correlation different that then average. To be able to compare with a background distribution, the red solid line is the Pearson correlation between every non-coding and coding pairs. The green dotted line is the threshold chose to consider those genes to be affected by the expression of their non-coding pair.

3.2.6 Differentially expressed proteins in *cdc2-M17as* time courses

Transcript levels give us information about the regulation of genes, however those changes are not necessarily translated into protein levels. To understand how the cell physiology is affected by the increased cell size, we need to also analyse proteins.The same algorithms (BETR and MaSigPro) were used to identify differentially expressed proteins during the time course. The intersection between the two algorithms produced 99 proteins that would be considered differentially expressed. This represents approximately 5% of the proteins detected in the study, compared to roughly 8% of the transcripts that are regulated. Most of the proteins do not change in fraction when cell size increases. Two very well defined clusters can be observed, one with the proteins whose proportion increases and the ones that decrease with time (figure 3.11B). In cluster 1, a increase in metabolic genes can be observed. These genes are also highly expressed in proliferating cells, with more copies per cell than average. This could be a product of the bias in the proteomics dataset, as lowly expressed proteins are not detected. However, comparing proteins in this cluster with only the proteins included in the dataset provides a similar result (figure 4.4). Cluster 2 includes ribosomal proteins and other factors involved in ribosome biogenesis, which also shows a reduction in the protein levels reflecting the reduction in transcripts. Histones and other nuclear proteins can be found in this cluster, suggesting that their concentration is not proportional to cell size. Histones, for instance, are only synthesised during S phase. As cells are arrested in G2, there is no production of new histones, producing a decreasing concentration as the rest of the proteome increases in number. This finding replicates what was also observed at the transcriptome level and will be studied more deeply in chapter 4.

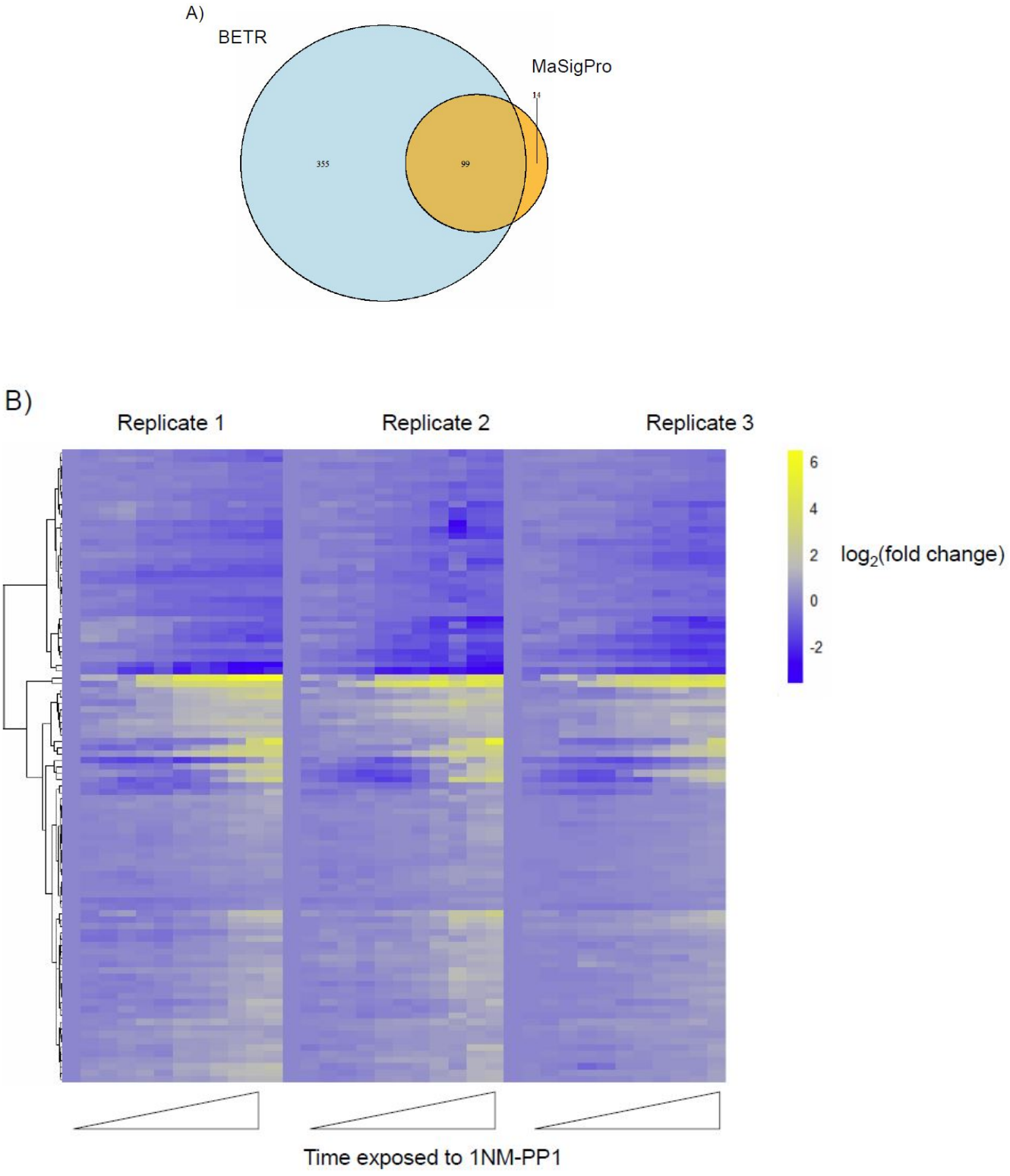


Figure 3.11: Proteins were extracted at each hour during the 11 hour exposition to 1NM-PP1 and quantified using mass-spectrometry in three biological replicates at each time point. The data was analysed using BETR and MaSigPro to produce a list of differentially regulated proteins during the time course. The overlap between the two algorithms is represented in A) as a Venn diagram. Proteins called DE by the two methods are represented in B) The expression levels in each replicate are normalised to time point zero

3.2.7 Proteome fractions

An anti-correlation between amounts of metabolic proteins and ribosomes has been described previously in prokaryotes [130, 131] in response to changes in the environment that lead to different division rates. This has been proposed to be due to the limited amount of cellular resources that the cell needs to allocate in the most efficient way depending on the environment. To visualise if there is a change in the proportion of these pathways as the cell grows, I calculated the proportion of each fraction at each time point. It is important to note that the fractions are over the proteins quantified and not the total proteome (figure 3.12). We observe a decrease in the proportion of ribosomal proteins is accompanied by an increment in the fraction of metabolic enzymes. Metabolic proteins were divided according to their annotation into carbohydrate metabolism and amino acid metabolism. The increase in the metabolic fraction is mostly driven by an increase in the proportion of enzymes related to carbohydrate metabolism. The fraction related to the metabolism of amino acids stays constant as the cell increases in size. A limitation in the cell resources could be responsible for this change in resource allocation, as the environment has an excess of nutrients.

The change in ribosomal fractions leads us to the question of whether there is also a change in the levels of ribosomal RNA. The RNA samples are subjected to a depletion of ribosomal RNA before library preparation, in order to be able to pool all the samples in the same lane. Due to this, information about these species from the transcriptomics dataset is unreliable. However, a change in these transcripts could be reflected in the proteins that produce them. rRNAs are transcribed by polymerases I and III, in opposition to coding transcripts that are generated by polymerase II. All of them are multimeric complexes, and in the case of fission yeast, share some subunits. I calculated the proportion of each complex over the total number of molecules that belong to any polymerase. Subunits that are shared by several complexes are categorised independently, as it is impossible to ascertain to which complex they belong to (figure 3.13). There is no observable consistent change in the fractions during the time course. The absence of a change in fractions for polymerase subunits is not enough evidence to confirm a change in the rRNA amount as the cell increases in size. Further evidence, using qPCR for instance, is

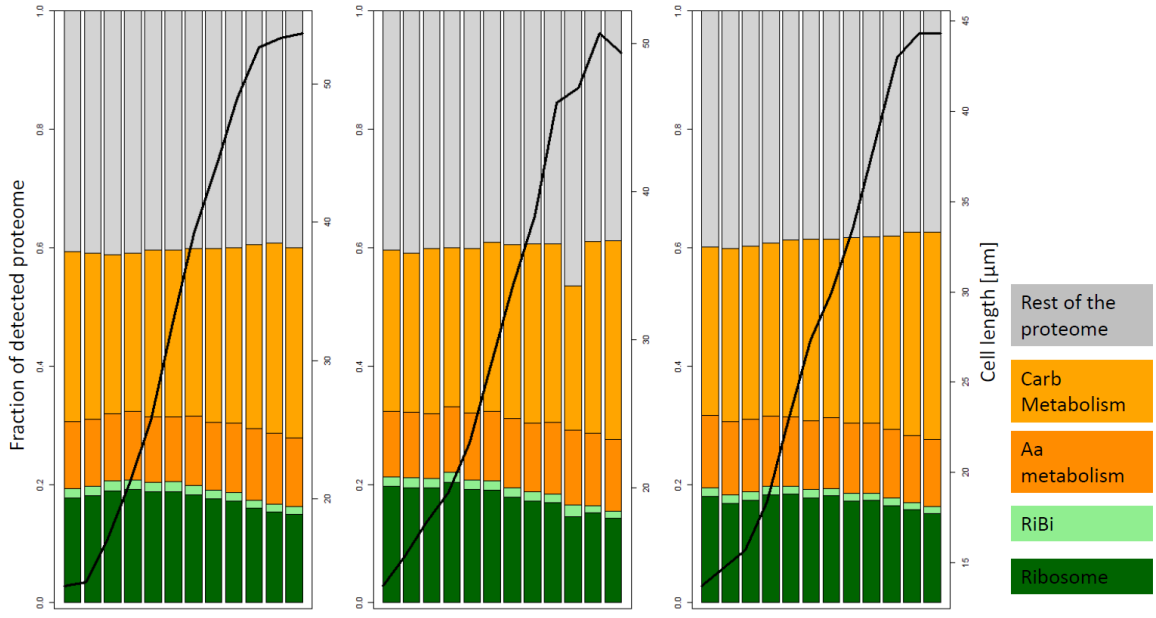


Figure 3.12: Proportion of ribosomal proteins and metabolic enzymes in the detected proteome. The proportion of each category was calculated dividing the sum of all members of each category by the sum of all proteins detected using mass-spectrometry. Proteins were categorised using their GO annotation and split into ribosomal and metabolic categories. The ribosome category was divided between ribosomal proteins and the ribosome biogenesis regulon. In a similar fashion, metabolic enzymes were separated in two categories: carbon and amino acid metabolism. The grey section is any protein not included in the aforementioned categories. The black line represents the cell length at different time points. Biological replicates were separated in three different panels.

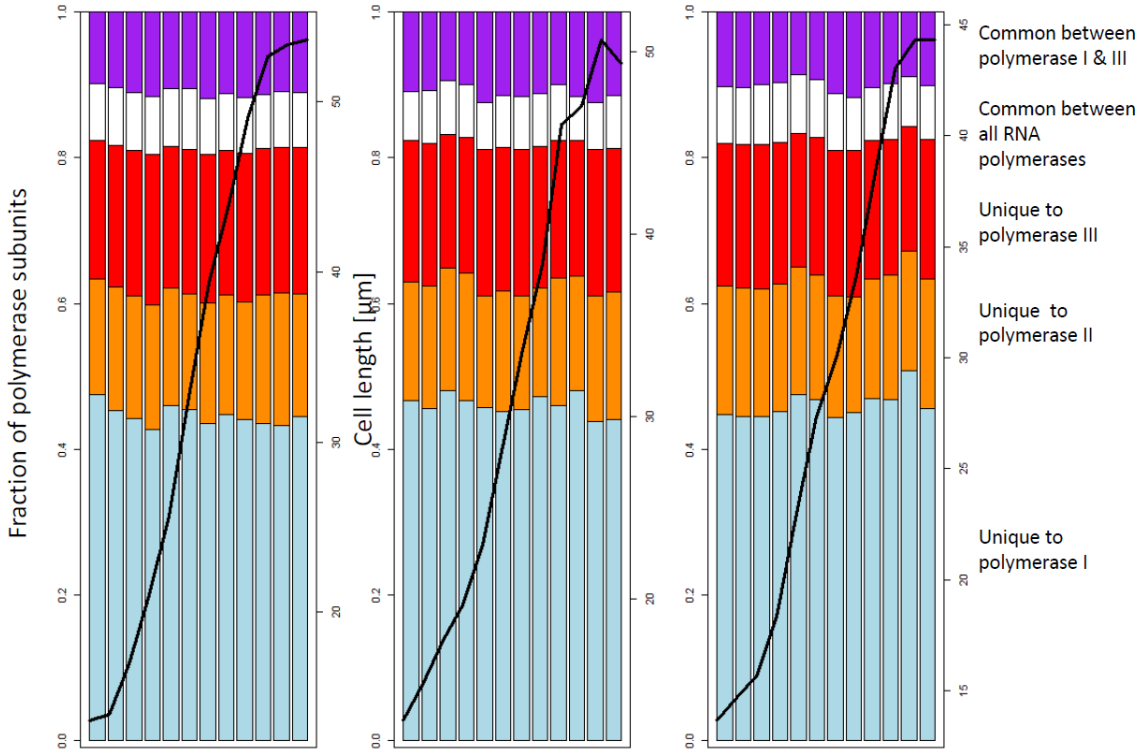


Figure 3.13: Proportion of subunits that belong to each of the polymerases. Annotation of each protein was obtained from pombase.org, and fractions calculated out of sum of the quantities of all the polymerase subunits. Shared subunits between different polymeres are considered independently, as it is impossible with this dataset to ascertain which complex they belong to. The black lines represent the increase in cell size in each of the time points. Each panel corresponds to a biological replicate

needed to support this hypothesis.

3.2.8 Comparison between proteome and transcriptome

The availability of proteomics and transcriptomics data for the same genes in the same time points, allows the analysis of the relationship between a transcript and its product. Correlation between transcripts and their corresponding gene product measures the difference between the behaviour of the two molecules. The distribution of this statistic between mRNA and protein is represented in figure 3.14A. The mean of the Pearson correlations is higher than zero, implying that the majority of transcripts and their corresponding proteins behave similarly during the time course. There is, however, a sub-population of transcripts and proteins whose levels of expression are opposite. We consider genes whose RNA-protein correlation is below -0.2 to

be potentially regulated at the post-transcriptional level. Protein and transcript expression data for candidates below the threshold is represented in figure 3.14A. Genes can be separated into two groups depending on their behaviour: Cluster 1, where the transcripts increase but the protein levels decrease, and cluster 2, where the protein increases as the transcript goes down in concentration (figure 3.14B). Cluster 1 does not show any functional enrichment, and it contains some subunits of the proteosome as well as chromatin remodellers such as Ino80. On the other hand, cluster 2 is enriched in ribosome biogenesis factors and RNA processing proteins. These includes proteins that contribute to splicing, as well tRNA-methyltransferases that contribute to the maturation of tRNAs.

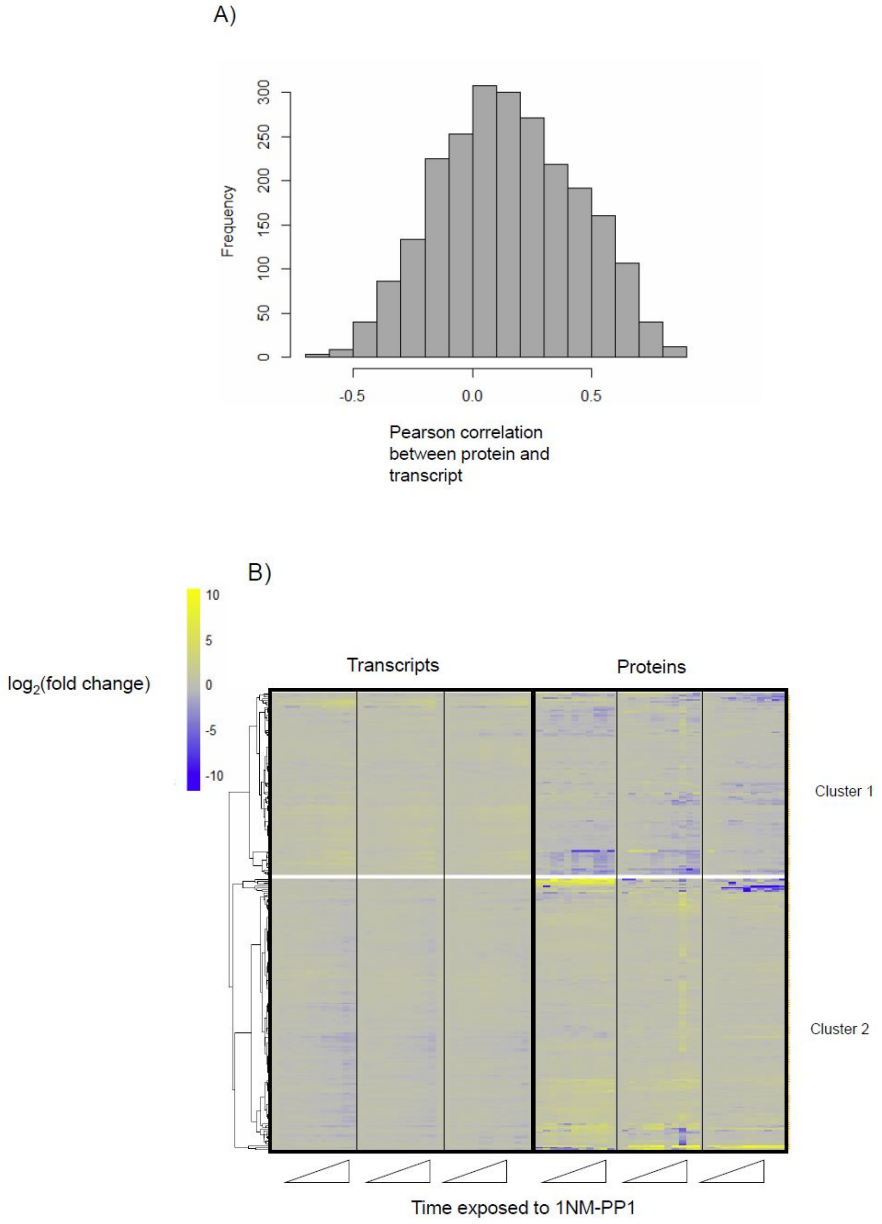


Figure 3.14: A) Distribution of the Pearson correlation between RNA-seq and the corresponding mass-spectrometry data for each protein. Correlations were calculated comparing the time course data for one transcript against the data for the same protein. The distribution is skewed to the right, showing that the majority of transcripts and proteins are correlated. B) Heatmap representation of transcript-protein pairs whose Pearson correlation is below -0.2, and therefore candidates for possible post-transcriptional regulation. The data is divided in two clusters: one where the transcript increases while the protein decreases, and the other one with the opposite behaviour.

3.2.9 Differentially expressed transcripts between growth and DPR-limited phases of the time course

The evolution of cell size in the *cdc2-M17as* strain when exposed to 1NM-PP1 can be divided in two different phases. First, a growth phase, in which the cells grow in size in an exponential fashion for approximately 6 hours. Afterwards, when the genome content becomes limiting, cells stop elongating. At the phenotypic level the two phases are clearly distinct, however if that translates into specific changes in gene expression is unclear from the analysis presented in previous chapters. This section aims to investigate the differences between the transcriptomics and proteomics profiles of the two phases. For instance, the cells could be launching a low-DPR specific response or stop growing without changing their regulation in the latest time points. Studying growth phases separately also will provide information about the metabolic needs in each particular process. Investigating genes and proteins that vary between the two phases will also provide information on what the limiting factor for gene expression could be when DPR is limiting.

Both phases were compared using MaSigPro. In this case, this method would find genes whose kinetics are significantly different between the growth and the limiting DPR phase. For example, a gene that increases in concentration during the first phase but stays constant at the end of the time course would be included in the differentially expressed gene list. In previous analyses, I also used another algorithm (BETR) and considered only those genes that were called as differentially expressed by both algorithms. Unfortunately, the package for this algorithm was not available at the time this analysis was carried out. Approximately 12% of genes show a significantly different behaviour between the early and late phases of the time course. To identify which genes have a common trend in gene expression, the data was clustered using k-means to produce 6 different clusters (figure 3.15 right). The clusters were characterised using the enrichment in KEGG pathways (figure 3.15 left). According to this analysis, 5 out of the 6 clusters are enriched in KEGG pathways. Each cluster is enriched in a different group of genes, making them biologically distinct between each other. As seen when taking the whole time course into account, there is an over representation of genes involved in metabolism in Cluster

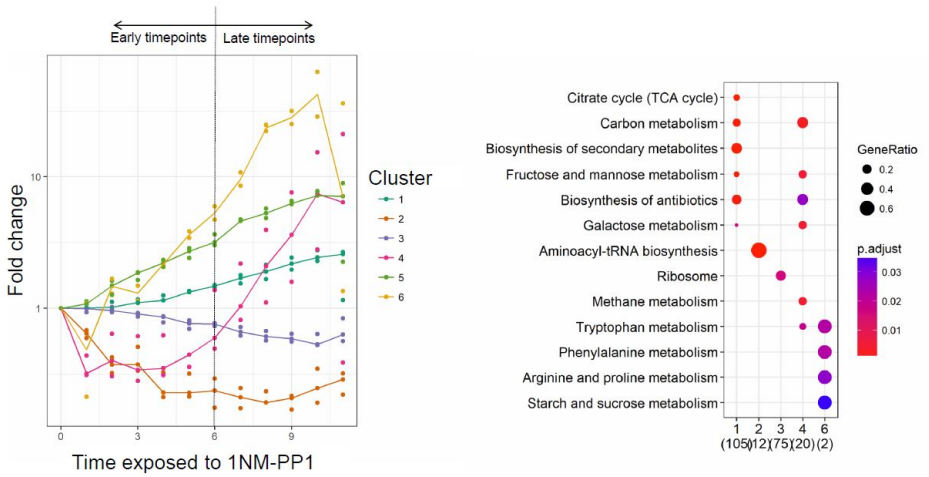


Figure 3.15: Clustering of the differentially expressed transcripts when comparing early and late phases of the time course. The clustering was performed using a k-means algorithm with a number 6 clusters. The panel on the right shows the average behaviour of each cluster in the RNA-seq dataset. Lines represent the median of the data, and the dots the median for each biological replicate. A comparison between the KEGG enrichment between the clusters is represented in the right panel. Enrichment values were calculated using the hypergeometric test, using the KEGG database available categories. Enrichment was performed using the *clusterProfiler* [169] function enrichKEGG.

1. In Cluster 2, only the enrichment in tRNAs can be observed. tRNAs contain modified nucleotides that affect the activity of the reverse transcriptase that transcribes the first strand in the library preparation method, making the quantification of individual tRNAs unreliable. However, the caveat in quantification would affect all tRNAs in the same way. It could still indicate that there is general reduction in the proportion of tRNAs in the trascriptome. tRNAs deliver amino acids to elongating peptides in the ribosome. A reduction in the concentration of tRNAs could severely impair transcription, stopping the process when proteins have that amino acid in their sequence. However, this finding must be considered still with reservations (Figure 3.15).

A representation of several metabolic pathways can be observed, together with ribosomes and tRNAs. Cluster 1 is enriched in genes related to carbon and sugar metabolism, whereas clusters 2 and 3 are translation related, with a majority of tRNA or ribosomal genes respectively. Cluster 4 contains genes that decrease during growth but are strongly induced when cells are at their lowest DPR. It is comprised of genes that are upregulated in stress conditions [25]. Non-coding

RNAs are overrepresented in Cluster 5, specially transcripts that are induced when *dbr1* is deleted. This gene encodes an enzyme that debranches the lariat, a product of splicing that contains the introns and is degraded after the debranching. The deletion of this gene in wild type cells has big effects on the global splicing efficiency, and induces a subset of non-coding RNAs. How this transcripts are induced is not clearly understood, it is speculated that it might be due to the presence of introns in their structure. Coding-genes in this cluster are mostly fission yeast specific proteins whose function has not been described. Finally, metabolism of several amino acids is enriched in cluster 6 (figure 3.15).

3.2.10 Differentially expressed proteins between growth and DPR-limited phases of the time course

A similar analysis can be applied to the proteomics data to find proteins whose behaviour is different between growth and DPR-limitation. To investigate what accumulation kinetics can be found at the protein level, k-means clustering was applied. Dividing the data into three different clusters gives the best result, producing groups with very distinct kinetics (figure 3.16). Cluster 1 contains proteins involved in metabolism as well as some ribosomal proteins that decrease in concentration as the cell grows larger. Some enzymes related to carbon metabolism decrease during the growth phase but increase when DPR becomes limiting (Cluster 2). One of these enzymes is a glucose-6-phosphate 1-dehydrogenase, the rate-limiting enzyme in the pentose phosphate pathway. This pathway is an alternative to glycolysis that produces ribose without spending ATP. Ribose is then used as a precursor for the synthesis of both ribonucleotides and deoxyribonucleotides. An increase in the concentration in this enzyme could reflect a shortage of nucleotides precursors when the cell reaches its limiting size. Finally, proteins related to glycolysis and secondary metabolism are also up regulated during growth to plateau when the DPR is low. An up regulation of glycolysis would produce more metabolic precursors that are used in metabolic pathways other than mitochondrial respiration. At the transcript level, all three clusters show a trend similar to the protein one (figure 3.16). Protein clusters were also subjected to an analysis of their promoters and chromatin environment, to address whether

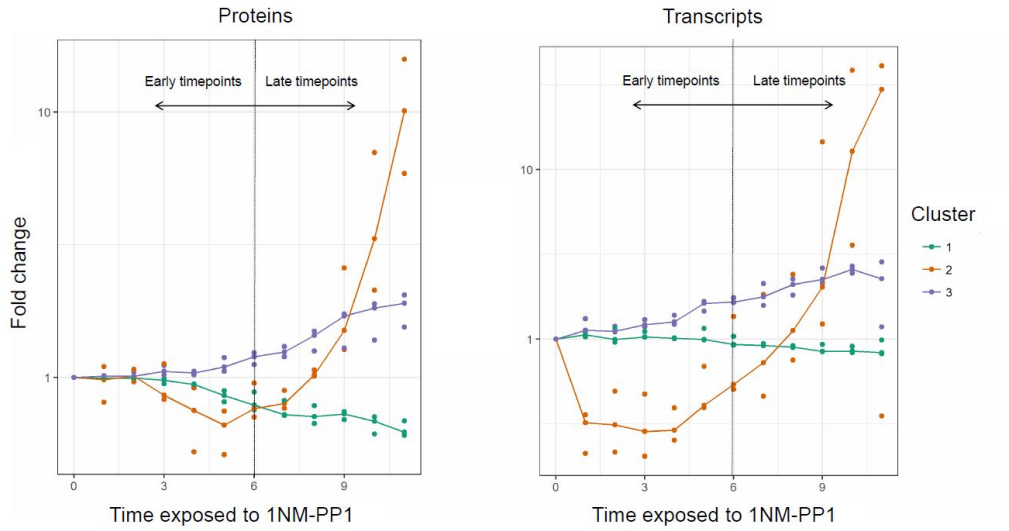


Figure 3.16: Differentially expressed proteins between the late and the early time points were divided in three clusters using k-means. The average behaviour of the proteins in each cluster is represented in the left panel, as the solid line. The dots are the average behaviour of each cluster in the three different biological replicates. To assess if there is evidence for post-transcriptional regulation, the corresponding transcripts were also plotted in the right panel.

they are under the same transcriptional regulation. No sequence was enriched in the promoter of the members of the different clusters. Regarding chromatin modifications, H3K4me2 seems to correlate with the gene expression levels. On the other hand, the repressive mark H3K9me2 is not overrepresented in any of the protein clusters. Genes enriched in H3K9me2 are lowly expressed and are not represented in the proteomics dataset.

3.3 Discussion

This chapter describes the acquisition and characterisation of the transcriptomics and proteomics dataset produced in this study. These datasets represent the dynamics of gene expression when the cell is challenged with an increase in cell size that leads to a lower limit in DPR.

The single cell elongation data shows a clear exponential trend before growth stops abruptly. However, single cell growth has been proposed to happen in bilinear pattern, with a change in growth rate that coincides with DNA synthesis in S phase [12]. Other authors argue that growth

is exponential, the larger the cell the faster it grows [31]. Both models fit the experimental data extremely well, with only small statistical differences between the two. These differences may reflect small changes in growth rate due to cell cycle progression. If growth were bilinear in G2-arrested cells, they would grow linearly after the change of rate during DNA synthesis. However, the data collected in this study shows a very strong exponential behaviour. This does not necessarily imply that cells during a normal cell cycle share the same growth pattern, as they are artificially arrested in one phase of the cell cycle. Bilinearity could be a product of having different cell cycle phases with different growth rates, and thus only observed in cycling cells. Yet, it provides evidence towards an exponential growth pattern during G2.

One of the most prominent responses to growth in both datasets is the decrease in expression of ribosomal proteins and ribosome biogenesis factors. These genes have been described to be major drivers of cell growth, correlating positively with growth rate in E. coli [130, 73]. Moreover, they have been described in *S. pombe* as involved in the growth programme described in [25]. This seems contrary to the observation of an increase in cell size in the genetic system used in this study, pointing to a possible role in limiting growth when DPR is low. On the other hand, transcripts related to metabolism are accumulating during the time course. It could be a reaction to the down-scaling of ribosomes, as these two genetic programmes have been shown to be anticorrelated in prokaryotes [130, 131, 73]. Different metabolic needs for a larger cell when the DPR starts to be limiting could also explain this behaviour. In this case, the cell as its grows would need to adjust metabolism to a lower DPR. As pointed out in [66], availability of nucleotides can be limiting for transcription so this might be a response to a higher need for precursors when transcription rates increase with growth. Fission yeast goes through a metabolic cycle that is synchronised with the cell cycle phase, so that different reactions are separated by time [134, 95]. During G2, several waves of gene expression have been correlated with oxygen consumption. Ribosome biosynthesis and amino acid biosynthesis genes peak at the beginning of G2, followed by an increase in the stress response. The up regulation of metabolic genes in the dataset could be explained by the metabolic cycle, however, it is not accompanied by a peak in ribosomal biogenesis.

Mining the transcriptomics data, information about different RNA degradation pathways and

splicing rates can be obtained. Non-coding RNAs are degraded by specific RNA pathways, thus, the accumulation or depletion of a certain type can be a proxy for the activity of certain degradation complex. The data does not provide any evidence that specific classes of non-coding transcripts are behaving in a similar way, points to an more specific regulation rather than a change in their correspondent RNA degradation pathway. Sequencing data can be used to calculate splicing efficiencies, using the ratio between reads that span two neighbouring exons and intronic reads. Calculating these values for all the genes with introns does not show any consistent change during time, either for specific genes or for the total splicing rate. The data does not provide enough evidence to consider splicing as a limiting factor when the DPR is low.

Taken together, the results of this chapter suggest a important role of ribosome function during cell growth and at a lower DPR. The single cell size data also provides some evidence towards an exponential model of cell growth, although more work is needed to confirm this trend.

In early time points, the cells are increasing in length exponentially until growth stops abruptly after approximately 6 hours. The molecules highlighted in this analysis provide a picture of the physiological differences between growth and limitation in genome content. Most differentially regulated transcripts and proteins show changes during growth to then stabilise after the cells have stopped increasing in size. Very few molecules are responding exclusively to the DPR-limitation. Growth seems to be the major driver in the regulation of the majority of genes. Metabolism seems to be playing a very important role in both phases of the time course. Both phases show a very striking difference regarding the kinetics of different metabolic routes. During growth, an increase in transcripts related to carbon and amino acid metabolism can be observed. However, that change is not extended to the late phase of the time course, where the proportion of this molecules stabilises. On the other hand, ribosomes decrease consistently during growth, stabilising when the cell's size stops increasing. However, there is a subset of transcripts and proteins that decrease during growth to increase afterwards. These species suggest the activation of pathways that produce metabolic precursors of amino acids and nucleotides. This could indicate a shortage of the building blocks necessary for transcription and translation. Potentially the increase in transcription rates that accompanies cell size would increase the consumption of nucleotides, inducing biosynthetic pathways. A

similar scenario could occur regarding translation, when the fast consumption of amino acid could trigger the synthesis of new ones. Some proteasome subunits are also induced. The degradation of existing proteins provides amino acids that can be used in the synthesis of new ones.. To asses if the availability of nucleotides and amino acids has an influence in the, growth of the *cdc2-M17as* strain, media conditions have been described that limit the synthesis of these molecules. Limiting the amount of nitrogen would produce a decrease in the number of amino acids that the cell is able to produce, whereas reducing phosphate would affect the synthesis of nucleotides [65]. If these metabolites are limiting in the experimental system studied, they would affect the elongation rate of the cell or the maximal attainable size.

Chapter 4

Characterisation of genes that do not scale with cell size

4.1 Introduction

I have shown in chapter 3 that the majority of transcripts and proteins scale with cell size. However, there are some molecules that escape this global trend. Unlike the rest of the genome, the concentration of these molecules is not proportional to the size of the cell. namely their concentration does remain constant as cell size increases. The mechanism through which these molecules manage to escape the global coordination of size and gene expression is not understood. Determining the differentiating characteristics of these transcripts and proteins would shed light in what could be the scaling mechanism for the rest of the genome.

In addition, non-scaling genes could play a role in regulating cellular processes. Their dynamic concentration makes them ideal candidates for concentration-dependent triggers. Several example of these genes and how their non-scaling properties influence cell physiology are described in literature; for instance, the protein Whi5, a G1/S transition inhibitor in budding yeast. Synthesis of Whi5 is cell-size independent, with cell of all sizes showing a similar number of molecules. The concentration, however, changes depending on the size of the cell. Smaller cells have a higher concentration of Whi5 at the start. The concentration of Whi5 decreases with

cell size until its below a threshold, when it activates the entry into G1/S phase. To reach the trigger concentration, smaller cells need to grow more in size, as they start with a more concentrated protein due to their size. Its non-scaling properties are crucial in regulating the size of the cell at G1/S [128]. Another example is the fission yeast gene *cdc25*a phosphatase that acts as a mitotic activator. Contrary to Whi5, the concentration of *cdc25* increases during the cell cycle until it reaches a threshold, inducing cell division [67]. Non-scaling molecules are not only involved in the control of cell size, they have other physiological roles. One example is the role of histone number in development in *Xenopus laevis*. After the egg is fecundated, a series of quick divisions occur without any production of biomass, including transcripts. As the cells divide without producing new histones, the decrease in histone number triggers the the midblastula transition [4]. Despite having been described as having key roles in the cell physiology, how any of these molecules scape scaling is still poorly understood.

In this chapter, I will explore the genes that do not scale with cell size in the time course dataset and the possible causes for this difference. The strain used (*cdc2-M17as*) is particularly apt for identifying proteins and transcripts of this kind, as the cells grow to an extreme size. This would make the differences in proportions more acute and easier to detect, improving the identification of these molecules. Besides, the majority of cells are arrested in G2, removing the effects of cell cycle specific transcription.

4.2 Results

4.2.1 Identification and characterisation of non-scaling genes

Identifying non-scaling transcripts and proteins

In chapter 3, I identified transcripts and proteins whose proportion changes significantly during the time course. Methods used and a more detailed explanation on how the analysis was done can be found in chapter 3. Briefly, two different algorithms were applied to the datasets to obtain molecules whose concentration changes significantly. The transcript or protein lists

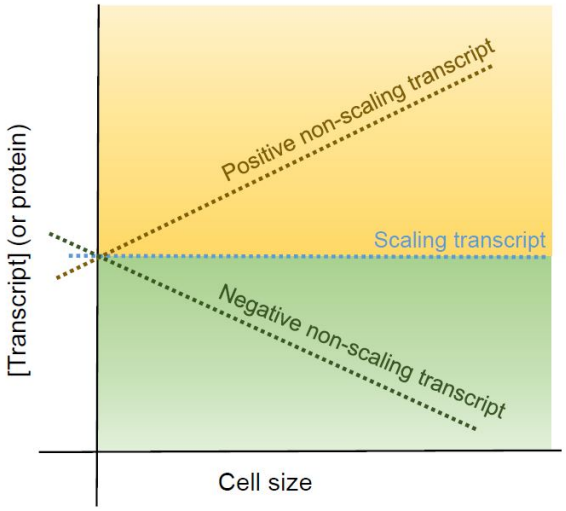


Figure 4.1: Schematic explaining the difference between positive and negative non-scaling transcripts. Most transcripts (or proteins) will fall under the scaling category, in which their numbers increase linearly with cell size maintaining concentrations. Positive non-scaling molecules show an increase in concentration with cell size. In turn, the concentration of negative non-scaling molecules decreases with growth.

obtained are then clustered to separate genes in sets depending if their proportion increases or decreases during the time course. From this point onwards the two groups will be designated as positive non-scaling molecules, meaning their concentration increases faster than cell-size does or negative non-scaling molecules, whose concentration decreases with cell size (figure 4.1).

There are several hypothesis as to why these genes are not subjected to the global coordination of cell size and gene expression. For example, cell cycle regulated genes are not expected to scale with cell size. These genes are only expressed in certain phases of the cell cycle. In the other phases, the transcripts are subjected only to degradation and thus get diluted. A similar logic could be applied to genes expressed in burst, such as stress regulated transcripts. For constitutively expressed transcripts, however, we expect their transcription rates to be size-independent (figure 4.2).

In the next section I will characterise positive and negative-scaling genes in terms of their transcriptional regulation, chromatin environment and other genomic features. The goal of this analysis is to identify what characteristics separate these genes from the rest of the genome, and how that could influence their scaling properties.

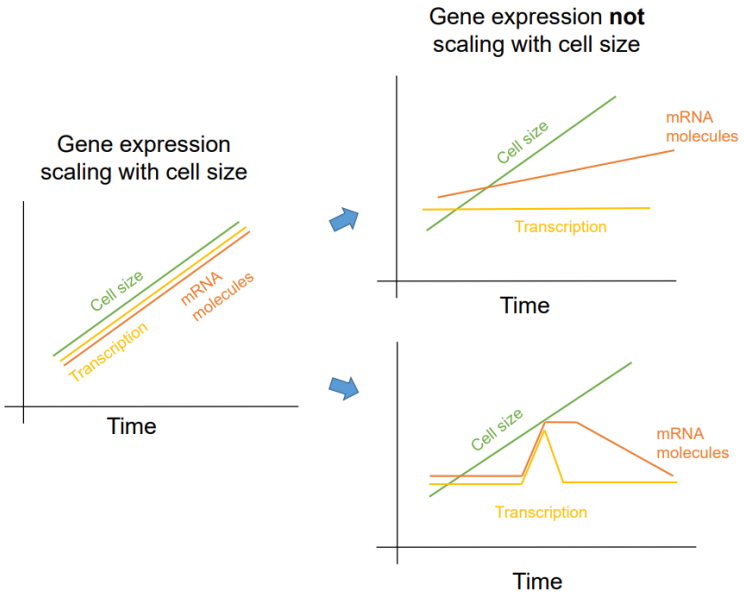


Figure 4.2: Diagrammatic illustration of the difference between cell cycle regulated transcripts and mRNAs whose transcription rates are not proportional to cell size. Transcripts that are not constitutively expressed peak at certain stages of the cell cycle, to be then only subjected to degradation. As they are only being degraded for the majority of time, their concentration is not proportional to the size of the cell. For constitutively expressed genes, there must be a mechanism that makes their transcription rates not scale with cell size.

4.2.2 Positive non-scaling genes

Gene enrichment and expression levels

453 transcripts and 68 proteins were predicted to increase in concentration as cell size increases. Full gene lists can be found in appendix tables A.1 and A.3. In order to find what pathways are overrepresented in those lists, GO terms and KEGG pathway enrichment were performed. Genes related to metabolism are overrepresented in both protein and transcripts, increasing faster than cell size does (table A.2 and A.4). An enrichment in pathways related to nucleotide metabolism can be observed in both datasets, suggesting a potential increase in the demand of nucleotides at larger cell sizes. Transcription rates have been shown to be proportional to cell size [173]. As transcription rates increase with cell size, so would the demand for nucleotides to sustain the rates of transcript production.

There is a possibility that the stress response could be activated as the cell is facing an non-physiological increase in size. To check if there is an overlap between the positive non-scaling

genes and the genes upregulated in stress, I used the data from Chen et al. 2003 [25]. In this article, the genome-wide response to several different stimuli was characterised by microarrays. The stress response is centralised in the CESR (Core Environmental Stress Response) that is activated in any kind of stress that the cell is subjected to. Proteins and transcripts categorised as non-scaling do show a significant overlap (Fisher test p-value ¡ 0.001) with stress-induced transcripts, pointing to a partial induction of stress upregulated genes when the DPR is low (figure 4.3).

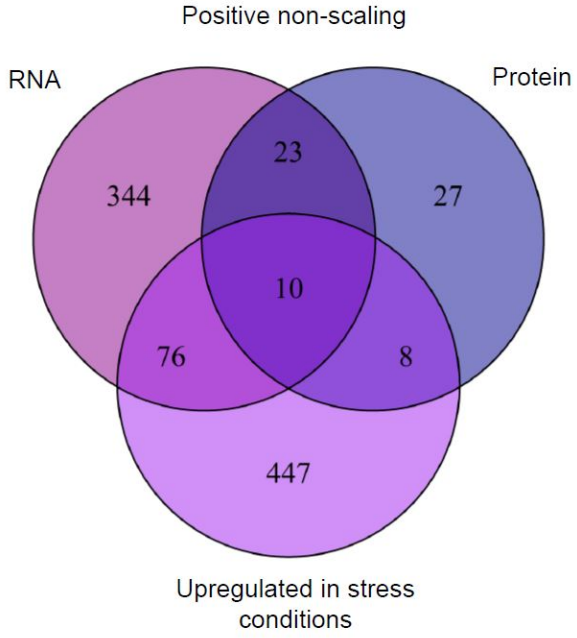


Figure 4.3: Non-scaling transcripts or proteins were predicted using the overlap between differentially expressed transcripts or proteins predicted by two algorithms, BETR and MaSigPro. Positive non-scaling molecules are the ones whose concentration increases with cell size. Here I draw a comparison between the positive non-scaling transcripts and proteins and the stress response described in [25]. There is a siginificant overlap between positive non-scaling transcripts and the stress response, suggesting there is an induction of the stress related transcripts when the DPR decreases.

I also used published protein and transcript data for wild type cells [96] to asses if there are any differences in the number of copies per cell between scaling and non-scaling genes. The data used was obtained in a culture of proliferating cells. When comparing the expression levels of these transcripts in a wild type situation with the rest of the transcriptome, transcripts that do not scale have significantly lower mRNA copies per cell (Wilcoxon test p-value = 4.729e-10). However, the protein copies of these genes are significantly higher that the average (Wilcoxon

test p-value = 0.006), highlighting the importance of post-transcriptional regulation. Positive non-scaling proteins are in a higher copy number per cell in proliferating cells (Wilcoxon test, p-value = 0.0003) (figure 4.4). Data for positive non-scaling proteins is compared against the distribution of detected proteins, not the whole proteome. This is to take into account the detection bias, as only a subset of proteins can be detected through proteomics.

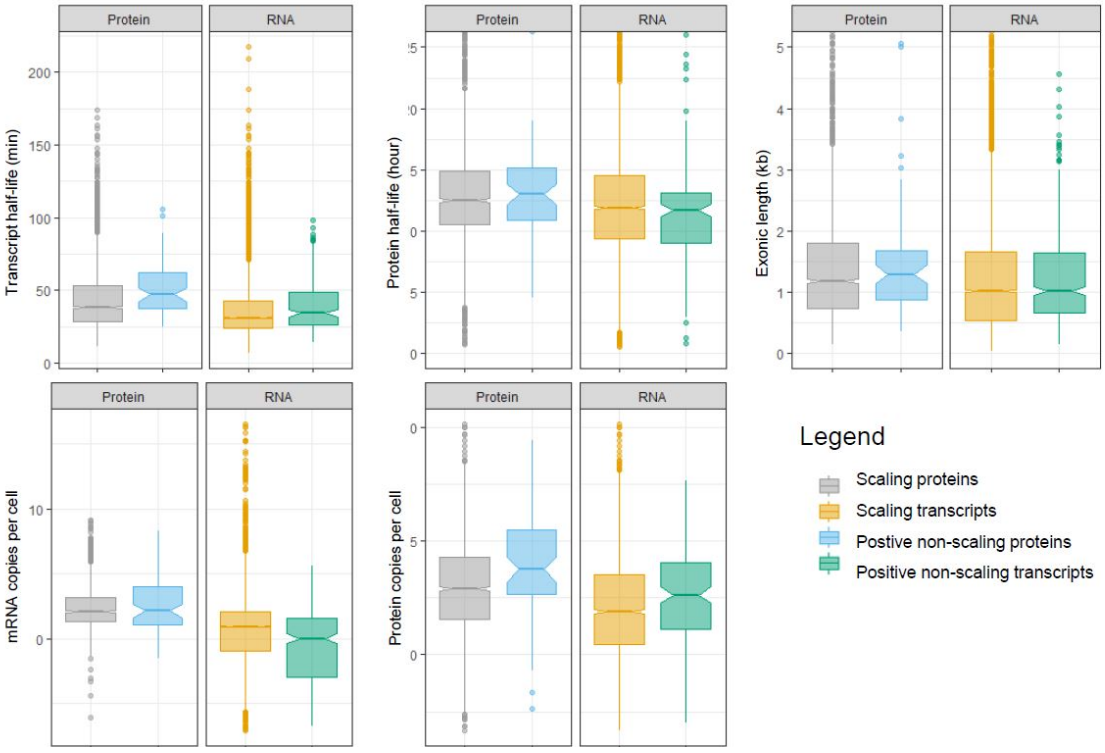


Figure 4.4: Comparison of different transcript features between positive non-scaling transcripts and proteins and the rest of the transcriptome or proteome. Non-scaling transcripts were compared against the rest of the transcriptome. Proteins, on the other hand, were compared only with proteins that were detected by mass-spectrometry in this experiment. Data for the half-life of transcripts and proteins were obtained from [52] and [27]. Copies of transcripts and proteins per cell were acquired form [96]. The data is represented using boxplot, with the notches around the median representing the 95% confidence interval around the median.

Promoters and chromatin landscape

The action of an specific transcription factor could be responsible for the ability of these molecules to escape the pervasive global regulation. To explore this possibility, both analyses of known transcription binding sites and the discovery of new motifs were carried out. Known transcription factors motifs were obtained from pombase.org. Specifically, three dif-

ferent algorithms were used: AME, CentriMo and MEME [10]. AME and Centrimo are both intended to find enrichment of known binding sites, however Centrimo will only call motifs that are enriched in a particular region of the sequences whereas AME does not take the position of the motif in the sequence into account to calculate enrichment values. MEME finds new motifs that appear more often that by chance in the sequences provided. Transcription binding sites can be at different distances from the promoter, a characteristic that depends on the specific transcription factor. In order to not miss any enrichment, sequences at 50, 100, 150, 200 and 500 bp around the TSS were analysed. TSS positions were obtained from the CAGE data published in[89]. The targeted approaches did not produce any significant result. There is small number of transcription factors that have a published motif in fission yeast, it is likely that none of them are responsible for the regulation of positive non-scaling genes. MEME did flag some motifs as being significantly enriched in the dataset, however those motifs seem to be very specific to a small subset of genes in the sample (figure 4.5), thus unlikely to be behind the regulation of the whole group of genes.

Genes predicted to be positively non-scaling are very varied in terms of their annotated sub-cellular compartment. Different compartments would have different scaling needs, for example the cytoplasmic membrane scales with the surface area of the cell instead of its volume. Nuclear proteins could also scale only with DNA amount, or with the size of the nucleus that has been shown to be proportional to cell size [116]. If this were the case, different compartments would be under a different regulation. To address this question, the gene lists were divided according to the compartment were the gene products are targeted (pombase.org) and rerun through the promoter analysis pipeline. The new analysis did not produce any significant result for any of the compartments.

The chromatin modification landscape also influences transcription and thus, the amount of transcripts and proteins [75]. ChIP-seq (Chromatin Inmunoprecipitation sequencing) allows for the mapping of histone modifications along the genome. Data for H3K9me2 and H3K4me3 were collected in [35, 68] for fission yeast. H3K9me2 has been strongly associated with repressed loci, whereas H3K4me3 correlates with active genes. Differences in the enrichment pattern of these two modifications could point to a role of chromatin structure in regulating the gene expression

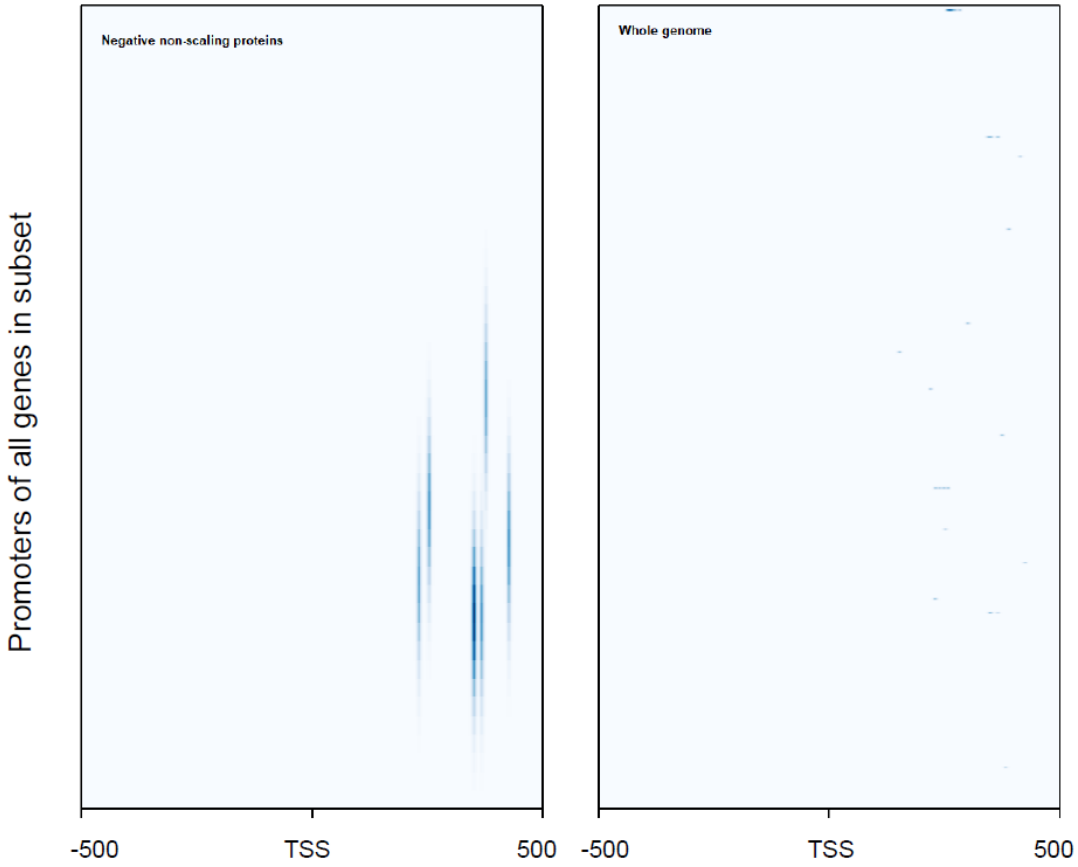


Figure 4.5: 500 bp around the transcription start site were analysed using AME to find statistically enriched motifs in the promoter sequence of positive non-scaling transcripts. These motifs could correspond to transcription binding sites regulating the expression of that subset of genes. The heatmap on the left shows the presence of an enriched motif in the positive non-scaling transcript subset. Every row is a sequence, with dark blue representing sequence stretches where that motif is enriched. The motif can only be found in a handful of transcripts in the subset. The panel on the left contains only the genes used for the analysis, whereas the one on the right represents the whole genome. Despite being flagged by the algorithm, this motif is only present in a handful of genes and it is likely not responsible for the non-scaling behaviour of that subset of genes.

response to an increased cell size. A comparison between the average modification in the DE transcripts and proteins and the whole genome is presented in figure 4.6. Genes that accumulate during the time course have a tendency to be highly expressed in wild type cells, also explaining the higher levels of H3K4me4. Regarding H3K9me2, genes that do not scale at the transcript levels seem to have a higher average presence of this modification. Using k-means clustering in the H3K9me2 levels, two very differentiated clusters can be found in the genome: a handful of genes show very high levels of the modification, while it is not very abundant in the rest of the genome. Genes with high H3K9me2 are close to centromeres and telomeres. All these genes are also present in the non-scaling genes at the RNA level (figure 4.7). However, these genes are not part of the described stress response in yeast. The coordinated deregulation of all the genes high in H3K9me2 could be due to an overall change in the chromatin modification pattern of the genome. The concentration of some proteins related to heterochromatin formation, for example *set1*, decreases with cell size. A shortage of these factors could be responsible for the up regulation of the transcripts high in H3K9me2. This could be a way for the cell to achieve higher transcription rates in response to a larger cell size.

Nucleosome positioning is another important determinant of gene expression patterns, especially the NFR (Nucleosome Free Region) that can be found upstream the TSS. Nucleosomes in that region are actively evicted to allow the polymerase access to the TSS. It has also been shown that expression levels correlate with the strength of the depletion [78] Nucleosome positioning can either be probed using H3 ChIP-seq or MNase-seq. Using available H3 ChIP-seq datasets [35] , non-scaling genes were compared to the whole genome looking for differences (figure 4.6). Genes that do not scale at the protein level show a deeper NFR , correlating with their higher average expression levels. On the other hand, at the transcript level the NFR is weaker, again reflecting lower average expression levels in the wild type cells.

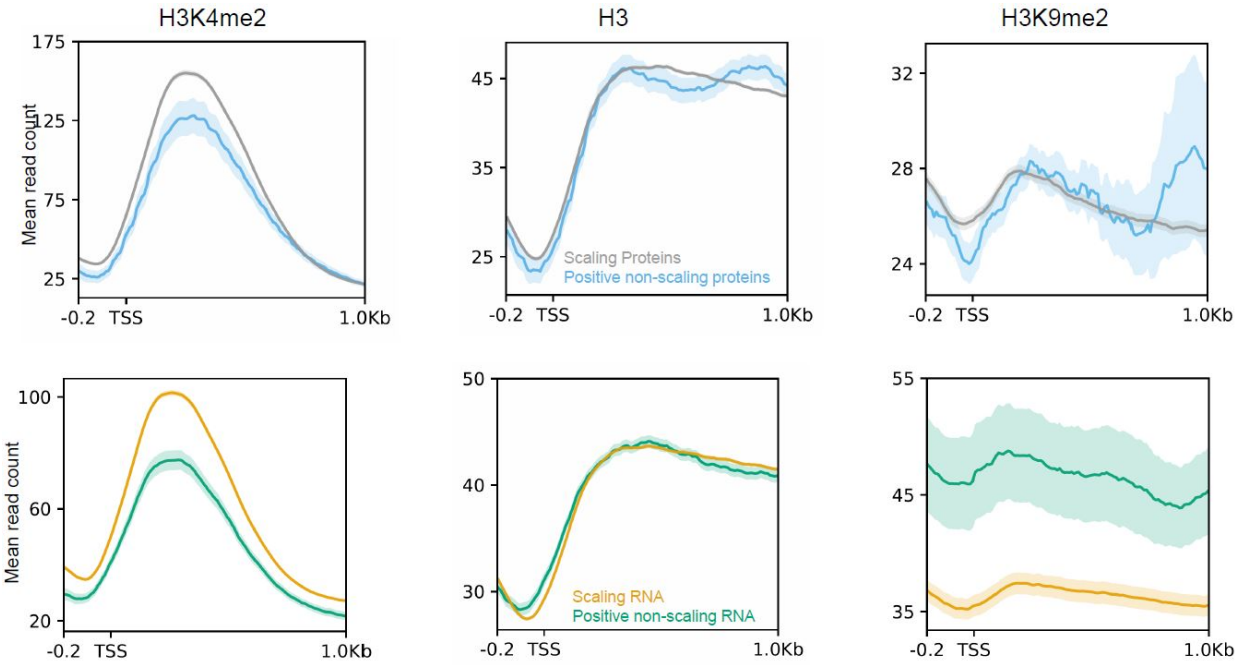


Figure 4.6: ChIP-seq data for different chromatin modifications was obtained from [35, 68] ans used to compare the chromatin environment of non-scaling proteins and transcripts with their scaling counterparts. Transcripts are compared with the rest of the genome, whereas proteins are only compared against proteins detected in this proteomics dataset. ChIP-seq reads across all gene subsets were averaged to produce the profiles shown in the figure. Standard deviation is represented as the shaded area around the solid line.

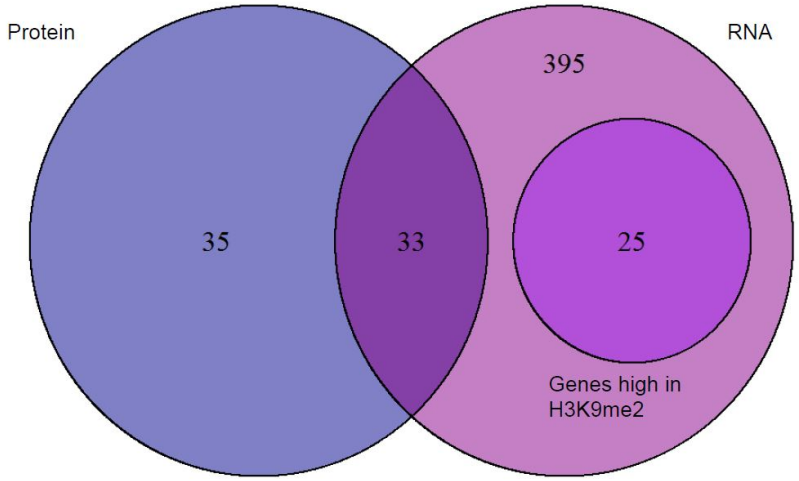


Figure 4.7: Genes with high levels of H3K9me2 were extracted using the dataset published in [35]. All those genes can be found in the list of positive non-scaling transcripts, suggesting a possible role of this modification in the scaling of transcription.

Other genomic features

Other genomic features were explored as well, such as gene length or GC content. Although some significant differences were found, the size of the effect is too small to account for the differences in the expression pattern of these transcripts or proteins.

The accumulation of these molecules could be due to their low degradation rates, making their concentration increase with cell size. Half lives for mRNAs and protein have been published in three different studies [52, 27, 40], all based on metabolic labelling. Positive non-scaling transcripts have a significantly higher half-life than average (Wilcoxon test p-value = 0.0002).

The genome is folded forming a 3D structure during interphase. As a result of this structure, genes that are far apart in the linear sequence can be adjacent in the 3D space. It has been shown in other organism that genes under similar regulation tend to be closer in space, forming what has been named TADs (Topologically Associated Domains). TADs are domains in the genome that are under a similar transcriptional regulation and are close in space. As non-scaling genes share a common regulation, it could be possible that they are located together in regulatory domains throughout the genome. Genome-wide 3D interactions are mapped using HiC, data for *S. pombe* was produced in [109]. This study revealed a genome that mostly displays interaction between sequences in the same chromosomal arm. Interactions seem to be stronger in a 100 kb window, suggesting a higher degree of local interactions, with not many long-range contacts between genes. If long-range contacts are uncommon, genes in the same TAD would also be closer than average in the linear sequence. Distances to the nearest neighbour were computed for both protein and mRNA positive non-scaling genes. To obtain the background distribution of distances, a random subset of genes was drawn a 100 times and the distance to the nearest gene in the same list was calculated. The distances between the gene of interests were compared against the background distribution using the Kolmogorov Smirnov test to compare the means, as the distributions are not normal. The statistical test produced p-values above the significant threshold, meaning non-scaling genes are not closer in the genome more that it is expected by chance (figure 4.8).

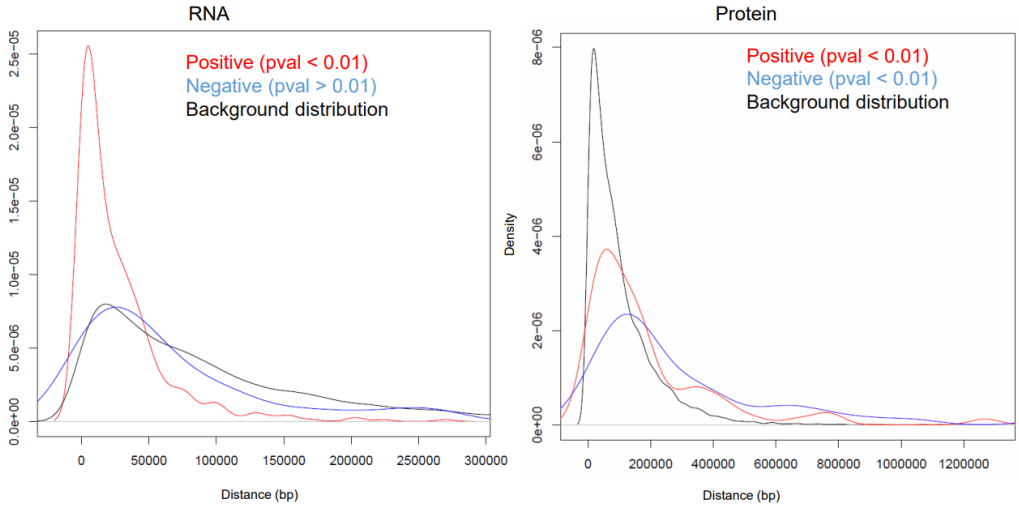


Figure 4.8: The regulation of non-scaling transcripts and proteins could depend on their location in the genome. If that were the case, genes in those subsets would be closer to each other than expected by chance. Plots in this figure show the distribution of distances between nearest neighbours for non-scaling transcripts and proteins. To obtain the background distribution to compare the gene subsets to, genes were sampled randomly from the genome and their distance to the nearest neighbour calculated. The procedure was repeated a 100 times. To compare the distributions I used the Wilcoxon test.

4.2.3 Negative non-scaling genes

Gene enrichment and gene expression levels

The DE pipeline found 128 transcripts and 44 proteins that scale negatively with cell size (gene lists are available in the appendix in tables A.5 and A.7). When applying GO enrichment analysis to the genes whose concentration goes down with cell size, a strong presence of genes related to translation and ribosomes can be observed (tables A.6 and A.8. As major drivers of growth, it is unexpected to have a decreasing proportion when the cell size increases. A decrease in ribosomal expression has been described to occur upon stress. However, the overlap between the non-scaling genes and the CERS downregulated genes is very small (figure 4.9). Thus, despite showing similar characteristics, the response to a decreasing DPR is sufficiently distinct from a stress response (Fisher test p-value >0.05).

Negative non-scaling transcripts are enriched also in transcripts involved in DNA packaging. These category includes histones, as well as enzymes involved in the post-translational modification of histones.

Nuclear proteins are also overrepresented in the negative non-scaling transcripts, specially those involved in heterochromatin spreading. Nuclear proteins mostly interact with DNA , whose quantity does not change in this experimental system because the cells are arrested in G2. It can be hypothesised that the expression of these genes might respond to the amount of DNA instead of cell size, in order to maintain a homeostatic proportion of this factors to nucleic acid content. A similar logic could also apply to membrane-bound transporters. The amount of membranes in the cell is not proportional to the increase in cell volume during growth. The cytoplasmic membrane, for example, scales with the surface area of the cell instead of its volume. Fission yeast cells are 3-4 µm wide, with constant width as they only grow from the tips. If we simulate the cell as a cylinder with two half spheres at the tips, the volume increases three times faster than the surface area (in a cell with 3 µm width). If membrane transporter respond to the amount of membranes in the cell, it could be the reason why some of them get diluted during the the time course

Copies per cell of negative non-scaling proteins and transcripts in proliferating cells were obtained from [96] as copies per cell. Proteins predicted to scale negatively with cell size show significantly higher copies per cell in the wild type (Wilcoxon test p-value = 0.0005). Non-scaling transcripts have higher copies per cell when cells are proliferating (Wilcoxon test p-value = 0.003). However, protein produced from these transcripts are in lower copy number than average in proliferating cells (Wilcoxon test p-value = 0.001) (figure 4.10).

Promoter sequences and chromatin landscape

In a similar fashion as in section 4.2.2, the promoters of genes in the subset were analysed looking for an enrichment in know transcription factor binding sites. A significant enrichment in HomolD and HomolE boxes was found. These motifs are associated with ribosomal genes and bound by the transcription factor Rrn7 [49, 125], which are overrepresented in this dataset. When removing these genes from the analysis no sequences came out as significant in either AME or Centrimo. To unravel if any uncharacterised motifs could be enriched in this subset of genes, MEME (Multiple Em for Motif Elicitation) [10] was used. This algorithm tries to find

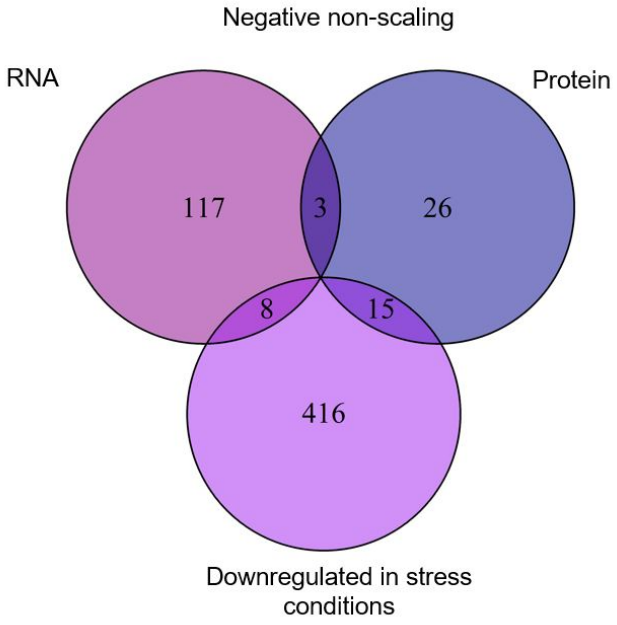


Figure 4.9: Venn diagram of the overlap of negative non-scaling transcripts and proteins with mRNAs downregulated in the stress response.

ungapped motifs of a fixed length that are occurring in several of the sequences tested. No statistically significant motif could be identified using this algorithm in the non-scaling genes both at the mRNA and protein level. Specific Transcription factors, therefore, are most likely not responsible for the behaviour of this gene subset.

The chromatin-modification landscape is also key in regulating gene expression levels [75]. As stated in the previous section, several ChIP-seq dataset for different modifications have been published in *S. pombe* [35, 68]. In figure 4.11, a comparison between the whole genome and non-scaling genes through different modifications is represented. Non-scaling genes at the protein level are enriched in the activating mark H3K4me2, as they are highly expressed. At the transcript level the same effect cannot be observed. Nucleosome positioning data, in this case H3 ChIP-seq, was also analysed [35]. As shown in figure 4.11, these molecules produce a similar pattern to the rest of the genome.

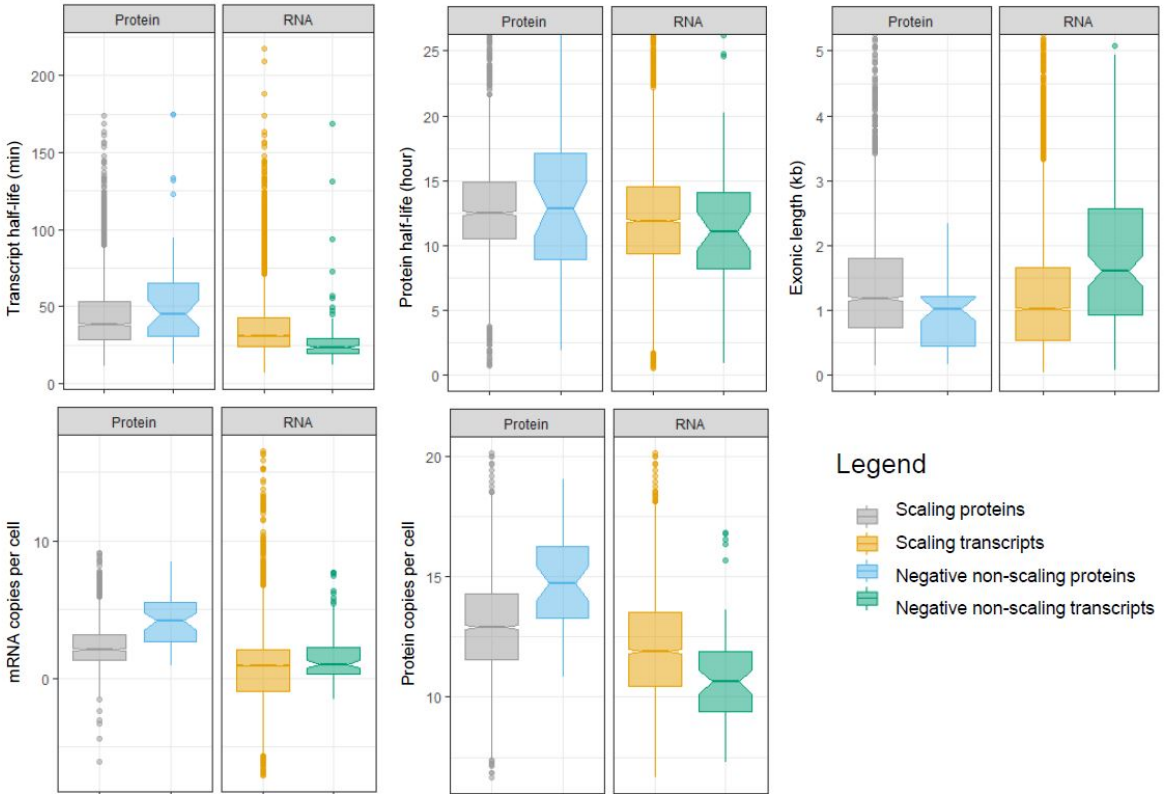


Figure 4.10: Distribution of different genomic features, such as gene lengths and molecule half-lives for non-scaling molecules compared to their scaling counterparts. Transcripts are compared to the rest of the genome. As only highly expressed proteins were detected in the proteomics dataset, proteins are only compared to the ones that have been detected instead of the whole proteome. Data for molecules half-lives comes from [52] and [27]. Marguerat and colleagues [96] quantified the number of copies per cell of transcripts and proteins in fission yeast.

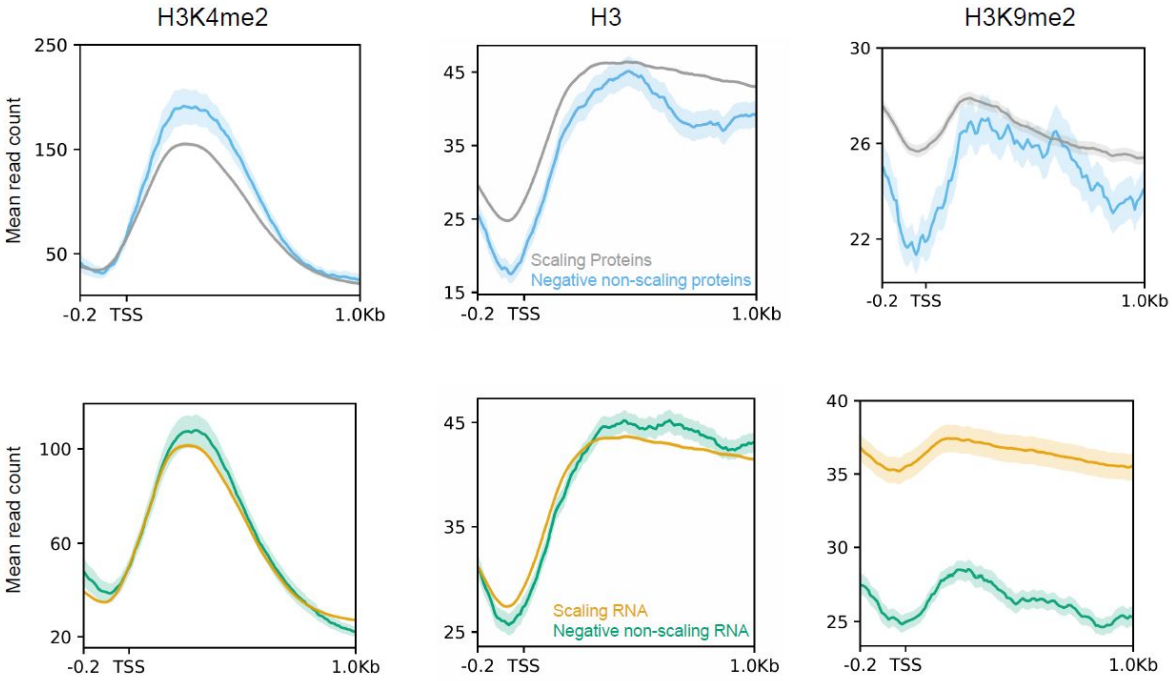


Figure 4.11: ChIP-seq profiles of H3K9me2, H3K4me2 and H3 in scaling and non-scaling gene groups. Profiles are obtained averaging the number of reads between all the genes in that group (solid line) using the data from [35, 68]. The shaded are around the solid lines represents standard deviation in each gene subset.

Other genomic features

Other sequence features could also be playing a role in producing this particular expression pattern, such as GC content and gene length. GC content for every gene as a percentage was obtained from pombase.org. Non-scaling genes, both at the protein and mRNA levels, have slightly higher GC content compared to the average in the genome (Mann-Whitney test, p-value <0.01). Albeit being statistically significant, the size of the difference between the two populations is small. Some studies point to a correlation between GC content and gene expression [76], whereas others report that there is no relationship between the two [132]. The relationship, if it exists, is hypothesised to come from differences in codon usage. The corresponding tRNAs for these codons are more abundant, making the translation process more efficient. Therefore, the higher GC content could be related to the increased expression of those genes.

Protein and mRNA half life could also play a role in changing the concentration of a molecule

at a changing cell size. Degradation rates, together with synthesis rates, set the concentration of proteins and transcripts at a given time. Thus, changes in them could affect the scaling properties of a molecule. Two different publication have produced half-life for transcripts and proteins using metabolic labelling [52, 27, 40]. I compared the half-life of transcripts for non-scaling proteins and mRNAs with the average of the genome. Only transcripts that do not scale with cell size show a significant difference (Wilcoxon test p-value = 1.8e-11). These transcripts have a shorter half-life on average than the rest of the transcriptome.

Apart from the information encoded in the sequence, the genome acquires a 3D structure during interphase that influences the regulation of transcription. When folded, regions of the genome that are far from each other in the sequence can be close together in the nucleus and be under the same regulation. In fission yeast the majority of contacts have been shown to occur in a 10 Kb window. This suggest a high number of contacts between adjacent genes, with a low proportion of long range or inter-chromosomal interactions. Using the methodology described in 4.2.2, I checked if genes in these subset are closer together than expected by chance. The analysis did not show significant results, these genes are (figure 4.8).

4.2.4 Validation of non-scaling candidates

Validation using published datasets

Genes that did not scale linearly with cell size were selected in this study using a strain that grows to a limiting size without dividing. Besides, the cells are arrested in G2, representing only one phase of the cell cycle. However, it is not clear how it will compare to strains that growth and divide and have different cell sizes compared to wild type. If these molecules show a similar behaviour in proliferating strains, it would support the evidence that classifies them as not scaling. In order to do this I will take advantage of one of the datasets generated in in Zhurinsky et al. 2010 [173]. Expression data was generated for two cell size mutants, one larger (*cdc25-22*) and one smaller(*wee1-50*) than the wild type. The strains have point mutations in the *cdc25-22* and *wee1-50* genes respectively that render the gene product temperature

sensitive. At 25°C cells show a similar size compared to wild type, as the gene product is still active. When the temperature is increased to 32°C, the activity of the protein is reduced. This provokes a change in the average cell size at division, larger for *cdc25-22* and smaller in the *wee1-50* strain. Their division rate, however, is similar to wild type cells. A negative non-scaling transcript, for example, would show a lower concentration in larger proliferating cells, as its concentration decreases with cell size. For positive non-scaling transcripts, because they accumulate with cell size increase, I would expect to see higher expression levels in larger cells. However, what to expect in *wee1-50* is less clear. The trend could be reversed, with positive non-scaling transcripts decreasing and negative ones increasing. Non-scaling genes could also show similar expression levels as the rest of the transcriptome, if the non-scaling effects were only triggered by growth, and now by a decrease in cell size.

The data obtained in Zhurinsky et al. 2010 [173] for the two strains was separated into scaling and non-scaling groups (figure 4.12). The data shows the expected behaviour for the *cdc25-22* strain, with an increase in positive non-scaling genes and a decrease in negative ones. In the *wee1-50* samples, on the other hand, differences between scaling and non-scaling genes are inconsistent between replicates. The data suggests that in this strain non-scaling and scaling genes show similar levels of expression. Further testing would be needed to assert that the non-scaling mechanism (or mechanisms) is related to growth in size. Transcripts that are predicted to not scale using the *cdc2-M17as* strain show, on average, a similar behaviour in proliferating cells. This suggest that the non-scaling behaviour is not a result of the genetic system used, but an inherent property of that group of genes.

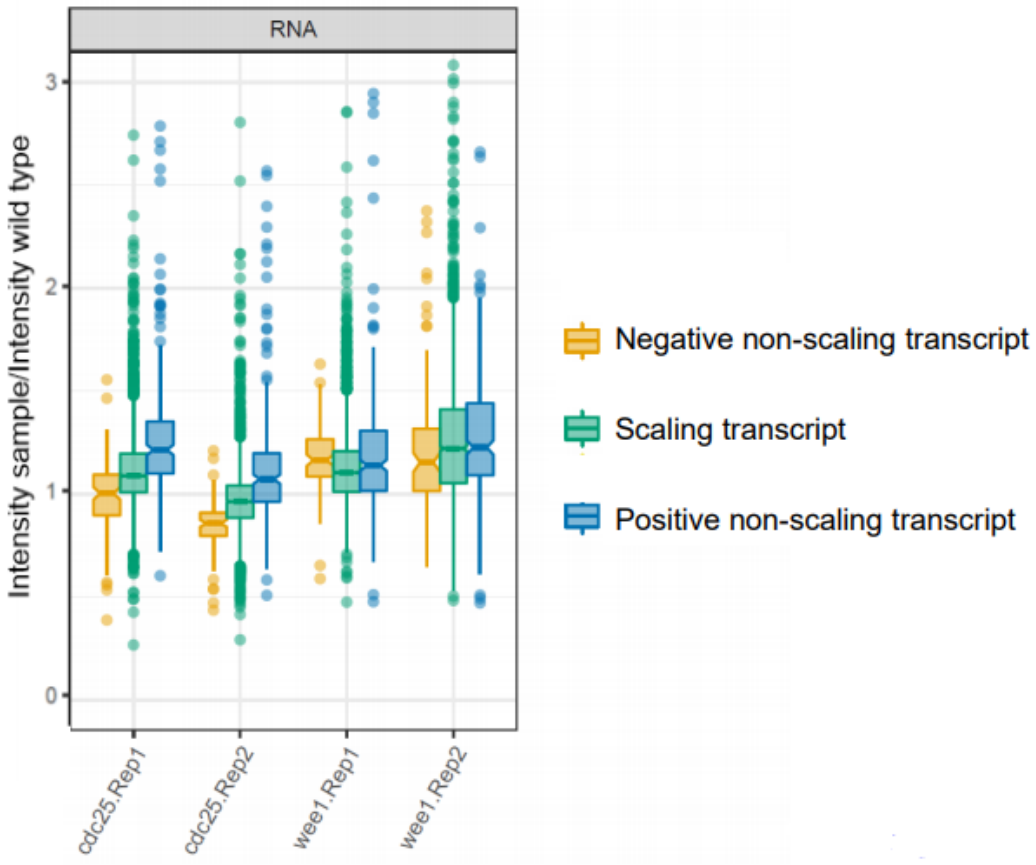


Figure 4.12: Microarray data from *cdc25* and *wee1* cells normalised against wild type published in [173]. *Cdc25-22* cells are larger than wild type, whereas *wee1-50* cells are smaller. Transcripts were separeted into different groups depending on their scaling properties. Negative non-scaling genes predicted in this study show a similar behaviour in the published dataset, presenting a decrease in expression in larger cells. This is accompanied by an increase in the *wee1-50* expression data.

mRNA quantification

After selecting candidates for genes that do not scale with cell size, these candidates need to be validated using other experimental means. Using smFISH, cell size and mRNA number can be quantified for single cells. Prior to imaging, cells need to be fixed with formaldehyde. Six candidates (*mst1,Nup107,prd1, rpb2, set1, srb10,tea1*) were chosen for their biological relevance to be probed further using this technique. As one of the hypothesis is that genes whose products act in an organelle or cell membrane are under different regulation, genes from the nuclear membrane and the cell tip were also chosen for smFISH confirmation.*mst1*, *set1* and *srb10* carry out their function inside the nucleus, whereas *prd1* is bound to the cytoplasmic membrane. *tea1* is transported to the cell tips, playing a role in polarising the cell's growth. As a negative control, I also quantified *rpb2* transcripts, a gene whose transcript number is proportional to cell size as shown by other FISH experiments carried out by other colleagues (data not shown). The transcriptomics data acquired confirms the non-scaling properties of *rpb2*, as well as suggest a decrease in concentration for the genes selected for FISH confirmation (figure 4.13). The results are presented in figure 4.14. Number of copies does not scale with cell size for any of the genes picked, as the slope of the linear regression is close to zero. The only exception is the negative control *rpb2* that shows the expected linear correlation of transcript number with cell length. However, all these genes seem to have a low copy number which combined with variability in gene expression may blur the linear relationship between cell size and mRNA number. For the two daughter cells to have the same number of transcripts as the mother, at some point the number of molecules needs to be duplicated during the cell cycle. This could skew the data towards a linear behaviour, as at the end of the cell cycle there would be higher number of molecules. It does not necessarily imply that the behaviour is linear throughout the whole range of cell sizes. For instance, *set1* shows an increase in number at larger sizes but it seems to not scale during the rest of the cell cycle. A low copy number could also be the mechanism behind non-scaling, although on average these transcripts are highly expressed.

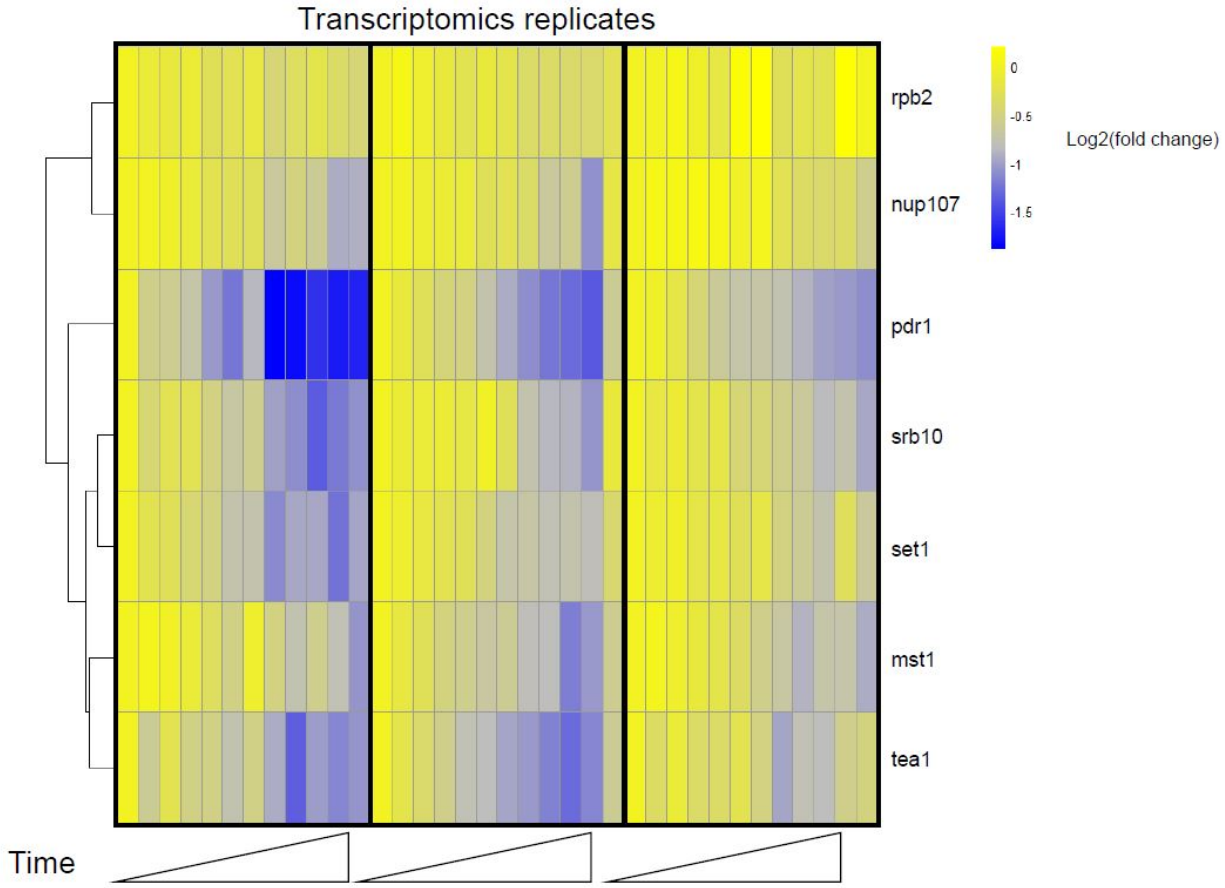


Figure 4.13: RNA-seq data for the transcripts used for FISH quantification, together with the negative control *rpb2*. Candidates were chose according to their subcellular localisation and potential role in regulating chromatin modifications. The data shows a decrease in concentration as the cell grows in size for all transcripts except *rpb2*

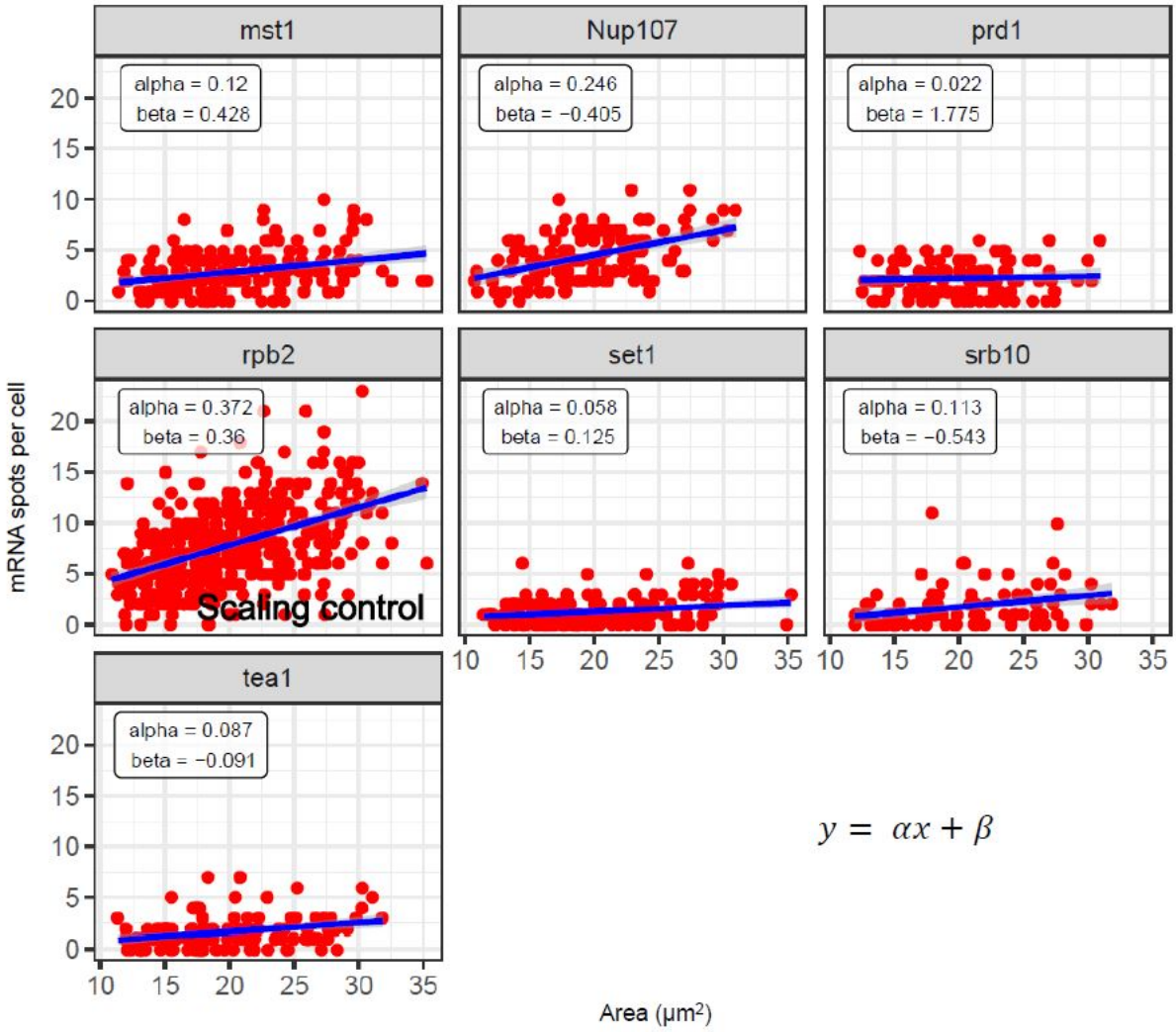


Figure 4.14: Transcript counts measured using smFISH for non-scaling candidates plus *rpb2* as an scaling control. *cdc2-M17as* cells without being exposed to 1NM-PP1 were fixed and treated using the protocol described in Chapter 2. Approximately a hundred cells were quantified for each transcript. Each red dot represents a cell, and the blue line is a linear regression of the counts with respect to cell area. The shaded area around it is the 95% confidence interval.

GFP quantification

Genes that do not scale with cell size could do so at either (or both) protein and mRNA levels. To study this behaviour at the protein level, I have taken advantage of the existence of strain library of 1000 genes tagged with GFP. I chose the genes that were studied by smFISH and also represented in the library, as well as other nuclear genes of interest with roles in the regulation of transcription.*taf12* is part of the SAGA complex, key to the initiation of transcription. A gene predicted to scale, *reb1*, was added as a negative control. Transcriptomics data for the chosen candidates is represented in 4.15, showing that the candidates show a decrease in concentration when cell size increases. The intensity of the GFP's fluorescence per cell is directly proportional to the absolute amount of protein, and it is easily quantified using light microscopy or ImageStream. ImageStream offers the advantage of being able to acquire both fluorescence data and cell size data for thousands of cells in a short amount of time. Moreover, it is equipped with an analysis software that allows easy extraction of different features of the images. Using this system, the concentration of the protein is calculated dividing the total GFP intensity by the cell's area, and then compared to see if concentrations are constant at all cell sizes. The data obtained for some GFP-tagged proteins can be found in figure 4.16. The running average of the data shows a downward trend, suggesting a small decrease in concentration as the cell grows larger for all the proteins studied. However, the scaling control *reb10* does not present a change in concentration different than the other genes. The data suggest that the concentration of these proteins scales with cell size, despite some of the transcript being non-scaling.

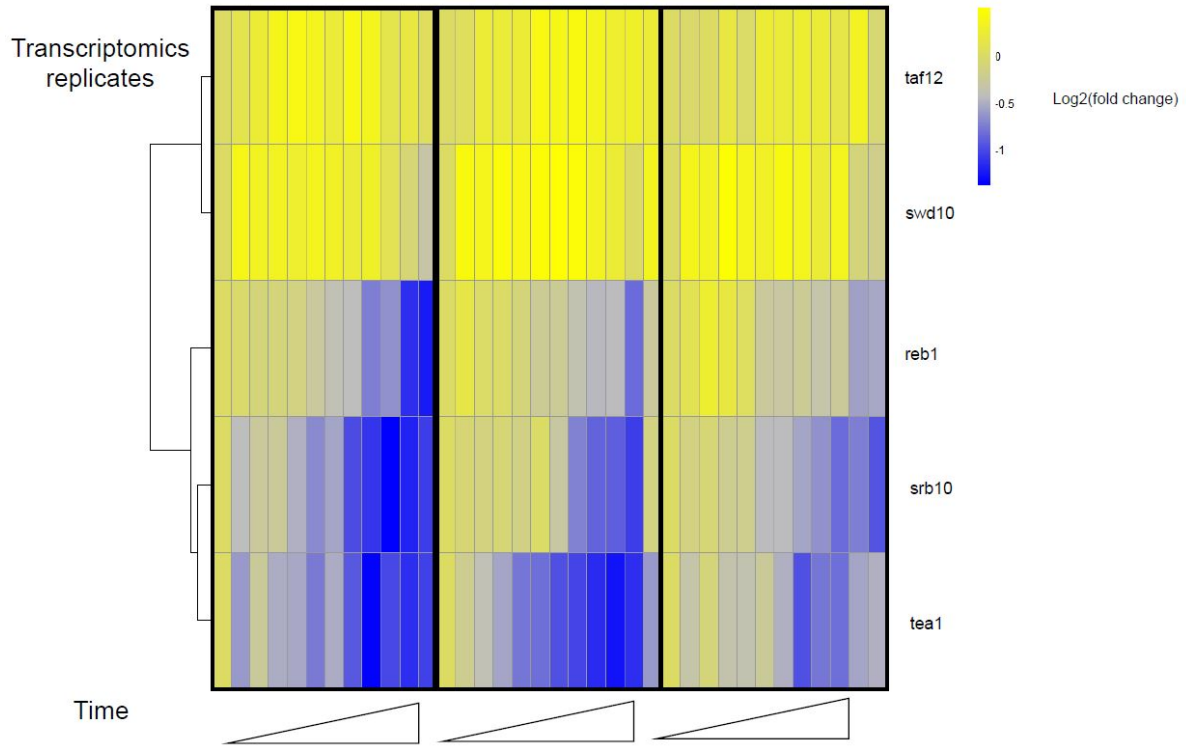


Figure 4.15: Transcriptomics data for the candidates genes that were studied in the next figure. Genes were chosen by their availability in the commercial GFP-tagged strain library and their subcellular localisation. I also included genes that were represented in figure 4.14

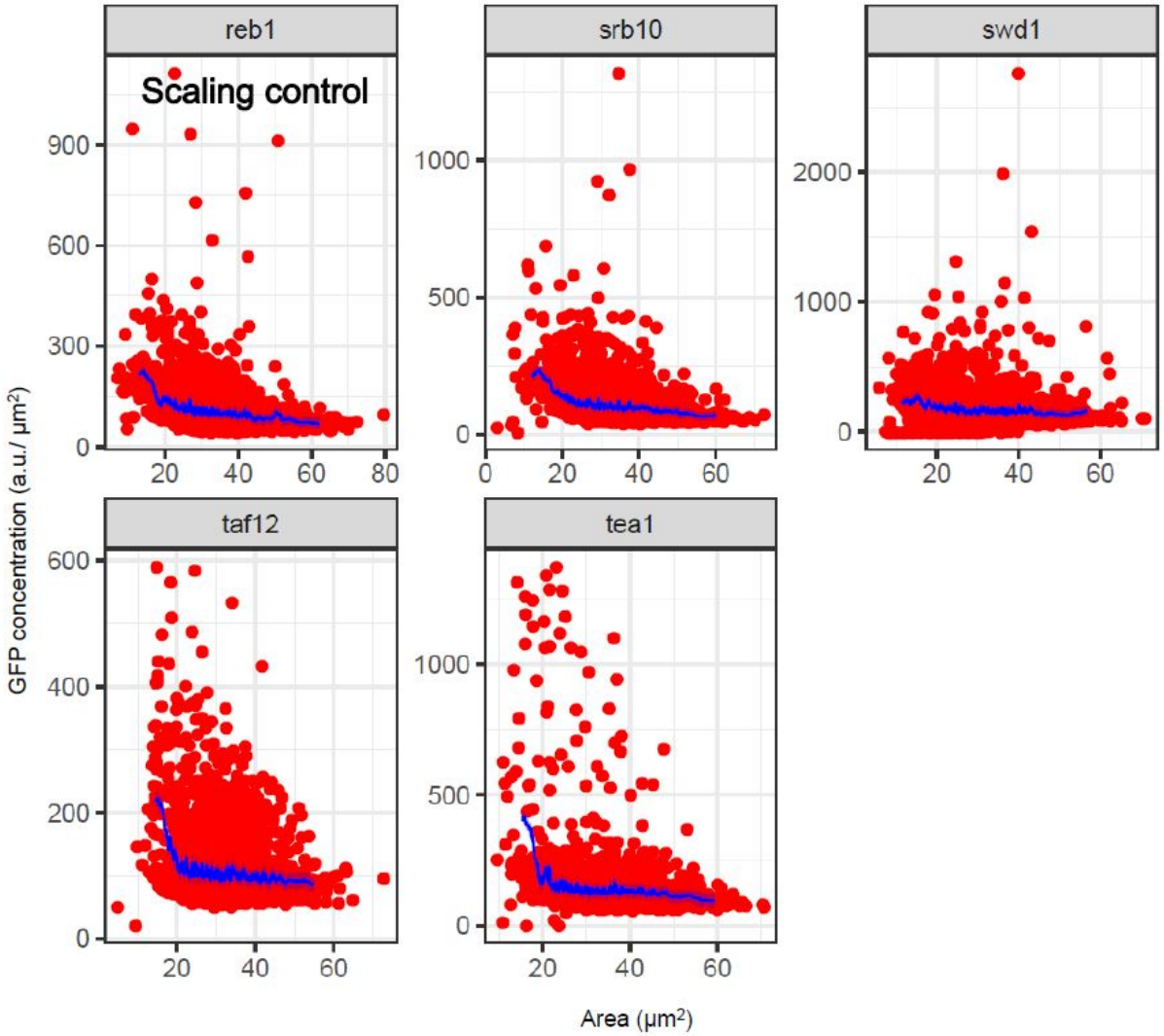


Figure 4.16: Cells with the gene of interest tagged with GFP were grown and the GFP intensity per single cells was quantified using ImageStream, together with cell size information. GFP concentration was calculated dividing the fluorescence intensity in the whole cell by its area. The dots represent single cell measurements. The solid line is a moving average of the data and the shaded line around it the 95% confidence interval obtained using bootstraping.

4.3 Conclusions

This chapter explores the behaviour and characteristics of DE expressed genes during the time course. As genes whose concentration changes with cell size, they have the potential to regulate size-dependent processes. Globally, the ribosomal content of the cell is going down whilst metabolic enzymes increase. A linear relationship between division rate and ribosomal fraction of the cell has been observed in bacteria [71, 131, 11] and is believed to be one of the fundamental laws of prokaryotic growth. It is unexpected to have a ribosome fraction that decreases when cells are growing in size exponentially, as shown by the cell size data. Together with the increase in metabolic enzymes could suggest a constraint in the cell resources. Resources in the cell are limited, and must be allocated to optimise division rate in response to the environment. Depending on the conditions, cells produce more ribosomes or metabolic enzymes but at the expense of the other category. This theory and its application in eukaryotes will be examined more deeply in chapter 5.

The coordinated response of a subset of genes points to a similar gene regulation mechanism. For instance, genes that accumulate faster than average seem to have a higher presence of heterochromatic marks than scaling genes. This, together with a decrease in the concentration of proteins required to maintain heterochromatin, suggests the possibility of a global decrease of H3K9me2 throughout the genome. Further experiments are needed to test this hypothesis. For instance, a western blot using anti H3K6me2 could confirm if there are changes in this modification related to cell size. Apart from the potential role of histone methylation, no other evidence of a specific mechanism regulation can be found in the data when looking at promoter sequences, chromatin landscape or other gene features. No common characteristic for all non-scaling genes has been found. This hints to an specific non-scaling mechanism for each gene. A negative feedback loop on transcription could keep the number of transcripts constant, for example. Cell-cycle regulated genes are also expected to not be proportional to the size of the cell, as they are transcribed in bursts. When they are not being transcribed, the transcripts and proteins produced by these genes are only subjected to degradation, making their concentration not constant as cell size increases.

Several non-scaling genes where chosen for experimental validation using smFISH and ImageStream, to quantify transcripts and proteins respectively. At the transcript level, two of the candidates (prd1, mst1) show the expected non-scaling behaviour. However, the same behaviour does not seem to be reproducible when looking at the gene product tagged with GFP. The difference in scaling properties between protein and transcript could be a result of the difference between the molecules half-life. A decrease in concentration can be observed for all proteins tested, included the negative control *reb1*.

Chapter 5

Exploring cellular economics at low DPR through mathematical modelling

5.1 Introduction

There is a plethora of proteins and transcripts involved in the cellular response to a decrease in DPR. In chapters 3 and 4, the behaviour of the individual molecules was studied. For the next section, I will take a different approach to understand how different pathways interact with each other. All pathways in the cell, specially the ones involved in gene expression, do not act in isolation and affect each others' behaviour. For example, changes in ribosome levels would affect the production of many proteins and alter, for instance, the polymerase levels and transcription rates. In turn, that will have an effect on the ribosome numbers and translation rates. Coarse-grained models reduce these complex pathways to a single reaction, allowing for an easier study of the interplay between complex cellular processes. For instance, translation involves hundreds of factors that interact in complex ways to produce a protein. In a coarse-grained model, all those reactions are modelled as one, in this case the ribosome binds to the transcript and produces a certain protein after consuming a number of precursors. This provides the advantage of having to deal with simple reactions and a fewer number of reactants, making the model easier to use and understand. Whole-cell coarse-grained models have been

used successfully to study the relationship between cellular allocation of resources and growth rate [130, 71, 56, 133].

In figure 3.12, I showed that the fraction of ribosomal proteins decreases during the time course. At the same time, the fraction of metabolic proteins expands. The anticorrelation between metabolic and ribosomal fractions has been described previously in prokaryotes [130, 56]), but not in the context of a changing cell size in a constant environment. A theoretical explanation for this behaviour has been reviewed in Scott et al. 2014 [131]. In summary, the cell has a limited number of proteins, and has to distribute them between different fractions in the most advantageous manner. In rich media, to maintain a faster growth rate protein production has to increase, hence a higher allocation of resources to produce ribosomes. This reduces the amount of resources that can be allocated for metabolic enzymes. When there are less external resources, the ribosomal fraction is reduced to be able to produce more metabolic proteins to sustain growth. However, for growing the cells used in chapter 3, I used a medium that has nutrients in excess. Therefore, fraction changes seem not to be due to external conditions. If this is the case, they could be revealing a change of how the cell needs to allocate its resources at extreme cell sizes.

Weiße et al. 2015 [158] proposed a model that simulates trade-offs in the cell due to a limited number of resources and specifically takes into account the interaction between several basic processes. The model simulates using ordinary differential equations (ODEs) the behaviour of a whole cell feeding on a substrate. The substrate is imported into the cell, where it is transformed by a metabolic enzyme into energy. Energy is consumed in translation, where all the proteins that form the cell are produced. Apart from the transporter and metabolic enzymes, the model also includes ribosomes that carry out translation and q-proteins. Q-proteins do not interact with any of the other molecules in the system, and represent housekeeping proteins. Transcripts are produced at a certain transcription rate. The key feature in the model is the implementation of three basic limitations in the cell. First, the levels of energy are finite, and need to be distributed between different kind of proteins. Ribosomes are also limited in number, with transcripts competing for being translated. Last, the cell has a finite mass. In other words, increasing the proportion of one kind of proteins implies that other species will

decrease.

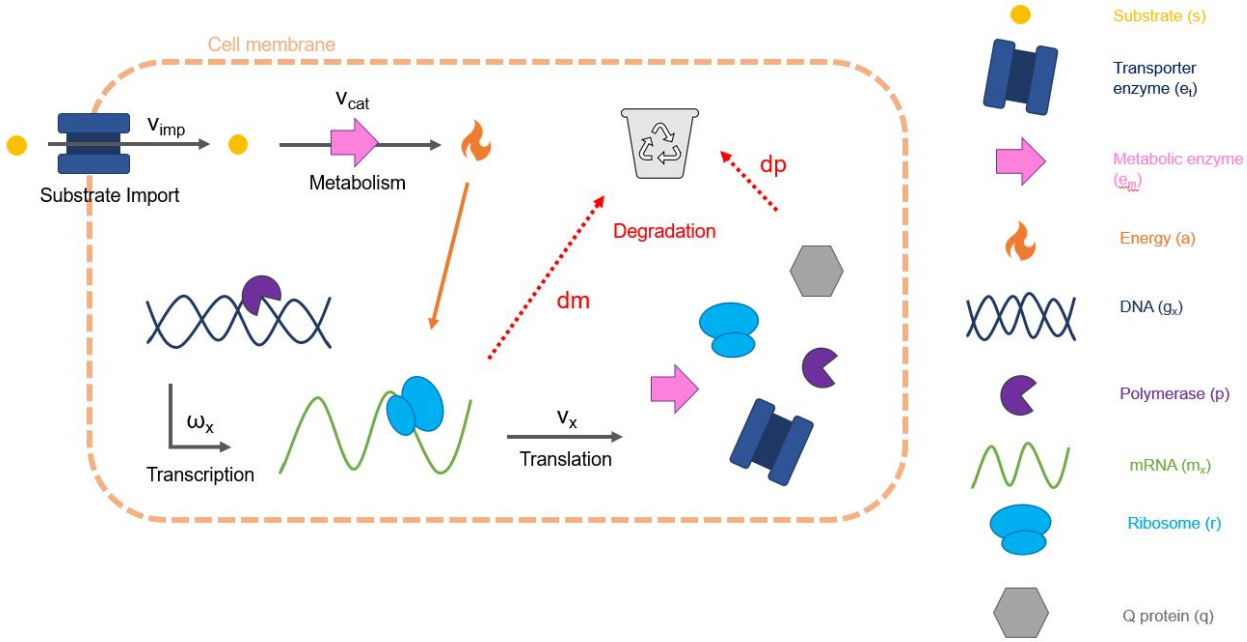
Using those three rules, they are able to recapitulate how cells allocate their resources in different environments, as well as the behaviour of exogenous genetic circuits when conditions change.

This chapter describes a new version of the model that includes the influence of cell size. The use of this model allows the exploration of several hypotheses, as well as generating predictions that can then be tested experimentally. How parameters affect the final behaviour of the model is also explored, as well as the hypothesis generated. This approach allows the study of how the processes considered change globally and interact when cell size increases and resources are limited.

5.2 Model description

I based the model on the one published in Weiße et al. 2015, but added some modifications to make it more suitable for the question at hand. First, reactions in the published model are dependent on the number of reactants. This assumes a constant size of the cell. As the system that we simulated increases in cell size, we modified the reactions so that they would depend on molecule concentrations instead of numbers, accounting for changes in reaction kinetics at different cell volumes. Moreover, dilution due to cell division is absent from our experimental system as *cdc2-M17as* cells do not divide when exposed to 1NM-PP1. All those terms were removed from the equations. I also modelled transcription more explicitly, simulating the binding kinetics of polymerases to genes. In the published version, transcripts are synthesised at certain transcription rate but it doesn't include polymerase molecules or genes. Including polymerases and genes, would allow the exploration the role of transcription in the allocation of resources at maximal sizes.

The model is composed of a series of ordinary differential equations that describe the concentration of the different species as they interact. Three fundamental physiological processes in the cell are included: metabolism, transcription and translation.



Transcription

$$g_x + p \overset{k_{2b},k_{2u}}{\longleftrightarrow} PG_x$$

$$PG_x \overset{\omega_x}{\rightarrow} g_x + p + m_x$$

Translation

$$m_x + r \overset{k_b,k_u}{\longleftrightarrow} c_x$$

$$n_r a + c_x \overset{v_x}{\rightarrow} r + m_x + x$$

Nutrient import and metabolism

$$s \overset{v_{imp}}{\longleftrightarrow} s_i$$

$$s_i \overset{v_{cat}}{\longrightarrow} n_s a$$

Protein and transcript degradation

$$m_x \overset{d_m}{\leftrightarrow} \emptyset$$

$$x \overset{d_p}{\rightarrow} \emptyset$$

Figure 5.1: Illustrative representation of the different species and reactions in the model, with the reactions specified below

Metabolism is simulated using two different reactions: transport of the substrate inside the cell, and its transformation through an enzymatic reaction into energy usable by the cell.

In the model, the cell is surrounded by a number of substrate molecules that represent nutrients that can be found in culture media. The substrate (s) is imported into the cell via a transporter protein (e_t). Nutrient import is simulated with Michaelis-Menten kinetics, with rate ν_{imp} :.

$$\nu_{imp}(e_t, s) = e_t \frac{v_t s}{K_t + s}$$

Once internalised, the substrate is referred to as s_i. After import, s_i is transformed into energy (a) by a metabolic enzyme (e_m), with rate ν_{cat}:

$$\nu_{cat}(e_m, s_i) = e_m \frac{v_m [s_i]}{K_m + [s_i]}$$

I modified of the original model to make the reaction rates inversely dependent on the volume of the cells. This adds the influence of cell size in biochemical kinetics. Cell mass has been shown to be proportional to cell size [11, 173], so I will use the total mass of the cell as a proxy for cell volume. In this case mass is calculated as the total number of amino acids that the simulated cell has at a certain point in time. Transcripts are included in the calculation of mass as well. To calculate the mass of transcripts in "amino acid" units, I use the ratio between the mass of an average amino acid and of an average nucleotide multiplied by 3.

$$mass = r + q + p + e_t + e_m + \sum PG_x + \sum C_x + \frac{3 * 324.3}{109} \sum m_x$$

For s_i, the concentration is calculated as follows:

$$[s_i] = \frac{s_i}{mass}$$

Energy units are an abstract representation of both energy (ATP, NADPH) and precursors

(amino acids). Energy is consumed only during translation, one unit per amino acid translated. Its expenditure in other processes is considered negligible How much energy is obtained per molecule of substrate depends on the parameter n_s or nutrient efficiency. It represents changes of substrate that are metabolised differently. Behaviour of a is described by the following equation:

$$\frac{da}{dt} = n_s \nu_{cat}(e_m, s_i) - \sum n_x \nu_x(c_x, a)$$

In order to study how polymerase concentration and gene dosage affects cell size and growth, transcription was represented explicitly. Again, this is modelled using two different reactions. First, polymerase molecules (p) bind genes (g_x) to form complexes (PG_x). Then this complexes will produce transcripts (m_x). Each protein in the model (ribosomes, transporter, metabolic enzymes and q-proteins) has its own corresponding transcript. The binding of polymerase and genes is a bidirecctional reaction controlled by two parameters: binding (k_{2b}) and unbinding (k_2u) rates. Polymerase-gene complexes produce transcripts at transcription rate (ω_x), releasing also the polymerase and gene molecule that can start another round of transcription. Despite not being consumed during transcription, the concentration of a still influences transcription rates (ω_x) through this expression:

$$\omega_x(a) = w_x \frac{[a]}{\theta_x + [a]}; [a] = \frac{a}{mass}$$

Transcription rates are also dependent on two parameters: maximal transcription rate (w_x) and the energy amount that produces half maximal transcription rates (θ_x). Q-protein transcription rates, however, follow a different expression. These represent housekeeping proteins, that do not change in number regardless of the external conditions. Microtubules or DNA-binding proteins would be examples of these proteins. To maintain their concentration throughout conditions, an auto-inhibition term is added to their transcription rate:

$$\omega_q(q,a) = w_q \frac{[a]}{(\theta_x + [a])(1 + (\frac{[q]}{K_q})^{h_q}}$$

For translation, transcripts bind to free ribosomes (r), forming translation complexes (c_x). They will then give rise to the 4 types of proteins that compose the model: transporters, metabolic enzymes, ribosomes and q-proteins.

Transcription rates (ν_x) are dependent on energy levels through the elongation rate γ:

$$\nu_x(c_x,a) = c_x \frac{\gamma(a)}{n_x}$$

The elongation rate, in turn, depends on the total energy levels of the cell with saturation kinetics:

$$\gamma(a) = \frac{\gamma_{max}[a]}{K_\lambda + [a]}$$

All proteins and transcripts are subjected to a constant rate of degradation, represented by $d_p$and d_m.

5.3 Results

5.3.1 Model behaviour with nominal parameters

The main purpose of this model is simulating the experimental system described in Chapter 2, to help understanding the relationship between the different pathways at play when cells reach maximal size. The strain used (*cdc2-M17as)* arrests in G2 when exposed to a drug that inhibits the activity of a mutant allele of *cdc2*. Cells keep growing from the tips reaching sizes larger than wild type. Single cell data shows that the cells stop growing after approximately 9 hours.The modelled cell has to also be able to increase in size, to then stop once the system

Parameter	Value	Description	Unit
w_r	930	maximal ribosome transcription rate	molecules/min.cell
$w_m = w_p = w_t$	4.14	maximal enzyme transcription rate	molecules/min.cell
w_q	948.93	maximal q-protein transcription rate	molecules/min.cell
d_m	0.1	transcript degradation rate	min^{-1}
d_p	0.001	protein degradation rate	min^{-1}
h_q	4	q-autoinhibition Hill coefficient	none
k_u	1	transcript-ribosome binding rate	cell/min.molecules
k_b	1	transcript-ribosome unbinding rate	min^{-1}
k_{2b}	1	transcript-gene binding rate	cell/min.molecules
k_{2u}	1	transcript-gene unbinding rate	min^{-1}
n_r	7459	ribosomal length	aa/molecule
n_x	300	metabolic proteins length	aa/molecule
n_p	3407	polymerase length	aa/molecule
n_s	0.5	nutrient efficiency	none
γ_{max}	1260	maximal translation elongation rate	aa/min.molecules
s	10000	external nutrient	molecules
v_t	726	maximal nutrient import rate	min^{-1}
K_t	1000	Nutrient import Michaelis constant	mocules
v_m	5800	maximal metabolic rate	min^{-1}
K_m	1000	Metabolism Michaelis constant	molecules
K	7	Translation elongation Michaelis constant	molecules/cell
$_r$	426.87	ribosome transcription threshold	molecules/cell
$_{nr}$	4.38	non-ribosomal transcription threshold	molecules/cell
K_q	152219	q autoinhibition threshold	molecules/cell

Table 5.1: Nominal parameters of the model. Obtained from [158] with some adjustments

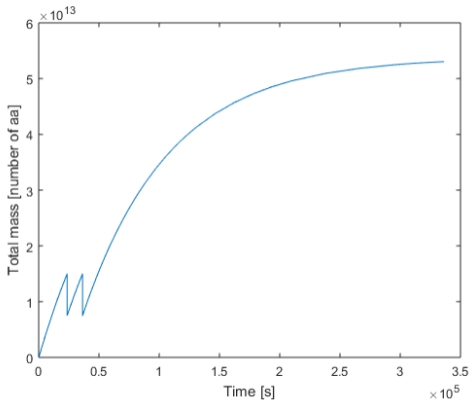


Figure 5.2: Behaviour of the total size of the cell during the simulations using the parameters in tables 5.1 and 5.2.

reaches steady state. In the experiment, cells divide similarly to wild type before arresting in G2 and starting to increase their mass. To include this in the model as well, 2 cell divisions are modelled before the arrest. The model chosen for cell division is a sizer, in which cells divide in two when reaching a certain size. *S. pombe* cells have been described to divide following a sizer-like mechanism [64]. A division size was chosen so that the difference between the limiting size and the division size was around 5-fold, as observed in the data. When the modelled cell reaches the division size, all the molecules are divided by two to give rise to a daughter cell that will continue growing. To simulate the arrest, no division is imposed, letting the system evolve (figure 5.2). Using the modifications mentioned previously, the *cdc2-M17as* behaviour can be reproduced qualitatively. After arrest, the modelled cell has a phase in which its total mass increases before reaching steady a constant mass at steady state.

The parameters used for this simulations are summarised in table 5.1. I used the ones obtained in [158] as a starting point to explore the model's behaviour and also due to the lack of experimental data describing some of the parameters in *S. pombe*. Table 5.2 presents the nominal parameters used for the simulations.

After checking that it is able to reach a limiting size, the model will be used to explore which parameters are controlling the maximal size that the cell arrives at. A sensitivity analysis was carried out, doubling the nominal parameters (5.1), and recording the maximal cell mass reached by these new parameter configurations (figure 5.3). Parameters involved in translation

Variable	Initial conditions [number of molecules]
e_t	1
e_m	1
r	10
a	1000
p	10
g_p	4
g_q	4
g_t	4
g_m	4
g_r	4

Table 5.2: Initial conditions for simulating the model

and molecule degradation are the ones that have a strongest influence in the largest size cells can attain. Total mass ultimately depends on the balance between molecule production and degradation. At the same synthesis capacity, higher degradation rates will lead to a lower total mass at the steady state, as it is shown by increasing both protein and transcript degradation rates. Increasing ribosome maximal transcription rate also leads to an increased limiting cell size (figure 5.3). Maximal size can also be influence by the efficiency of the formation of ribosome-transcript complexes. Increasing the binding efficiency of the reaction increases cell size, whereas a higher unbinding constant reduces it. The maximal elongation rate also regulates the extent of how much the cell can grow. An increase in this parameter would increase the potential output of ribosomes. Both results highlight the importance of not only the number of ribosomes, but also their activity levels. I also simulated the ODE system doubling the initial amount of molecules, with constant parameters. Regarding the initial conditions at which the simulation starts, only the number of ribosomal genes affects cell size, suggesting a possible limitation in ribosomal translation at extreme cell sizes (figure 5.4). It has been reported in the literature that an increase in the expression of transcription factors that control ribosomal translation produces a higher average cell size [149, 124]. In bacteria, evidence also points out to a direct relationship between the number of ribosomes, cell size and growth rate [130, 131, 73, 72].

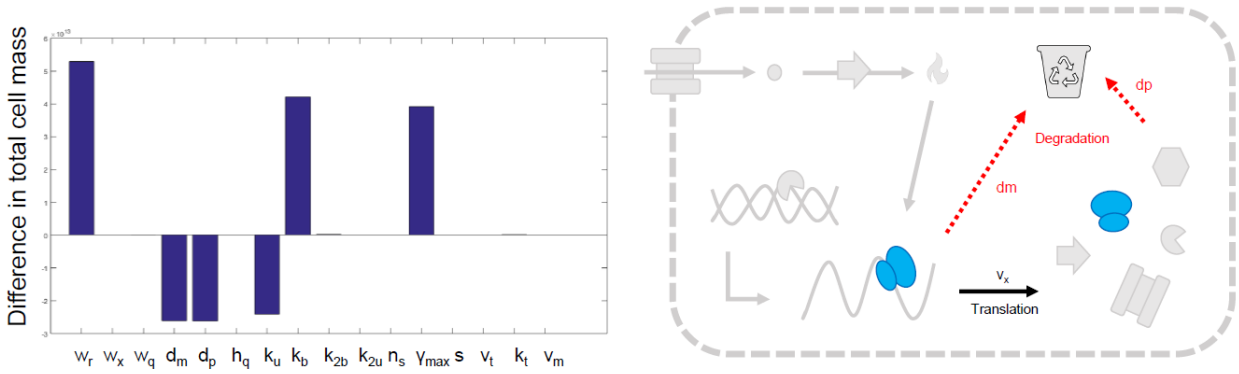


Figure 5.3: In order to understand the effect that the different parameters have in the maximal size the cell can attain, I carried out a sensitivity analysis. For the analysis, I simulated the model using the doubled value of each parameter, to then calculate the mass of the cell in the last time point of the simulation. The difference between the mass obtained with the new parameters and the one calculated with the nominal ones is represented in the barplot on the left. Parameters involved in translation and the degradation of molecules are the ones that increase the maximal size of the cell. The diagram on the right highlights the processes influencing maximal size in the context of the full model.

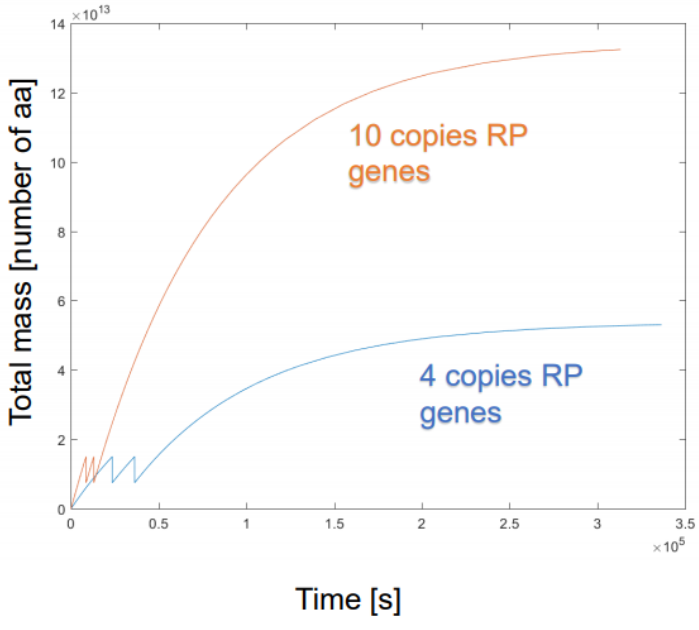


Figure 5.4: As increasing parameters related to translation also increase the maximal size the cell can attain, I simulated the model increasing the number of copies of ribosomal genes. An increase in copies is also accompanied by an enlarged maximal size, suggesting that a limitation in the transcription of ribosomes could influence size.

5.3.2 Parameter Fitting

The parameters used in the previous simulations were the ones used in Weiße et al. 2015. [158]. Those parameters were obtained by fitting the model to available *E. coli* data. Coming from a different organism, it is not clear how well these numbers apply to a fission yeast model. To get a new set of parameters applicable to *S. pombe*, I applied a parameter estimation technique, in this case an Aproximate Bayesian Computing (ABC) approach [144, 90, 97]. I used the Julia language implementation kindly provided and developed by Anthony Bowman (software not published). The algorithm needs "priors", or an initial distribution of the parameters to fit. Priors incorporate the knowledge that we have a priori of what the parameter value could be. From these priors, the algorithm samples a new set of parameters that are used to simulate the model. Finally the difference between the simulation and the experimental data is calculated. If it falls below a certain threshold the parameters are accepted, if not the algorithm starts again. At the end, we end up with the "posterior" distributions of the parameters. They give information not only about the possible values of the parameters, but also about their uncertainty. One of the challenges is sampling the parameter space extensively enough to get an accurate distribution of the posteriors. In this case, an Sequential Monte Carlo (SMC) technique is applied for examining the parameter space. This approach uses a series of intermediate distributions to approximate the posterior. In each iteration the acceptance threshold is reduced, in order to narrow down the posterior distribution to the tolerance needed [90, 83, 97, 144].

I started by fitting the model to the fraction of ribosomal and metabolic proteins calculated in chapter 3 (figure 5.5). The posterior distribution of the majority of the parameters spans the same values as the priors given. This means that the algorithm could not find a tighter posterior distribution, suggesting that the model behaviour may not be very sensitive to these parameters. Other parameters, like γ_{max}, are tightly constrained. Moreover, there is no correlation of the parameters between each other (figure 5.6). However, for the fractions to change independently, it would be more adequate to have different degradation rates for different species. Another possibility would be to make protein degradation for each fraction to be responsive to another

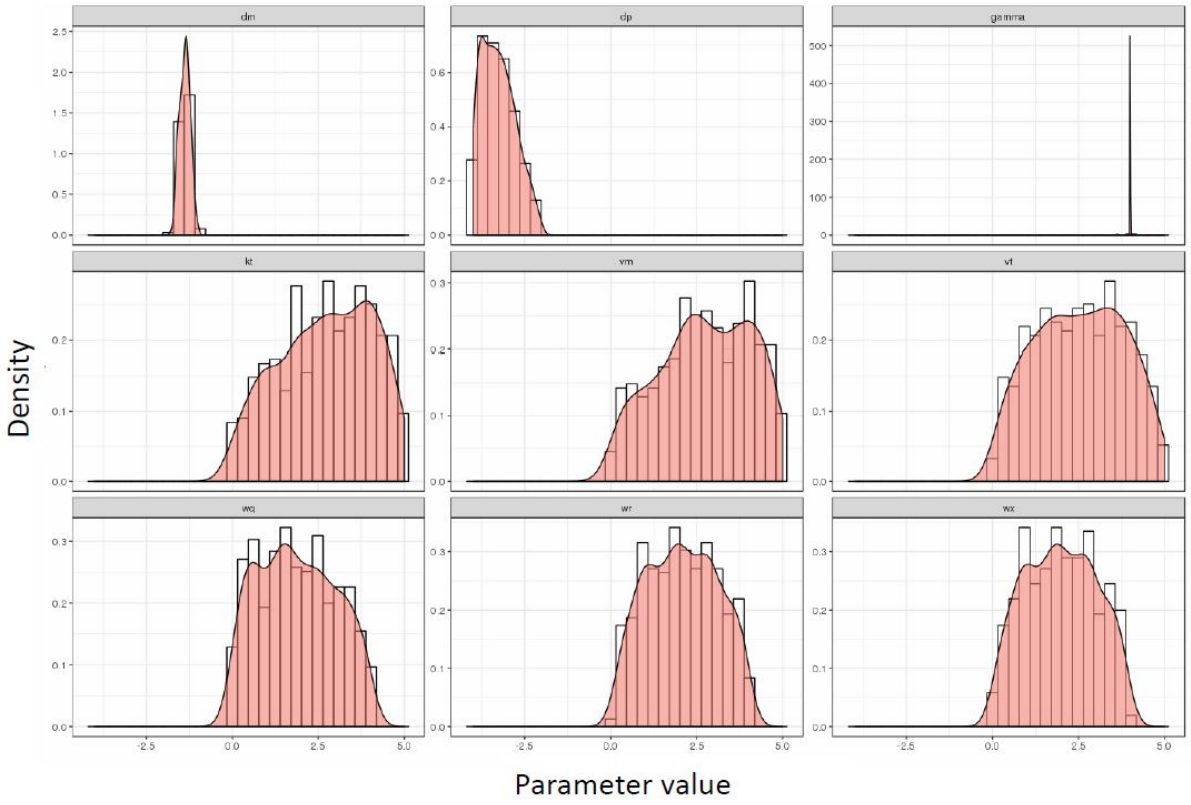


Figure 5.5: Distribution of the posterior probability after applying the ABC approach to fit the model to the fraction data from figure 3.12

characteristic of the cell. The ABC strategy developed also has model selection implemented, and could be used to decide between different implementation of the model to accomodate differences in degradation rates.

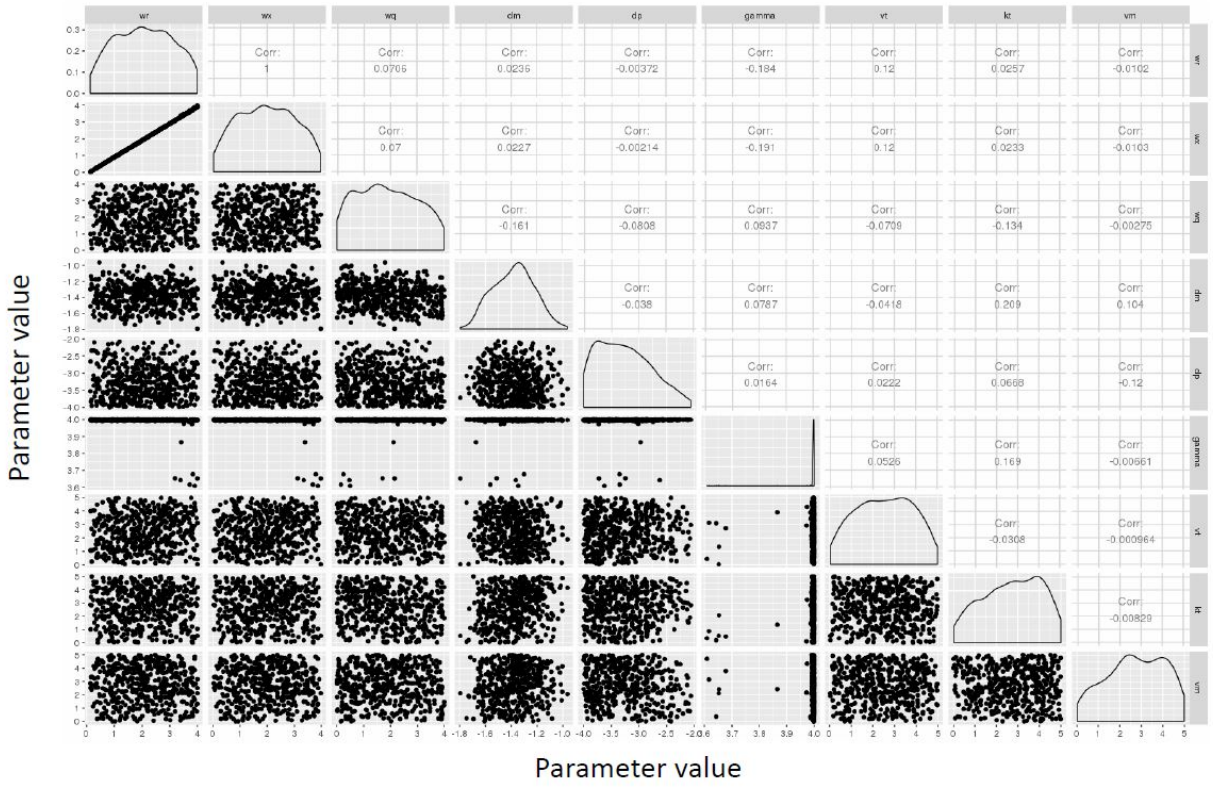


Figure 5.6: Correlation between the posterior distributions of all possible parameter pairs. There does not seem to be any correlation between parameters, except the transcription rates of ribosomes and metabolic enzymes. Larger ribosomal pools require more precursors to be efficient, thus the correlation between the two. 3.12

5.4 Conclusions

This chapter describes an ordinary differential equations model adapted to study the interaction between different processes in the experimental system described in Chapter 3. The model is able to qualitatively reproduce the cell size trend observed in experimental data. When checking the influence of the parameters in the final cell size attained by the cell, the one related to ribosomes and their transcription are the ones having the biggest influence in the size of the cell. Taking this with the experimental data that reveals a decrease in the proportion of ribosomes in the cell as the size increases, it can be hypothesised that there is a limitation on the number of ribosomes during growth. An experimental system where ribosome amounts could be increased or reduced would be able to test this possibility. The mechanism that regulates ribosome production has mostly been described in *S. cerevisiae*. It is under the control of *TORC1*, with two proteins downstream, *SFP1* and *SCH9*. *SFP1* is a transcription factor that is phosphorylated when *TORC1* is activated, inducing ribosomal genes and other genes involved in ribosomal biogenesis. *SCH9*, on the other hand, is a kinase that is also activated by *TORC1* phosphorylation. In turn, this kinase phosphorylates an array of targets that induce the expression of rRNAs, ribosomal proteins and the ribosomal biogenesis regulon [62, 82]. Overexpression of the fission yeast counterparts of these proteins gives rise to larger cells [149, 124]. Seeing how overexpressing these protein in the *cdc2-M17as* strain affects the limiting cell size, would test if the ribosome number is what is limiting at a low DPR. Going back to the data presented in Chapters 3 and 4, ribosomes are mostly found in the negatively non-scaling genes. A limitation in their number could be the explanation for the limitation in the growth capacity of the cell, as the model shows how they are key in setting the growth potential of the cell.

The transcriptomics and proteomics datasets show an increase in molecules related to the proteasome during the time course. The proteasome is a complex that degrades proteins in the cytoplasm to single aminoacids, which can then be used by metabolic reactions or to make new proteins. An increase in its numbers could be related to an increase in the demand of precursors, that can be recycled from other proteins. These hypothesis can be studied by modifying the

current model to produce energy units after the degradation of molecules.

Despite its caveats, the model still allows for a theoretical exploration of the influence of different parameters in maximal cell size, making predictions that can then be tested experimentally.

Chapter 6

Discussion

In this thesis, I explored in the possible mechanisms behind the coordination of cell size and gene expression. To that end, I used a genetic system in fission yeast: the strain *cdc2-M17as*. In this strain, the cells arrest in G2 after being exposed to a drug that inhibits the activity of Cdc2. When arrested, these cells keep growing in length without replicating their genomes. I studied the accumulation of transcripts and proteins as the cell increases in size. Quantitative data for transcripts and proteins was obtained using RNA-seq and mass-spectrometry based proteomics respectively. Mining this data allowed the identification of pathways affected by the increase in cell size, as well as genes whose concentration is not proportional to size. Studying the characteristics of genes not subjected to the coordination of cell size and gene expression, could shed light on the molecular mechanisms behind this fact. To better understand the system, I developed a published mathematical model to replicate growth without division, based on a system of ordinary differential equations. The model was used to investigate how different physiological processes interact when resources are limited.

6.1 Effects of low DPR in the transcriptome and proteome

In chapter 3, I presented a time series proteomics and trancriptomics dataset. Samples come from the *cdc2-M17as* strain, that upon the addition of 1NM-PP1, arrests in G2 and keeps growing in size without dividing. The most striking characteristic of these datasets is the seemingly reduction in the concentration of ribosomes as the cell accumulates mass. In proliferating cells, the ribosomal fraction has been shown to correlate with cell size and elongation rate [130, 129]. The reduction of these proteins could be due to an activation of the stress response, as it has been described that the growth and stress modules are anti-correlated [91] . However, the overlap between the negative non-scaling transcripts and the mRNAs downregulated in stress conditions is not significant. Results from the modelling approach also hint to a potential limitation in the transcription of ribosomal proteins. Metzl-Raz and colleagues [105] have described how resources are allocated in budding yeast when exposed to different media. Strikingly, they discovered that part of the ribosome pool stays inactive depending on the conditions. This would allow the cell to respond rapidly when the environment changes, without needed to synthesis more ribosomes if the growth rate increases. This piece of evidence raises another possible explanation for the behaviour of ribosomes as the cell increases in size. The cell does not increase its number, but it might be increasing the proportion of active ribosomes, which can not be measured using proteomics. A polysome profiling experiment would provide information about the proportion of active ribosomes, by calculating the polysome to monosome ratio. This question could also be explored using the theoretical model, by adding an inactive ribosome species that can switch to the active form. Studies in *S. cerevisae* have shown that ribosome biogenesis plays an important role in the coordination of cell size and division [62, 61]. Ribosome biogenesis could integrate signals from the environment and transduce them into a size threshold by limiting the amount of proteins that can be produced.

It is unclear what happens to the cell's viability after it reaches its lowest DPR limit. Visual inspection hints to a decrease in viability at maximal size, however it is still untested if these cells

have metabolic activity, suggesting a possibly terminally differentiated state, or they are not alive after some time at maximal size. Quantifying the oxygen consumed by the culture, using for instance the Agilent Seahorse, would provide information about the activity of metabolism at different sizes and if cells are viable at maximal size.

The evidence suggesting a decrease in cell viability raises another point, if cells lose viability because there is a physical limitation to how much they can elongate. It could be argued that the microtubule system that transports growth related-factors to the tips cannot be sustained at larger sizes. Physical properties of the cell wall could also prevent cells from reaching such sizes. Creating a diploid strain with two *cdc2-M17as* alleles and quantify the maximal size it can reach, would shed light on the question. If the limitation is due to the physical properties of the cell, the maximal size would be similar to the one of a haploid cell. A larger maximal size would indicate a role of ploidy in setting the maximal synthesis capacity of the cell.

All histone subunits can be found in the negative non-scaling subset. Histone amount has been described to drive transcriptional activation in zebrafish and *Xenopus* eggs [4, 63]. Fertilised eggs in both organisms go through several rounds of division without growth, relying on the proteins and nutrients previously contained in the egg. In every division, the number of histones per cell is reduced. The titration of histones against DNA has been shown to control the transition from those rapid divisions to the elongation of the cell cycle in the MBT. Zygotic transcription is also activated in this transition. The experiments in zebrafish show that the amount of histones bound to the genome does not change, but there is a decrease in free histones as the cell divisions progress. Their evidence points to a competition between transcription factors and histones for the binding of DNA. When the number of free histones decreases, chromatin becomes more accessible to transcription factors that initiate zygotic transcription [63]. The reduction in histone concentration in the dataset could suggest a similar mechanism, with less free histones to compete for the binding to DNA. Further experiments would be needed to confirm this hypothesis. A ChIP-seq of histones during the time course would produce data on the amount of histones bound to DNA at a particular cell size. If it is constant, the histone decrease shown in both omics dataset would be due to a decrease in free histones. Overexpressing histones and quantify their effect in cell

Positive non-scaling genes paint another picture. The proportion of metabolic proteins increases with cell size, specially enzymes related to the synthesis of amino acids and nucleotides. This could hint to a potential shortage of monomers needed to produce new proteins and transcripts when cell size becomes extreme. The anti-correlation of metabolic enzymes and ribosomes has been described in prokaryotes. A theoretical approach to describe this behaviour has been detailed in a series of papers authored by T. Hwa [131]. The theory relies on the assumption that the cell has limited resources that need to be distributed to maximise division rate. In poor conditions, cells devote more of this resources to create the needed metabolic proteins at the expense of ribosomes. Richer environment allow for a higher expression of ribosomes, as not as many enzymes are needed to maintain growth. The fact that those fractions show a similar behaviour in the enlarged *cdc2-M17as* could point to a shortage of external nutrients. However, the media used for the experiments has nutrients in excess. Experiments in our laboratory (data not shown) have shown that some nutrients need to be diluted a thousand times to have an effect in the division rates of wild type cells. Since the external conditions are most likely not responsible for the changes in metabolism, it could be due to changes inside the cell. As the cell gets larger, the concentration of certain internal metabolites could decrease. This could trigger the activation of pathways related to starvation, despite the rich environment. For instance, the TOR pathway responds to the concentration of amino acids inside the cell. Besides, higher rates of transcription and translation that accompany larger sizes would consume more precursors and deplete the cell reserves at a higher rate.

The increase in cell size through elongation also has an effect in the concentration of some membrane transporters. Some of these proteins seem to decrease in concentration as the cell grows. One of their roles is importing nutrients from the environment and sensing their concentration. The decrease in concentration of these receptors could seem to the cell a reduction in the concentration of external nutrients. However, proteomics evidence for membrane proteins must be taken with reservations, as membrane proteins are more difficult to solubilise and would be under-represented in the dataset.

Testing the validity of this hypothesis would require additional information about the accumulation of metabolites at larger cell sizes. A mass-spectrometry based approach could reveal how

the pools of different metabolites are affected by growth and size. Additionally, if metabolism is key in the interaction between cell size and gene expression, the exposure to varied environments could affect the accumulation of mass and the maximal size the cell can reach. An array of media containing different nitrogen sources has been characterised in fission yeast [22], as well as media that promote respiration instead of fermentation [95]. Examining how the cell size if affected by the external environment would show the importance of metabolism in maintaining the proportionality between gene expression and cell growth.

Metabolism and chromatin are very tightly connected, as the precursors used by chromatin modifiers come from enzymatic reactions in the cytoplasm [101, 171, 57]. For instance, histone acetylases get the acetyl group from acetyl-CoA directly reflecting the levels inside the cell [142, 171]. Histone methylation depends on the availability of S-Adenosyl methionine (SAM). The transcriptomics data shows an upregulation of all genes present in regions of the genome where histones are highly methylated. This modification is usually confined to telomeres and centromeres, inhibiting their transcription and compacting their structure. Genes close to them are also affected by the modification. The proportion of some histone methylases also decreases during the time course. The two findings together suggest a demethylation of the genome when cell size increases. To further prove this hypothesis, global histone methylation levels will be tested using western blot in chromatin extracts. There are also available commercial Enzyme-Linked ImmunoSorbent Assays (ELISA) kits that make the detection of methylation quantitative. In fission yeast and multicellular organisms (*Drosophila melanogaster*, for example) small RNAs have shown to play an important role in establishing and maintaining heterochromatic regions. Through small RNA sequencing, I could test if there are changes in the small RNAs population in the cell, given there is a variation in the chromatin methylation levels. The regulation of lncRNAs during the time course could also be a sign of a change in the chromatin status of the cell. The role of histone methylation in cell size increase could be also proved using drug that inhibit the activity of methylases, with two compounds identified with this activity in fission yeast [24].

Histone acetylation plays a key role in regulating transcription globally [142, 171]. Enzymes that carry out this modification have been traditionally associated with regulating specific

loci, however one of then has been proposed to have also an untargeted role. The enzyme proposed is *mst1*, the catalytic subunit of the NuA4 acetylating complex. A similar complex with less subunits, picNuA4, is the one thought to be responsible of untargeted acetylation. In addition, *mst1* is the only acetylase that is necessary for the cell to be viable, highlighting its crucial role in the cell physiology. The untargeted activity of the PicNuA4 complex would maintain global levels of acetylation in the genome [45]. There is a potential role for this modification to influence the global regulation of transcription rates. Western blot would be the first experiment to perform, to quantify the changes in chromatin acetylation at different cell sizes. If the changes in global acetylation of histones are confirmed, the use of an *mst1* thermosensitive allele would provide data on the effect of removing this enzyme at different cell sizes.

Despite the information obtained from the genome-wide expression dataset, interesting targets need to be confirmed. If a gene influences the maximal size the cell can attain, then removing it from the genome would produce a change in the phenotype, either in the elongation rate or in the largest size the cells can reach. The influence of all possible deletion mutants in non-essential genes can be investigated using a synthetic genetic array (SGA). The technique takes advantage of a commercial library that contains all deletion mutants in non-essential genes (Bioneer). The strain of interest, *cdc2-M17as* in this case, is crossed with each of the deletion strains in the library with the help of a robot. Double mutants are selected using the two antibiotic markers. Once all double mutants are selected, their influence in the phenotype can be scored. Either single double mutants of candidates of interest can be selected for further tests, or high-throughput microscopy can be used to quantify cell size in all strains in the limiting conditions. The thermosensitive allele of *cdc2* could also be used to facilitate the scoring of the double mutants. Thanks to this allele, double mutants can be grown at 25°C as if they were wild type cells until a colony is formed. Switching the cells to 36.5°C would inhibit the protein and trigger the arrest. Colony size change would then be quantified as a proxy for cell elongation. This approach would require less equipment, and allows for the collection of time course data of the colony size. As the analogue needs to be added to the plates, it does not allow the initial growth of the cells to form a colony whose size can be measured.

I attempted to create a strain suitable for an SGA analysis, but so far I have not succeeded. Since the commercial deletion library has a kanamycin resistant cassette, the query strain needs to have another marker. In this case, I tried to add a nurseothricin resistance cassette at the end of the *cdc2* locus.

6.2 Non-scaling transcripts and proteins

I used the transcriptomics and proteomics datasets to identify genes whose concentration does not scale with cell size. I divided them in two groups: positive, when the proportion increases, and negative when the molecule gets diluted. Studying what makes these genes different from the rest of the scaling molecules, could give us clues of what mechanism could be behind the coordination between cell size and genome output. Furthermore, the non-scaling proteins described in literature have very important roles in the regulation of cell size and triggering developmental events. Investigating non-scaling transcripts and proteins could also uncover new factors that act as concentration dependent triggers in the cells physiology. Characterising non-scaling proteins and transcripts didnt produce a conclusive answer as to why they are able to escape the ubiquitous regulation that affects the majority of the genome. There is no evidence of a particular transcription factor having an influence or a strong relationship with the chromatin structure. It is possible that each molecule is regulated independently to achieve its non-scaling property, instead of all of them sharing particular characteristics. Even though there does not seem to be an overarching characteristic between promoters of non-scaling genes, the behaviour could still be encoded in individual promoters. Swapping promoter sequences between a scaling and a non-scaling gene would further test if promoter sequence has an influence in the scaling behaviour of a transcript. Intriguingly, negative non-scaling genes contain DNA and membrane binding proteins. Neither of them scale linearly with cell size in the genetic system used. When adding 1NM-PP1 to the strain *cdc2-M17as*, DNA content is constant since there is no replication of the genome when cells are arrested in G2. There is a possibility that proteins that interact with DNA are proportional to the ploidy of the cell, and not to the size of the cytoplasm. For instance, transcription factors could be

sensitive to the number of binding sites in the DNA. A simple negative feedback loop in which transcription factors inhibit their own transcription could produce a constant amount of this proteins in the nucleus.

When cell size increases, cytoplasmic membrane increases proportionally to the surface volume of the cell instead of the total volume. Therefore, membrane-bound proteins would not need to scale with cytoplasmic volume. Indeed there is a decrease in concentration of these proteins as cell size increases. Experiments in mice livers has shown a response in the lipidome when size increases, with an increase in species related to the mevalonate pathway [108]. These molecules are involved in the synthesis of cholesterol, an integral part of the eukaryotic cytoplasmic membrane. In Miettinen et al. 2014, metabolites involved in the mevalonate pathway are hypothesised to act as a proxy for the ratio between the surface area and the cell volume. This could provide a possible mechanism is which the cell regulates the production of proteins that are attached to the membrane.

Fission yeast only elongates from the tips, where several complexes that control cell polarity reside. It is interesting to note that a component of these complexes, Tea1, presents a negative non-scaling behaviour. Despite the increase in cell size, the cell still only needs two of each complexes sitting on the cell tips. Their concentration would be diluted as the cell size increases together with the number of proteins of the rest of the proteome.

Confirming all non-scaling transcripts by experimental means would be a daunting task. Quantifying their single cell copy number through smFISH is possible, but achieving the necessary number of cells for an accurate quantification is a time consuming task. Interesting targets would then have to be selected for further validation. In the case of non-scaling proteins, quantification of GFP fluorescence and cell size in high number of cells can be achieved using ImageStream for a high number of samples. However, the challenge here is to be able to grow numerous GFP tagged strains in a satisfactory manner for posterior quantification.

6.3 Theoretical approach

I developed a model based on the one published in Weiße et al. 2015 to try to understand how lmitations in the cells' resources could affect the increase in cell size. The new version of the model includes the influence of cell size in reaction rates, as well as a more explicit description of transcription. In the published model, transcription was represented as a constant rate of transcript production. I included the binding of polymerases to genes, to investigate the effect of a limiting template in transcription. Changing the model parameters highlighted the influence of ribosomes in setting the maximal mass the cell can reach. Moreover, increasing the number of ribosomal genes also makes the cell grow larger. This could suggest a limitation in the transcription of ribosomes in the system. Ribosomes decrease in the experimental data. Their decrease in concentration could be behind the limitation of synthesis capacity of a single genome in the experimental system. To test this hypothesis, I tried to create a strain with an increased amount of ribosomes in the *cdc2-M17as* background. There are two proteins that have been shown to control the amount of ribosomes in the cell: *sck2* and *sfp1*. Both are effectors of the TOR pathway and when their transcription is increase, produce cells larger than wild type. To overexpress them, I will switch their promoters for the one from the gene *nmt1*. The expression of *nmt1* is dependent on the external concentration of thiamine. Presence of thiamine inhibits the activity of the promoter, although minimal concentrations of thiamine can be used for a partial repression of transcription. In absence of thiamine, the gene downstream the promoted is overexpressed. Another option would be to add an extra copy of those genes somewhere else in the genome. This approach would not affect the endogenous regulation of the genes, as their wild type promoter is kept. If the overexpression of these factors and subsequent increase in ribosome content has an effect in how cell size changes in the *cdc2-M17as* strain, it would suggest that there is a limitation of ribosomes. The model used can also be further modified to explore other questions. For instance, the model presented relies only on energy that represents both amino acids consumed in translation and ATP. Separating energy into two different molecules, precursors and ATP would allow to study the influence of protein recycling in size limitation. Amino acids would be produced when proteins degrade, as

well as by normal metabolism. The data shows an upregulation of the proteasome, a complex that degrades proteins to aminoacids that can then be recycled to produce new polypeptides. Energy and precursor consumption by transcription could also be implemented to investigate if transcription and translation are competing in terms of resources. In the model degradation rates are constant, regardless of the metabolic status of the cell. It would be interesting also to make them dependent on the energy status of the cell. This could replicate the cells switching to a more catabolic metabolism when nutrients are perceived to be low.

6.4 Conclusions

In this work, I have investigated the mechanisms behind the coordination of cell size and gene expression. Furthermore, I have characterised the effects of growth in both the transcriptome and the proteome. The major findings of this document are the following:

- Possible limitation of ribosomes when cells reach their low DPR limit
- The majority of proteins concentrations stay constant when cell size increases
- The mechanism behind non-scaling molecules is most likely specific for each species, as there is no common characteristics for non-scaling transcripts or proteins.
- A theoretical modelling approach predicts a limitation in the transcription of ribosomes when cell size is maximal.

Appendix A

Gene lists

A.1 Positive non-scaling transcripts

A.1.1 Gene List

Systematic ID	Name	Product description
SPBC106.04	ada1	adenosine deaminase Ada1
SPBPB21E7.07	aes1	phenazine biosynthesis PhzF protein family
SPAPB24D3.10c	agl1	maltose alpha-glucosidase Agl1
SPAC26F1.14c	aif1	mitochondrial inner membrane anchored oxidoreductase, apoptosis-inducing factor homolog Aif1 (predicted)
SPAC22F3.15	aim41	mitochondrial aspartyl/glutamyl-tRNA amidotransferase subunit B-related Aim41
SPAC15A10.08	ain1	alpha-actinin
SPCC1281.04	akr7	pyridoxal reductase (predicted)
SPAC13D6.05	alp11	tubulin specific chaperone cofactor B
SPAC8E11.07c	alp31	tubulin specific chaperone cofactor A, Alp31
SPNCRNA.1309	alr2-antisense-1	antisense RNA (predicted)
SPCC1919.03c	amk2	AMP-activated protein kinase beta subunit Amk2

Systematic ID	Name	Product description
SPNCRNA.771	arb1-antisense-1	antisense RNA (predicted)
SPAC6F6.10c	arc2	ARP2/3 actin-organizing complex subunit Arc34
SPAC17G8.04c	arc5	ARP2/3 actin-organizing complex subunit Arc5
SPBC4F6.18c	arf1	ADP-ribosylation factor, Arf family Arf1
SPAC3H1.07	aru1	arginase Aru1
SPBC800.05c	atb2	tubulin alpha 2
SPAC22F3.02	atf31	transcription factor Atf31
SPAC6F6.12	atg24	autophagy associated protein Atg24 (predicted)
SPBP8B7.24c	atg8	autophagy associated protein Atg8
SPNCRNA.980	ayr1-antisense-1	antisense RNA (predicted)
SPAC1556.03	azr1	serine/threonine protein phosphatase Azr1
SPBC1289.16c	cao2	copper amine oxidase-like protein Cao2
SPAC20G8.05c	cdc15	extended Fer/CIP4 (EFC) domain protein Cdc15
SPBC14C8.07c	cdc18	MCM loader
SPNCRNA.1413	cft2-antisense-1	antisense RNA (predicted), possible alternative UTR
SPAC8C9.03	cgs1	cAMP-dependent protein kinase regulatory subunit Cgs1
SPAC26A3.05	chc1	clathrin heavy chain Chc1 (predicted)
SPAC6C3.04	cit1	citrate synthase Cit1
SPBC337.15c	coq7	ubiquinone biosynthesis protein Coq7
SPAC19G12.11	coq9	ubiquinone biosynthesis protein Coq9 (predicted)
SPAC1420.04c	cox1101	cytochrome c oxidase assembly protein Cox1101/ mitochondrial ribosomal protein Rsm22 fusion protein
SPAC19B12.13	cox1102	cytochrome c oxidase assembly protein Cox1102/ mitochondrial ribosomal protein Rsm2202, fusion protein (predicted)
SPAC16C9.05	cph1	Clr6 histone deacetylase associated PHD protein-1 Cph1

Systematic ID	Name	Product description
SPAC22A12.08c	crd1	cardiolipin synthase/ hydrolase fusion protein Crd1 (predicted)
SPCC757.07c	ctt1	catalase
SPNCRNA.653	cut3-antisense-1	antisense RNA (predicted)
SPBC337.06c	cwf15	complexed with Cdc5 protein Cwf15
SPBP8B7.25	cyp4	cyclophilin family peptidyl-prolyl cis-trans isomerase Cyp4
SPAC22A12.11	dak1	dihydroxyacetone kinase Dak1
SPAC977.16c	dak2	dihydroxyacetone kinase Dak2
SPAC167.05	dbs2	Usp (universal stress protein) family protein
SPAC2C4.07c	dis32	3'-5'-exoribonuclease activity Dis3L2
SPAC1805.08	dlc1	dynein light chain Dlc1
SPAC8E11.03c	dmc1	RecA family ATPase Dmc1
SPAC3A11.10c	dpe1	dipeptidyl peptidase, unknown specificity, implicated in glutathione metabolism (predicted)
SPNCRNA.947	dph4-antisense-1	antisense RNA (predicted)
SPCC285.11	dsc5	UBX domain containing protein required for Sre1 cleavage
SPAC644.05c	dut1	deoxyuridine 5'-triphosphate nucleotidohydrolase (predicted)
SPBC1271.04c	dys1	eIF-5A-deoxyhypusine synthase Dys1 (predicted)
SPCC1223.10c	eaf1	RNA polymerase II transcription elongation factor SpEAF
SPAC1705.03c	ecm33	extracellular leucine-rich repeat domain, receptor L domain-like Ecm33
SPBC1604.01	egt1	Ergothioneine biosynthesis protein Egt1
SPBPB21E7.01c	eno102	enolase (predicted)

Systematic ID	Name	Product description
SPNCRNA.1302	eno102-antisense-1	antisense RNA (predicted)
SPBC337.09	erg28	Erg28 protein (predicted)
SPAC26F1.04c	etr1	enoyl-[acyl-carrier protein] reductase (predicted)
SPBPB8B6.03	fah1	fatty acid amide hydrolase Fah1 (predicted)
SPBC1773.01	far8	SIP/FAR complex striatin subunit, Far8/Csc3
SPBC1198.14c	fbp1	fructose-1,6-bisphosphatase Fbp1
SPBC11G11.01	fis1	mitochondrial fission protein Fis1 (predicted)
SPNCRNA.1514	fnx1-antisense-1	antisense RNA (predicted)
SPAC977.01	ftm1	sub-telomeric 5Tm protein family Ftm1
SPAC977.02	ftm2	sub-telomeric 5Tm protein family Ftm2
SPAC750.05c	ftm4	sub-telomeric 5Tm protein family Ftm4
SPBC1348.02	ftm5	sub-telomeric 5Tm protein family Ftm5
SPBC1348.03	ftm6	sub-telomeric 5Tm protein family Ftm6
SPBP4G3.03	fub2	PI31 proteasome regulator Fub2 (predicted)
SPCC18.18c	fum1	fumarate hydratase (predicted)
SPBPB2B2.13	gal1	galactokinase Gal1
SPNCRNA.1697	gal1-antisense-1	antisense RNA (predicted)
SPBPB2B2.12c	gal10	UDP-glucose 4-epimerase/aldose 1-epimerase Gal10
SPBPB2B2.10c	gal7	galactose-1-phosphate uridylyltransferase Gal7
SPCC794.01c	gcd1	glucose dehydrogenase Gcd1
SPNCRNA.875	gcv1-antisense-1	antisense RNA (predicted)
SPAC186.05c	gdt1	Golgi Ca(2+)/H(+) antiporter Gdt1
SPCC1672.12c	get4	GET complex (ER membrane insertion) subunit Get4 (predicted)
SPCC548.07c	ght1	hexose transmembrane transporter Ght1
SPAC1F8.01	ght3	hexose transmembrane transporter Ght3
SPBC1683.08	ght4	hexose transmembrane transporter Ght4

Systematic ID	Name	Product description
SPBC1348.14c	ght7	hexose transmembrane transporter Ght7 (predicted)
SPAC343.04c	gid7	GID complex subunit Gid7 (predicted)
SPAC17A5.09c	glc8	protein phosphatase regulatory subunit Glc8 (predicted)
SPAC13F5.03c	gld1	mitochondrial glycerol dehydrogenase Gld1
SPAC824.07	glo2	glyoxalase II
SPAC17H9.11	gmf1	cofilin/tropomyosin family Glia Maturation Factor homolog Gmf1
SPBC660.16	gnd1	phosphogluconate dehydrogenase, decarboxylating
SPBPB8B6.04c	grt1	transcription factor Grt1 (predicted)
SPNCRNA.1301	grt1-antisense-1	antisense RNA (predicted)
SPAC1039.11c	gto1	alpha-glucosidase (predicted)
SPBC2F12.14c	gua1	IMP dehydrogenase Gua1
SPBC1734.08	hse1	STAM like protein Hse1
SPNCRNA.1100	hsp3101-antisense-1	antisense RNA (predicted)
SPNCRNA.613	hsp3102-antisense-1	antisense RNA (predicted)
SPAC24H6.04	hxk1	hexokinase 1
SPAC11G7.03	idh1	isocitrate dehydrogenase (NAD+) subunit 1 Idh1
SPBC902.05c	idh2	isocitrate dehydrogenase (NAD+) subunit 2
SPBC3B9.17	isa2	mitochondrial iron-sulfur cluster assembly protein Isa2
SPBC365.12c	ish1	nuclear envelope LEA domain protein Ish1
SPNCRNA.606	isp3-antisense-1	antisense RNA (predicted)
SPAC4A8.04	isp6	vacuolar serine protease Isp6
SPAC20G8.03	itr2	MFS myo-inositol transmembrane transporter
SPAC144.08	jac1	mitochondrial DNAJ domain protein, Co-chaperone Hsc20 Jac1

Systematic ID	Name	Product description
SPBC776.15c	kgd2	dihydrolipoamide S-succinyltransferase, e2 component of oxoglutarate dehydrogenase complex Kdg2 (predicted)
SPBC2F12.13	klp5	kinesin-8 family plus-end microtubule motor Klp5
SPBC119.05c	lsb1	Wiskott-Aldrich syndrome homolog binding protein Lsb1 (predicted)
SPBC16E9.16c	lsd90	Lsd90 protein
SPCC1795.01c	mad3	mitotic spindle checkpoint protein Mad3
SPBC1683.07	mal1	maltase alpha-glucosidase Mal1
SPAC11E3.06	map1	MADS-box transcription factor Map1
SPBC23G7.09	mat1-Mc	mating-type m-specific polypeptide Mc
SPBC1711.02	mat3-Mc	mating type M-specific HMG-box transcription factor Mc at silenced MAT3 locus
SPMTR.04	mat3-Mc	silenced mating-type m-specific polypeptide Mc
SPAC15A10.10	mde6	Muskelin homolog, kelch repeat, expressed during meiotic cell cycle (predicted)
SPCC306.08c	mdh1	malate dehydrogenase Mdh1 (predicted)
SPNCRNA.1467	mdl1-antisense-1	antisense RNA (predicted), possible alternative UTR
SPAC27D7.03c	mei2	RNA-binding protein involved in meiosis Mei2
SPAC869.07c	mel1	alpha-galactosidase, melibiase
SPCC1223.12c	meu10	GPI anchored cell surface protein involved in ascospore wall assembly Meu10
SPBC1347.03	meu14	sporulation specific PIL domain protein Meu14
SPBC27.03	meu25	Schizosaccharomyces specific protein Meu25
SPAC25H1.05	meu29	calcium transport regulatory factor (predicted)
SPNCRNA.868	mfc1-antisense-1	antisense RNA (predicted)
SPCC338.05c	mms2	ubiquitin conjugating enzyme Mms2
SPBC13G1.02	mpg2	mannose-1-phosphate guanyltransferase (predicted)

Systematic ID	Name	Product description
SPBC2G2.16	mpi1	mannose-6-phosphate isomerase Mpi1 (predicted)
SPBC530.09c	mrl1	cation dependent mannose-6-phosphate cargo receptor Mrl1 (predicted)
SPAC12G12.08	mrpl1602	mitochondrial ribosomal protein subunit L16 (predicted)
SPAC644.17c	mrpl9	mitochondrial ribosomal protein subunit L9 (predicted)
SPNCRNA.700	mss1-antisense-1	antisense RNA (predicted)
SPAC29A4.12c	mug108	Schizosaccharomyces specific protein Mug108
SPBC19C2.06c	mug124	Schizosaccharomyces pombe specific protein
SPBC359.06	mug14	adducin, involved in actin cytoskeleton organization
SPBC56F2.06	mug147	Schizosaccharomyces specific protein Mug147
SPAC3C7.05c	mug191	alpha-1,6- mannanase (predicted)
SPBC106.08c	mug2	mug2/mug135/meu2 family
SPAC343.07	mug28	RNA-binding protein Mug28
SPCC4G3.05c	mus81	Holliday junction resolvase subunit Mus81
SPBC947.15c	nde1	mitochondrial NADH dehydrogenase (ubiquinone) Nde1 (predicted)
SPAC806.07	ndk1	nucleoside diphosphate kinase Ndk1
SPAC1952.13	ned1	lipin, phosphatidate phosphatase Ned1
SPBC12D12.08c	ned8	ubiquitin-like protein modifier for cullin Ned8
SPNCRNA.1450	ngg1-antisense-1	antisense RNA (predicted)
SPBC651.02	nit1	bis(5'-adenosyl)-triphosphatase Nit1 (predicted)
SPBC16A3.07c	nrm1	MBF complex corepressor Nrm1
SPBC215.10	odr1	HAD superfamily hydrolase, unknown role
SPAP14E8.04	oma1	metallopeptidase Oma1 (predicted)
SPNCRNA.130	omt3	intergenic RNA Omt3
SPAC22F8.06	pam1	20S proteasome complex subunit beta 6 Pam1
SPBPB2B2.09c	pan5	2-dehydropantoate 2-reductase Pan5 (predicted)

Systematic ID	Name	Product description
SPAC869.08	pcm2	protein-L-isoaspartate O-methyltransferase Pcm2 (predicted)
SPBC16D10.09	pcn1	PCNA
SPAC186.09	pdc102	pyruvate decarboxylase (predicted)
SPAC13A11.06	pdc202	pyruvate decarboxylase (predicted)
SPNCRNA.1386	pdf1-antisense-1	antisense RNA (predicted)
SPAC1F5.02	pdi1	protein disulfide isomerase (predicted)
SPBC359.04c	pfl7	cell surface glycoprotein, flocculin Pfl7, DIPSY family
SPAC1F8.06	pfl8	cell surface glycoprotein, flocculin Pfl8
SPAC186.01	pfl9	cell surface glycoprotein, flocculin Pfl9, DIPSY family
SPAC4D7.02c	pgc1	phosphatidylglycerol phospholipase C Pgc1 (predicted)
SPCC1322.16	phb2	prohibitin Phb2 (predicted)
SPCC16C4.03	pin1	peptidyl-prolyl cis-trans isomerase Pin1
SPAC23C11.16	plo1	Polo kinase Plo1
SPBC776.18c	pmh1	transcription factor TFIIK complex complex ubiquitin-protein ligase E3 subunit, Pmh1
SPNCRNA.1628	ppm1-antisense-1	antisense RNA (predicted)
SPCC1795.04c	pre10	20S proteasome complex subunit alpha 7, Pre10
SPBC4C3.10c	pre3	20S proteasome complex subunit beta 1 Pre3
SPBC577.10	pre4	20S proteasome complex subunit beta 7, Pre4
SPCC1442.06	pre8	20S proteasome complex subunit alpha 2, Pre8
SPAC13C5.01c	pre9	20S proteasome complex subunit alpha 3 Pre9
SPNCRNA.19	prl19	non-coding RNA, poly(A)-bearing RNA (predicted)
SPNCRNA.03	prl3	non-coding RNA, poly(A)-bearing RNA (predicted)
SPNCRNA.51	prl51	non-coding RNA, poly(A)-bearing (predicted)
SPNCRNA.724	prp22-antisense-1	antisense RNA (predicted)
SPAC1006.01	psp3	vacuolar serine protease Psp3 (predicted)

Systematic ID	Name	Product description
SPAC23D3.07	pup1	20S proteasome complex subunit beta 2 Pup1
SPCC63.12c	pup3	20S proteasome complex subunit beta 3, Pup3
SPNCRNA.1091	put4-antisense-1	antisense RNA (predicted)
SPBC4F6.12	pxl1	paxillin-like protein Pxl1
SPBC13E7.11	rbd1	mitochondrial rhomboid protease (predicted)
SPNCRNA.1492	rct1-antisense-1	antisense RNA (predicted), possible alternative UTR
SPAC6F12.06	rdi1	Rho GDP dissociation inhibitor Rdi1 (predicted)
SPAC4H3.07c	rdl2	mitochondrial outer membrane, thiosulfate sulfurtransferase/rhodanise like domain Rdl2, human TSTD1 ortholog (predicted)
SPAC25G10.04c	rec10	meiotic recombination protein Rec10
SPNCRNA.760	rec12-antisense-1	antisense RNA (predicted)
SPBC1711.14	rec15	meiotic recombination protein Rec15
SPBC2F12.11c	rep2	MBF transcription factor activator Rep2
SPAC22F3.12c	rgs1	regulator of G-protein signaling Rgs1
SPAC20H4.11c	rho5	Rho family GTPase Rho5
SPAC926.03	rlc1	myosin II regulatory light chain Rlc1
SPBP8B7.23	rnf10	ubiquitin-protein ligase E3 (predicted)
SPAC31G5.02	rot1	ER chaperone Rot1 (predicted)
SPAC19G12.07c	rsd1	RNA-binding protein Rsd1 (predicted)
SPBC119.14	rti1	Rad22 homolog Rti1
SPNCRNA.1680	rtt109-antisense-1	antisense RNA (predicted), possible alternative UTR
SPAC17C9.06	sam50	mitochondrial sorting and assembly machinery complex subunit Sam50 (predicted)
SPBPB2B2.02	say1	ER sterol deacetylase Say1 (predicted)
SPBC646.16	scl1	20S proteasome complex subunit alpha 1
SPAC1556.02c	sdh1	succinate dehydrogenase Sdh1 (predicted)

Systematic ID	Name	Product description
SPAC140.01	sdh2	succinate dehydrogenase (ubiquinone) iron-sulfur protein subunit (predicted)
SPAC22E12.03c	sdj1	glyoxylase III sdj1
SPNCRNA.1082	sec16-antisense-1	antisense RNA (predicted)
SPAC959.02	sec17	alpha SNAP (predicted)
SPAC1834.11c	sec18	secretory pathway protein Sec18 (predicted)
SPNCRNA.1586	sec2302-antisense-1	antisense RNA (predicted)
SPAC24B11.11c	sid2	NDR kinase Sid2
SPNCRNA.637	sif3-antisense-1	antisense RNA (predicted)
SPAC1486.01	sod2	mitochondrial superoxide dismutase Sod2
SPNCRNA.611	SPAC11D3.18c-antisense-1	antisense RNA (predicted), possible alternative UTR
SPNCRNA.932	SPAC15E1.10-antisense-1	antisense RNA (predicted)
SPNCRNA.788	SPAC17H9.12c-antisense-1	antisense RNA (predicted)
SPNCRNA.1033	SPAC19B12.01-antisense-1	antisense RNA (predicted)
SPNCRNA.1037	SPAC1B3.04c-antisense-2	antisense RNA (predicted)
SPNCRNA.981	SPAC23D3.17-antisense-1	antisense RNA (predicted)
SPNCRNA.832	SPAC23H3.15c-antisense-1	antisense RNA (predicted), possible alternative UTR
SPNCRNA.955	SPAC27E2.02-antisense-1	antisense RNA (predicted)

Systematic ID	Name	Product description
SPNCRNA.612	SPAC5H10.01-antisense-1	antisense RNA (predicted), possible alternative UTR
SPNCRNA.905	SPAC8E11.01c-antisense-1	antisense RNA (predicted)
SPNCRNA.987	SPAC9E9.17c-antisense-1	antisense RNA (predicted)
SPNCRNA.1403	SPBC146.02-antisense-1	antisense RNA (predicted)
SPNCRNA.1573	SPBC15D4.05-antisense-1	antisense RNA (predicted)
SPNCRNA.1574	SPBC15D4.09c-antisense-1	antisense RNA (predicted)
SPNCRNA.1647	SPBC1604.09c-antisense-1	antisense RNA (predicted)
SPNCRNA.1338	SPBC1773.12-antisense-1	antisense RNA (predicted)
SPNCRNA.1601	SPBC2G2.17c-antisense-1	antisense RNA (predicted), possible alternative UTR
SPNCRNA.1548	SPBC336.13c-antisense-1	antisense RNA (predicted)
SPNCRNA.1667	SPBC56F2.07c-antisense-1	antisense RNA (predicted)
SPNCRNA.1395	SPBC713.13-antisense-1	antisense RNA (predicted)
SPNCRNA.1695	SPBP4G3.03-antisense-1	antisense RNA (predicted)
SPNCRNA.1307	SPBPB10D8.02c-antisense-1	antisense RNA (predicted)

Systematic ID	Name	Product description
SPNCRNA.1160	SPCC1393.09c-antisense-1	antisense RNA (predicted)
SPNCRNA.1143	SPCC1672.03c-antisense-1	antisense RNA (predicted)
SPNCRNA.1294	SPCC1827.07c-antisense-1	antisense RNA (predicted)
SPNCRNA.1176	SPCC4B3.06c-antisense-1	antisense RNA (predicted)
SPNCRNA.1298	SPCC569.02c-antisense-1	antisense RNA (predicted)
SPNCRNA.1101	SPCC757.13-antisense-1	antisense RNA (predicted)
SPAC144.04c	spe1	ornithine decarboxylase Spe1 (predicted)
SPBC1289.03c	spi1	Ran GTPase Spi1
SPCC188.12	spn6	meiotic (sporulation) septin Spn6
SPBC19F8.01c	spn7	meiotic septin Spn7
SPAC3F10.15c	spo12	Spo12 family nuclear protein
SPBC19G7.04	spr1	DNA-protein crosslink removal protease (predicted)
SPNCRNA.1078	spt20-antisense-1	antisense RNA (predicted), possible alternative UTR
SPNCRNA.1517	srm1-antisense-1	antisense RNA (predicted), possible alternative UTR
SPBC660.13c	ssb1	DNA replication factor A subunit Ssb1
SPCC74.03c	ssp2	AMP-activated protein serine/threonine kinase alpha subunit Ssp2
SPBC32C12.02	ste11	transcription factor Ste11
SPAC24B11.06c	sty1	MAP kinase Sty1
SPAC16E8.16	sua7	transcription factor TFIIB
SPBC25D12.04	suc22	ribonucleotide reductase small subunit Suc22

Systematic ID	Name	Product description
SPBC1711.18	tam9	mitochondrial ribosomal protein subunit L36, MrpL36/YmL36 (predicted)
SPCP31B10.06	tcb2	tricalbin, C2 domain protein (phospholipid binding) ER-plasma membrane tethering protein Tcb2 (predicted)
SPBPB2B2.11	tgd1	nucleotide-sugar 4,6-dehydratase (predicted)
SPBP23A10.16	tim18	TIM22 inner membrane protein import complex anchor subunit Tim18
SPAC139.03	toe2	transcription factor, zf-fungal binuclear cluster type (predicted)
SPCC1753.04	tol1	3'(2'),5'-bisphosphate nucleotidase/inositol-1,4- bisphosphate 1-phosphatase
SPCC24B10.21	tpi1	triosephosphate isomerase
SPAP8A3.12c	tpp2	tripeptidyl-peptidase II Tpp2
SPNCRNA.643	trm112-antisense-1	antisense RNA (predicted)
SPNCRNA.952	trm5-antisense-1	antisense RNA (predicted), possible alternative UTR
SPBC800.07c	tsf1	mitochondrial translation elongation factor EF-Ts Tsf1
SPBC9B6.04c	tuf1	mitochondrial translation elongation factor EF-Tu Tuf1
SPCC4B3.01	tum1	thiosulfate sulfurtransferase, involved in tRNA wobble position thiolation Tum1 (predicted)
SPCC126.06	twf1	twinfilin (predicted)
SPCC1259.15c	ubc11	ubiquitin conjugating enzyme E2, Ubc11/UbcP4
SPAC11E3.04c	ubc13	ubiquitin conjugating enzyme E2 Ubc13
SPBC1105.09	ubc15	ubiquitin conjugating enzyme E2 Ubc15
SPBP16F5.04	ubc7	Hrd1 ubiquitin ligase complex ubiquitin conjugating enzyme E2 Ubc7
SPBC211.07c	ubc8	ubiquitin conjugating enzyme E2 Ubc8 (predicted)
SPNCRNA.1137	ubp16-antisense-1	antisense RNA (predicted)

Systematic ID	Name	Product description
SPCC188.08c	ubp5	ubiquitin C-terminal hydrolase Ubp5
SPAC13A11.04c	ubp8	SAGA complex ubiquitin C-terminal hydrolase Ubp8
SPBC27.04	uds1	septation protein Uds1
SPAC644.08	utr4	methionine salvage haloacid dehalogenase-like hydrolase Utr4 (predicted)
SPNCRNA.1687	vps10-antisense-1	antisense RNA (predicted)
SPNCRNA.1640	vps13a-antisense-1	antisense RNA (predicted)
SPBC4B4.06	vps25	ESCRT II complex subunit Vps25
SPBC32H8.09	wdr8	mitosis-specific spindle pole body WD repeat protein Wdr8
SPNCRNA.1036	wis2-antisense-1	antisense RNA (predicted)
SPCC1281.08	wtf11	wtf element Wtf11
SPCC553.05c	wtf6	wtf element Wtf6
SPAC869.02c	yhb1	nitric oxide dioxygenase Yhb1
SPBP35G2.12	ysa1	ADP-ribose diphosphatase, NudF subfamily Ysa1 (predicted)
SPBC543.09	yta12	mitochondrial m-AAA protease Yta12 (predicted)
SPAC22H10.13	zym1	metallothionein Zym1
SPAC1002.07c	spermidine/spermine N1-acetyltransferase (predicted)	
SPAC1071.13	-	dubious
SPAC11E3.12	-	mitochondrial thioredoxin family protein
SPAC11E3.14	-	protein with a role in clearing protein aggregates (predicted)
SPAC12G12.07c	-	conserved fungal protein

Systematic ID	Name	Product description
SPAC144.17c	-	6-phosphofructo-2-kinase (predicted)
SPAC14C4.04	-	ThiJ domain protein (predicted)
SPAC1635.01	-	mitochondrial outer membrane voltage-dependent anion-selective channel (predicted)
SPAC1687.21	-	fructose-2,6-bisphosphatase, human TIGAR ortholog (predicted)
SPAC1705.02	-	SERF family protein, DUF, human 4F5S homolog
SPAC186.04c	-	N-terminal of transmembrane channel, truncated
SPAC186.08c	-	L-lactate dehydrogenase (predicted)
SPAC1952.09c	-	acetyl-CoA hydrolase (predicted)
SPAC1B3.06c	-	UbiE family methyltransferase (predicted)
SPAC1B3.20	-	Schizosaccharomyces specific protein
SPAC1F8.08	-	Schizosaccharomyces pombe specific protein
SPAC212.01c	-	S. pombe specific DUF999 family protein 2
SPAC212.02	-	Schizosaccharomyces pombe specific protein
SPAC212.04c	-	S. pombe specific DUF999 family protein 1
SPAC212.06c	-	DNA helicase in rearranged telomeric region, truncated
SPAC212.08c	-	S. pombe specific GPI anchored protein family 1
SPAC212.12	-	S. pombe specific GPI anchored protein family 1
SPAC222.17	-	conserved fungal protein
SPAC22H12.03	-	mitochondrial hydrolase (predicted)
SPAC23A1.14c	-	pyridoxal phosphate-dependent transferase, unknown specificity (predicted)
SPAC23D3.17	-	protease inhibitor 178 family
SPAC23H3.15c	-	Schizosaccharomyces specific protein
SPAC24B11.05	-	pyrimidine 5'-nucleotidase (predicted)
SPAC24H6.08	-	Schizosaccharomyces specific protein

Systematic ID	Name	Product description
SPAC29A4.17c	-	mitochondrial FUN14 family protein involved in mitophagy
SPAC29B12.13	-	S-(hydroxymethyl)glutathione synthase activity (predicted)
SPAC30C2.08	-	UPF0662 family conserved fungal protein
SPAC31G5.21	-	DUF1754 family, human FAM32A homolog, implicated in splicing
SPAC3G6.03c	-	Maf-like protein, human ASMTL ortholog
SPAC4A8.14	-	ribose-phosphate pyrophosphokinase (predicted)
SPAC4H3.04c	-	MEMO1 family, human MEMO1 ortholog, ancient conserved protein, in some organisms this domain is present with ammecr1 which interacts with mediator complex and is implicated in transcription, implicated in signalling and localization to the plasma membrane
SPAC4H3.08	-	3-hydroxyacyl-CoA dehydrogenase (predicted)
SPAC56E4.07	-	N-acetyltransferase (predicted)
SPAC637.03	-	DUF1774 family multi-spanning conserved fungal membrane protein
SPAC644.13c	-	Rab GTPase binding (predicted)
SPAC688.03c	-	human AMMECR1 homolog
SPAC6F6.11c	-	pyridoxine-pyridoxal-pyridoxamine kinase (predicted)
SPAC750.02c	-	transmembrane transporter (predicted)
SPAC750.06c	-	S. pombe specific DUF999 protein family 4
SPAC750.07c	-	S. pombe specific protein
SPAC750.08c	-	NAD-dependent malic enzyme (predicted), partial
SPAC806.05	-	mitochondrial ANC9 family protein
SPAC869.01	-	hydrolase activity (predicted)

Systematic ID	Name	Product description
SPAC869.03c	-	urea transmembrane transporter (predicted)
SPAC8E11.04c	-	palmitoyl-(protein) hydrolase (predicted)
SPAC977.04	-	truncated C terminal region of membrane transporter
SPAC977.05c	-	Velum formation protein 1-like conserved fungal family
SPAC977.09c	-	phospholipase (predicted)
SPAC977.17	-	MIP water channel (predicted)
SPAC977.18	-	conserved fungal plasma membrane protein
SPAC9E9.06c	-	threonine synthase (predicted)
SPACUNK4.17	-	NAD binding dehydrogenase family protein, human DHDH ortholog, unknown biological role, implicated in carbohydrate metabolic process
SPAPB24D3.07c	-	Schizosaccharomyces pombe specific protein
SPAPJ691.02	-	yippee-like protein
SPAPJ695.02	-	Schizosaccharomyces pombe specific protein
SPBC1198.03c	-	DUF4646 family conserved fungal protein
SPBC1198.05	-	guanylate kinase (predicted)
SPBC11C11.06c	-	Schizosaccharomyces specific protein
SPBC1348.01	-	S. pombe specific DUF999 protein family 5
SPBC1348.05	-	transmembrane transporter (predicted)
SPBC1348.06c	-	Velum formation protein 1-like conserved fungal family
SPBC1348.07	-	S. pombe specific DUF999 protein family 6
SPBC1348.10c	-	phospholipase (predicted)
SPBC1348.12	-	transcription factor (predicted)
SPBC1652.01	-	transcription factor (predicted)
SPBC16A3.14	-	superoxide dismutase, mitochondrial ribosomal protein subunit (predicted)

Systematic ID	Name	Product description
SPBC16G5.07c	-	stomatin, mitochondrial membrane organization protein (predicted)
SPBC16H5.15	-	DUF3807 family conserved fungal protein, domain associated with Mpn1 in some fungi
SPBC1773.12	-	transcription factor, zf-fungal binuclear cluster type (predicted)
SPBC17G9.06c	-	siderophore-iron biosynthesis protein (predicted)
SPBC216.03	-	NADP binding superfamily conserved fungal protein
SPBC21B10.08c	-	antibiotic biosynthesis monooxygenase-like domain (predicted)
SPBC21D10.08c	-	conserved fungal protein
SPBC26H8.11c	-	acyl-coenzyme A thioesterase
SPBC32F12.10	-	phosphoglucomutase (predicted)
SPBC337.02c	-	mug2/mug135/meu2 family
SPBC428.10	-	Schizosaccharomyces pombe specific protein
SPBC460.01c	-	amino-acid permease, unknown
SPBC460.04c	-	sulfonate/alpha-ketoglutarate dioxygenase (predicted)
SPBC557.02c	-	DUF2458 conserved fungal protein
SPBC725.03	-	pyridoxamine 5'-phosphate oxidase (predicted)
SPBC725.10	-	mitochondrial transport protein, tspO homolog (predicted)
SPBC8E4.04	-	alditol NADP+ 1-oxidoreductase activity (predicted)
SPBCPT2R1.01c	-	S. pombe specific DUF999 protein family 9
SPBCPT2R1.02	-	Schizosaccharomyces pombe specific protein
SPBCPT2R1.04c	-	S. pombe specific DUF999 protein family 10
SPBCPT2R1.06c	-	
SPBPB21E7.04c	-	O-methyltransferase, human COMT catechol homolog 2

Systematic ID	Name	Product description
SPBPB21E7.08	-	
SPBPB21E7.11	-	Schizosaccharomyces pombe specific protein
SPBPB2B2.01	-	amino acid transmembrane transporter (predicted)
SPBPB2B2.05	-	class I glutamine amidotransferase family protein
SPBPB2B2.06c	-	extracellular 5'-nucleotidase, human NT5E family (predicted)
SPBPB2B2.15	-	Velum formation protein 1-like conserved fungal family
SPBPB2B2.16c	-	transmembrane transporter (predicted)
SPBPB2B2.18	-	conserved fungal plasma membrane protein
SPBPB2B2.19c	-	S. pombe specific 5Tm protein family
SPBPB8B6.02c	-	urea transmembrane transporter (predicted)
SPCC1235.17	-	dubious
SPCC132.03	-	Schizosaccharomyces specific protein
SPCC1322.10	-	cell wall protein, Kre9/Knh1 family Pwp1
SPCC1393.12	-	Schizosaccharomyces specific protein
SPCC1442.04c	-	meiotic recombination protein (predicted)
SPCC1450.15	-	pig-F/3-ketosphinganine reductase fusion protein (predicted)
SPCC16A11.15c	-	Schizosaccharomyces specific protein
SPCC1739.08c	-	short chain dehydrogenase (predicted)
SPCC191.01	-	Schizosaccharomyces specific protein
SPCC24B10.02c	-	NAD/NADH kinase (predicted)
SPCC285.04	-	transthyretin/hydroxyisourate hydrolase (predicted)
SPCC285.10c	-	SPRY domain protein
SPCC569.01c	-	mug2/mug135/meu2 family
SPCC569.02c	-	S. pombe specific UPF0321 family protein 2
SPCC569.09	-	Schizosaccharomyces specific protein

Systematic ID	Name	Product description
SPCC576.06c	-	mitochondrial tyrosine-tRNA ligase (predicted)
SPCC584.15c	-	arrestin involved in ubiquitin-dependent endocytosis
SPCC61.05	-	Schizosaccharomyces specific multicopy membrane protein family 1
SPCC663.15c	-	DUF3818 and PXA domain conserved fungal protein
SPCC736.13	-	short chain dehydrogenase (predicted)
SPCC777.06c	-	hydrolase, conserved in fungi, bacteria, plants, protazoa (predicted)
SPCC794.04c	-	amino acid transmembrane transporter (predicted)
SPCC965.14c	-	cytosine deaminase (predicted)
SPCP20C8.01c	-	mug2/mug135/meu2 family
SPCP20C8.02c	-	S. pombe specific UPF0321 family protein 1
SPCP20C8.03	-	
SPNCRNA.1027	-	antisense RNA (predicted)
SPNCRNA.1092	-	intergenic RNA (predicted)
SPNCRNA.1116	-	intergenic RNA (predicted)
SPNCRNA.1145	-	antisense RNA (predicted)
SPNCRNA.1212	-	antisense RNA (predicted)
SPNCRNA.1218	-	intergenic RNA (predicted), possible alternative UTR
SPNCRNA.1259	-	intergenic RNA (predicted), possible alternative UTR
SPNCRNA.1269	-	intergenic RNA (predicted)
SPNCRNA.1290	-	intergenic RNA (predicted)
SPNCRNA.1304	-	intergenic RNA (predicted)
SPNCRNA.1390	-	intergenic RNA (predicted)
SPNCRNA.1532	-	intergenic RNA (predicted)
SPNCRNA.1604	-	intergenic RNA (predicted)
SPNCRNA.1605	-	intergenic RNA (predicted)

Systematic ID	Name	Product description
SPNCRNA.1625	-	intergenic RNA (predicted)
SPNCRNA.1694	-	intergenic RNA (predicted)
SPNCRNA.249	-	non-coding RNA (predicted)
SPNCRNA.289	-	non-coding RNA (predicted)
SPNCRNA.293	-	intergenic RNA (predicted)
SPNCRNA.329	-	non-coding RNA (predicted)
SPNCRNA.332	-	non-coding RNA (predicted)
SPNCRNA.389	-	non-coding RNA (predicted)
SPNCRNA.601	-	intergenic RNA (predicted)
SPNCRNA.623	-	intergenic RNA (predicted), possible alternative UTR
SPNCRNA.626	-	intergenic RNA (predicted), possible alternative UTR
SPNCRNA.667	-	intergenic RNA (predicted)
SPNCRNA.671	-	intergenic RNA (predicted)
SPNCRNA.672	-	intergenic RNA (predicted)
SPNCRNA.679	-	intergenic RNA (predicted)
SPNCRNA.767	-	intergenic RNA (predicted)
SPNCRNA.779	-	intergenic RNA (predicted)
SPNCRNA.780	-	intergenic RNA (predicted)
SPNCRNA.853	-	antisense RNA (predicted)
SPNCRNA.871	-	antisense RNA (predicted)
SPNCRNA.965	-	intergenic RNA (predicted)
SPNCRNA.992	-	antisense RNA (predicted), possible alternative UTR

A.1.2 Gene Ontology enrichment

External ID	GeneSet Name	over represented/under represented	Corrected p-value
GO:0010467;	gene expression	Underrepresented	3.5011199999999999E-12
GO:0043170;	macromolecule metabolic process	Underrepresented	1.2029799999999999E-9
GO:0044260;	cellular macromolecule metabolic process	Underrepresented	1.2029799999999999E-9
GO:0002181;	cytoplasmic translation	Underrepresented	1.16433E-8
GO:0006396;	RNA processing	Underrepresented	9.7909599999999999E-8
GO:0044712;	single-organism catabolic process	Enriched	5.6858499999999995E-7
GO:0090304;	nucleic acid metabolic process	Underrepresented	2.9180900000000002E-6
GO:0016070;	RNA metabolic process	Underrepresented	6.8910599999999997E-6
GO:0009059;	macromolecule biosynthetic process	Underrepresented	8.4887700000000004E-6
GO:0034645;	cellular macromolecule biosynthetic process	Underrepresented	1.0957100000000001E-5
GO:0034660;	ncRNA metabolic process	Underrepresented	1.2182500000000001E-5
GO:0006412;	translation	Underrepresented	2.4672200000000001E-5
GO:0016052;	carbohydrate catabolic process	Enriched	2.4878599999999999E-5
GO:0044724;	single-organism carbohydrate catabolic process	Enriched	1.115E-4
GO:0019318;	hexose metabolic process	Enriched	3.46492E-4
GO:0006099;	tricarboxylic acid cycle	Enriched	3.5021299999999999E-4
GO:0005996;	monosaccharide metabolic process	Enriched	4.5268499999999998E-4
GO:0044249;	cellular biosynthetic process	Underrepresented	4.7469199999999997E-4
GO:0006091;	generation of precursor metabolites and energy	Enriched	8.3168499999999998E-4
GO:0006012;	galactose metabolic process	Enriched	8.7093499999999998E-4

External ID	GeneSet Name	over represented/under represented	Corrected p-value
GO:0006101;	citrate metabolic process	Enriched	9.1333799999999998E-4
GO:1901576;	organic substance biosynthetic process	Underrepresented	9.8093700000000008E-4
GO:0009056;	catabolic process	Enriched	1.1163E-3
GO:0009058;	biosynthetic process	Underrepresented	1.30657E-3
GO:0072350;	tricarboxylic acid metabolic process	Enriched	1.30657E-3
GO:1901575;	organic substance catabolic process	Enriched	1.3821E-3
GO:0055086;	nucleobase-containing small molecule metabolic process	Enriched	2.3265600000000001E-3
GO:0044281;	small molecule metabolic process	Enriched	3.7626700000000001E-3
GO:0032787;	monocarboxylic acid metabolic process	Enriched	4.0433800000000001E-3
GO:0044238;	primary metabolic process	Underrepresented	4.3848699999999999E-3
GO:0006753;	nucleoside phosphate metabolic process	Enriched	9.3269600000000005E-3
GO:0009060;	aerobic respiration	Enriched	9.3269600000000005E-3

A.2 Positive non-scaling proteins

A.2.1 Gene list

Systematic ID	Name	Product description
SPAC144.03	ade2	adenylosuccinate synthetase Ade2
SPCC1322.13	ade6	phosphoribosylaminoimidazole carboxylase Ade6
SPBPB21E7.07	aes1	phenazine biosynthesis PhzF protein family

Systematic ID	Name	Product description
SPAC8E11.07c	alp31	tubulin specific chaperone cofactor A, Alp31
SPAC26A3.05	chc1	clathrin heavy chain Chc1 (predicted)
SPCC757.07c	ctt1	catalase
SPAC22A12.11	dak1	dihydroxyacetone kinase Dak1
SPAC513.06c	dhd1	D-xylose 1-dehydrogenase (NADP+) (predicted)
SPBC713.12	erg1	squalene monooxygenase Erg1 (predicted)
SPCC1322.04	fyu1	UTP-glucose-1-phosphate uridylyltransferase Fyu1
SPBPB2B2.13	gal1	galactokinase Gal1
SPBC29A10.08	gas2	1,3-beta-glucanosyltransferase Gas2
SPCC794.01c	gcd1	glucose dehydrogenase Gcd1
SPAPB1E7.05	gde1	glycerophosphoryl diester phosphodiesterase Gde1 (predicted)
SPCC1672.12c	get4	GET complex (ER membrane insertion) subunit Get4 (predicted)
SPBC660.16	gnd1	phosphogluconate dehydrogenase, decarboxylating
SPBC354.12	gpd3	glyceraldehyde 3-phosphate dehydrogenase Gpd3
SPAP7G5.02c	gua2	GMP synthase [glutamine-hydrolyzing] Gua2 (predicted)
SPAC4F8.07c	hxk2	hexokinase 2
SPBC56F2.12	ilv5	acetohydroxyacid reductoisomerase (predicted)
SPBC3H7.03c	kgd1	2-oxoglutarate dehydrogenase (lipoamide) (e1 component of oxoglutarate dehydrogenase complex) (predicted)
SPBC16E9.16c	lsd90	Lsd90 protein
SPAC31G5.04	lys12	homoisocitrate dehydrogenase Lys12
SPAC29E6.05c	mxr1	peptide-methionine (S)-S-oxide reductase MsrA
SPBC2G2.11	myr1	N-myristoyltransferase Myr1 (predicted)
SPAC22F8.06	pam1	20S proteasome complex subunit beta 6 Pam1
SPAC1F8.07c	pdc101	pyruvate decarboxylase (predicted)

Systematic ID	Name	Product description
SPBC16H5.02	pfk1	6-phosphofructokinase pfk1
SPAC4D7.02c	pgc1	phosphatidylglycerol phospholipase C Pgc1 (predicted)
SPBC14F5.04c	pgk1	phosphoglycerate kinase Pgk1 (predicted)
SPAC6F12.06	rdi1	Rho GDP dissociation inhibitor Rdi1 (predicted)
SPAC1556.02c	sdh1	succinate dehydrogenase Sdh1 (predicted)
SPAC6G9.10c	sen1	Nrd1 complex ATP-dependent 5' to 3' DNA/RNA helicase Sen1
SPBC1289.03c	spi1	Ran GTPase Spi1
SPBC25D12.04	suc22	ribonucleotide reductase small subunit Suc22
SPCC24B10.21	tpi1	triosephosphate isomerase
SPAC19G12.15c	tpp1	trehalose-6-phosphate phosphatase Tpp1
SPAP8A3.12c	tpp2	tripeptidyl-peptidase II Tpp2
SPBP16F5.03c	tra1	SAGA complex phosphatidylinositol pseudokinase Tra1
SPAC23C4.17	trm402	tRNA (cytosine-5-)-methyltransferase (predicted)
SPCC4B3.01	tum1	thiosulfate sulfurtransferase, involved in tRNA wobble position thiolation Tum1 (predicted)
SPCC126.06	twf1	twinfilin (predicted)
SPAC644.08	utr4	methionine salvage haloacid dehalogenase-like hydrolase Utr4 (predicted)
SPAC767.01c	vps1	dynamin family protein Vps1
SPAC227.14	yfh7	uridine kinase Yfh7 (predicted)
SPAC869.02c	yhb1	nitric oxide dioxygenase Yhb1
SPBP35G2.12	ysa1	ADP-ribose diphosphatase, NudF subfamily Ysa1 (predicted)
SPAC14C4.04	-	ThiJ domain protein (predicted)
SPAC1B3.06c	-	UbiE family methyltransferase (predicted)

Systematic ID	Name	Product description
SPAC23A1.14c	-	pyridoxal phosphate-dependent transferase, unknown specificity (predicted)
SPAC977.14c	-	aldo/keto reductase, predicted calcium channel regulator
SPAC9E9.06c	-	threonine synthase (predicted)
SPBC106.03	-	mitochondrial Rossman fold DUF1776 family protein
SPBC2G5.05	-	transketolase (predicted)
SPBC32F12.10	-	phosphoglucomutase (predicted)
SPBC4.06	-	acid phosphatase Fmp10 (predicted)
SPBC800.14c	-	DUF1772 family protein, multimembrane spanning anthrone oxygenase-like
SPBC83.16c	-	protein with a role in clearing protein aggregates (predicted)
SPBPB2B2.06c	-	extracellular 5'-nucleotidase, human NT5E family (predicted)
SPCC1827.03c	-	acetyl-CoA ligase (predicted)
SPCC777.06c	-	hydrolase, conserved in fungi, bacteria, plants, protazoa (predicted)
SPCC965.14c	-	cytosine deaminase (predicted)

A.2.2 Gene Ontology Enrichment

External ID	GeneSet Name	over represented/under represented	Corrected p-value
GO:0044281;	small molecule metabolic process	Enriched	6.4330900000000002E-10

External ID	GeneSet Name	over represented/under represented	Corrected p-value
GO:0006753;	nucleoside phosphate metabolic process	Enriched	3.5630100000000001E-9
GO:0055086;	nucleobase-containing small molecule metabolic process	Enriched	3.5630100000000001E-9
GO:0019693;	ribose phosphate metabolic process	Enriched	9.4953500000000007E-9
GO:0009117;	nucleotide metabolic process	Enriched	1.6455199999999998E-8
GO:0019637;	organophosphate metabolic process	Enriched	4.9245399999999997E-8
GO:0009119;	ribonucleoside metabolic process	Enriched	2.4383900000000002E-7
GO:0009259;	ribonucleotide metabolic process	Enriched	2.4911900000000002E-7
GO:0044724;	single-organism carbohydrate catabolic process	Enriched	2.7324999999999999E-7
GO:0006793;	phosphorus metabolic process	Enriched	3.5206899999999997E-7
GO:0009116;	nucleoside metabolic process	Enriched	4.00344E-7
GO:0046365;	monosaccharide catabolic process	Enriched	5.0697600000000005E-7
GO:0044710;	single-organism metabolic process	Enriched	5.1006600000000001E-7
GO:1901657;	glycosyl compound metabolic process	Enriched	5.1006600000000001E-7
GO:0009161;	ribonucleoside monophosphate metabolic process	Enriched	5.5792599999999996E-7
GO:0006796;	phosphate-containing compound metabolic process	Enriched	5.6049000000000001E-7
GO:0016052;	carbohydrate catabolic process	Enriched	5.6565300000000002E-7
GO:0009123;	nucleoside monophosphate metabolic process	Enriched	7.6845199999999998E-7
GO:0019320;	hexose catabolic process	Enriched	9.51419E-7

External ID	GeneSet Name	over represented/under represented	Corrected p-value
GO:1901135;	carbohydrate derivative metabolic process	Enriched	1.0127399999999999E-6
GO:0019752;	carboxylic acid metabolic process	Enriched	2.09582E-6
GO:0005996;	monosaccharide metabolic process	Enriched	2.2091300000000001E-6
GO:0043436;	oxoacid metabolic process	Enriched	2.40361E-6
GO:0006082;	organic acid metabolic process	Enriched	2.4452400000000002E-6
GO:0042278;	purine nucleoside metabolic process	Enriched	3.8192699999999996E-6
GO:0072521;	purine-containing compound metabolic process	Enriched	4.2709000000000001E-6
GO:0019318;	hexose metabolic process	Enriched	9.0880100000000008E-6
GO:0009126;	purine nucleoside monophosphate metabolic process	Enriched	9.0965899999999996E-6
GO:0009167;	purine ribonucleoside monophosphate metabolic process	Enriched	9.0965899999999996E-6
GO:0006096;	glycolytic process	Enriched	9.3211400000000007E-6
GO:0032787;	monocarboxylic acid metabolic process	Enriched	1.6088800000000001E-5
GO:0006090;	pyruvate metabolic process	Enriched	2.2963400000000001E-5
GO:0006091;	generation of precursor metabolites and energy	Enriched	2.47012E-5
GO:0009150;	purine ribonucleotide metabolic process	Enriched	2.47012E-5
GO:0046128;	purine ribonucleoside metabolic process	Enriched	2.47012E-5

External ID	GeneSet Name	over represented/under represented	Corrected p-value
GO:1901564;	organonitrogen compound metabolic process	Enriched	2.6242999999999999E-5
GO:0006163;	purine nucleotide metabolic process	Enriched	3.9957899999999999E-5
GO:0044723;	single-organism carbohydrate metabolic process	Enriched	3.9957899999999999E-5
GO:0046496;	nicotinamide nucleotide metabolic process	Enriched	4.0686599999999998E-5
GO:0051156;	glucose 6-phosphate metabolic process	Enriched	4.0686599999999998E-5
GO:0061615;	glycolytic process through fructose-6-phosphate	Enriched	4.0686599999999998E-5
GO:0061620;	glycolytic process through glucose-6-phosphate	Enriched	4.0686599999999998E-5
GO:0061621;	canonical glycolysis	Enriched	4.0686599999999998E-5
GO:0044712;	single-organism catabolic process	Enriched	4.3216600000000002E-5
GO:0019362;	pyridine nucleotide metabolic process	Enriched	4.44208E-5
GO:0006735;	NADH regeneration	Enriched	5.7204399999999997E-5
GO:0005975;	carbohydrate metabolic process	Enriched	8.2492199999999996E-5
GO:0072524;	pyridine-containing compound metabolic process	Enriched	1.46036E-4
GO:0006007;	glucose catabolic process	Enriched	1.63356E-4
GO:0006733;	oxidoreduction coenzyme metabolic process	Enriched	2.21869E-4
GO:0006006;	glucose metabolic process	Enriched	2.252E-4

External ID	GeneSet Name	over represented/under represented	Corrected p-value
GO:0044260;	cellular macromolecule metabolic process	Underrepresented	4.2129100000000002E-4
GO:0043170;	macromolecule metabolic process	Underrepresented	8.5601999999999998E-4
GO:0006734;	NADH metabolic process	Enriched	1.1661099999999999E-3
GO:0006732;	coenzyme metabolic process	Enriched	5.0319800000000001E-3
GO:0090304;	nucleic acid metabolic process	Underrepresented	5.1366900000000002E-3
GO:0019674;	NAD metabolic process	Enriched	8.0772699999999992E-3

A.3 Negative non-scaling transcripts

A.3.1 Gene list

Systematic ID	Name	Product description
SPAC30.04c	abc4	glutathione S-conjugate-exporting ATPase Abc4
SPBC713.06	adl1	DNA ligase (predicted)
SPAC14C4.09	agn1	glucan endo-1,3-alpha-glucosidase Agn1
SPBC691.03c	apl3	AP-2 adaptor complex alpha subunit Alp3
SPBC3D6.10	apn2	AP-endonuclease Apn2
SPAC1250.04c	atl1	alkyltransferase-like protein Atl1
SPAC977.08	ayr2	1-acyl DHAP reductase Ayr2 (predicted)
SPAC24C9.07c	bgs2	spore wall 1,3-beta-glucan synthase catalytic subunit Bgs2
SPBC582.05c	brc1	BRCT domain protein Brc1
SPCC1919.15	brl1	ubiquitin-protein ligase E3 Brl1

Systematic ID	Name	Product description
SPBC29A3.01	ccc2	copper transporting ATPase Ccc2 (predicted)
SPAC23H4.09	cdb4	curved DNA-binding protein Cdb4, peptidase family
SPAC17D4.02	cdc45	DNA replication pre-initiation complex subunit Cdc45
SPBC1709.08	cft1	cleavage factor one Cft1 (predicted)
SPAC3G6.11	chl1	ATP-dependent DNA helicase Chl1 (predicted)
SPCC663.12	cid12	poly(A) polymerase Cid12
SPBC2D10.17	clr1	SHREC complex subunit Clr1
SPCC306.03c	cnd2	condensin complex non-SMC subunit Cnd2
SPCC338.13	cog4	Golgi transport complex subunit Cog4 (predicted)
SPBC342.05	crb2	DNA repair protein Rad9 homolog Crb2
SPNCRNA.762	crp79-antisense-1	antisense RNA (predicted)
SPNCRNA.1374	cta3-antisense-1	antisense RNA (predicted)
SPCC5E4.04	cut1	separase/separin
SPBC106.09	cut4	anaphase-promoting complex, platform subcomplex scaffold subunit Apc1
SPAC30D11.09	cwf19	complexed with Cdc5 protein Cwf19
SPBC23E6.01c	cxr1	mRNA splicing factor
SPBC19C7.03	cyr1	adenylate cyclase
SPAC5H10.01	dgc1	mitochondrial D-glutamate cyclase Dgc1 (predicted)
SPAC14C4.15c	dpp1	dipeptidyl peptidase (predicted)
SPAC13C5.02	dre4	splicing associated factor Dre4
SPBC337.07c	ecm14	carboxypeptidase Ecm14 (predicted)
SPBC947.11c	elg1	DNA replication factor C complex subunit Elg1
SPAC821.09	eng1	endo-1,3-beta-glucanase Eng1
SPCC622.16c	epe1	Jmjc domain chromatin associated protein Epe1
SPCC830.11c	fap7	nucleoside-triphosphatase involved in SSU-rRNA maturation Fap7 (predicted)

Systematic ID	Name	Product description
SPBC12C2.13c	fnx1	vacuolar membrane amino acid transmembrane transporter Fnx1
SPBP8B7.12c	fta3	kinetochore protein, CENP-H ortholog Fta3
SPBC1706.03	fzo1	mitofusin, mitochondrial dynamin family fusion GTPase protein (predicted)
SPAC1952.05	gcn5	SAGA complex histone acetyltransferase catalytic subunit Gcn5
SPCC1682.09c	ggc1	mitochondrial guanine nucleotide transmembrane transporter Ggc1 (predicted)
SPBC3D6.07	gpi3	pig-A, phosphatidylinositol N-acetylglucosaminyltransferase subunit Gpi3 (predicted)
SPCC1223.03c	gut2	glycerol-3-phosphate dehydrogenase Gut2 (predicted)
SPAC1834.03c	hhf1	histone H4 h4.1
SPBC1105.12	hhf3	histone H4 h4.3
SPAC1834.04	hht1	histone H3 h3.1
SPBC1105.11c	hht3	histone H3 h3.3
SPAC1783.05	hrp1	ATP-dependent DNA helicase Hrp1
SPAC19G12.06c	hta2	histone H2A beta
SPCC191.11	inv1	external invertase, beta-fructofuranosidase Inv1
SPCC1393.06c	ipi1	rRNA processing protein Ipi1
SPAC4A8.07c	lcb4	sphingoid long chain base kinase (predicted)
SPAC17C9.02c	lys7	alpha-aminoadipate reductase phosphopantetheinyl transferase Lys7
SPCC74.06	mak3	histidine kinase Mak3
SPBC9B6.09c	mdl1	mitochondrial peptide-transporting ATPase
SPBP35G2.10	mit1	SHREC complex ATP-dependent DNA helicase subunit Mit1

Systematic ID	Name	Product description
SPBC1703.04	mlh1	MutL family protein Mlh1 (predicted)
SPCC1682.08c	mpf2	meiotic pumilio family RNA-binding protein Mpf2
SPAC23C11.03	mpp1	U3 snoRNP-associated protein Mpp1 (predicted)
SPBC2G2.15c	mrm2	mitochondrial 2' O-ribose methyltransferase Mrm2 (predicted)
SPAC8F11.03	msh3	MutS protein homolog 3
SPAC222.05c	mss1	mitochondrial tRNA wobble uridine modification GTPase Mss1 (predicted)
SPAC637.12c	mst1	KAT5 family histone acetyltransferase Mst1
SPAC323.05c	mtq2	eRF1 methyltransferase Mtq2 (predicted)
SPCC645.11c	mug117	conserved fungal protein Mug117
SPBC146.10	mug57	FAS1 domain protein Mug57
SPAC6B12.02c	mus7	DNA repair protein Mus7/Mms22
SPAC637.08	nbp35	iron-sulfur cluster assembly ATPase Nbp35 (predicted)
SPBC28F2.10c	ngg1	SAGA complex subunit Ngg1/Ada3
SPCC645.04	nse3	Smc5-6 complex non-SMC MAGE family subunit Nse3
SPAC29E6.10c	nst1	conserved fungal NST1 family protein
SPBC17D11.04c	nto1	histone acetyltransferase complex PHD finger subunit Nto1 (predicted)
SPBC428.01c	nup107	nucleoporin Nup107
SPAC1805.04	nup132	nucleoporin Nup132
SPBC13A2.02	nup82	nucleoporin Nup82
SPAC959.04c	omh6	alpha-1,2-mannosyltransferase Omh6 (predicted)
SPBC685.09	orc2	origin recognition complex subunit Orc2
SPAC664.03	paf1	RNA polymerase II associated Paf1 complex (predicted)
SPAC22G7.04	pan2	PAN complex (poly(A)-specific ribonuclease) ubiquitin C-terminal hydrolase subunit Pan2 (predicted)

Systematic ID	Name	Product description
SPAC12B10.04	pby1	tubulin-tyrosine ligase Pby1 (predicted)
SPAC26H5.03	pcf2	CAF assembly factor (CAF-1) complex subunit B, Pcf2
SPAC6G9.06c	pcp1	pericentrin/kendrin Pcp1
SPBC530.12c	pdf1	palmitoyl protein thioesterase/ dolichol pyrophosphate phosphatase fusion protein Pdf1
SPCC1840.08c	pdi5	protein disulfide isomerase (predicted)
SPAPB24D3.09c	pdr1	ABC transmembrane transporter Pdr1
SPAC3G9.12	peg1	CLASP family microtubule-associated protein
SPCC1183.04c	pet127	mitochondrial RNA metabolism pathway protein Pet127
SPCC188.09c	pfl4	cell surface glycoprotein, flocculin Pfl4
SPAC22F8.11	plc1	phosphoinositide phospholipase C Plc1
SPAC2F3.12c	plp1	thioredoxin fold protein Plp1 (predicted)
SPAC3G9.08	png1	ING family homolog Png1
SPBC14C8.14c	pol5	DNA polymerase phi
SPAC3H1.13	ppk13	serine/threonine protein kinase Ppk13, involvedin negative regulation of vacuole fusion (predicted)
SPBC336.14c	ppk26	PAN complex protein phosphotransferase subunit Ppk26 (predicted)
SPNCRNA.61	prl61	antisense RNA (predicted)
SPBC1711.17	prp16	ATP-dependent RNA helicase Prp16
SPAC10F6.02c	prp22	ATP-dependent RNA helicase Prp22
SPBC20F10.10	psl1	cyclin pho85 family Psl1 (predicted)
SPAC1071.01c	pta1	mRNA cleavage and polyadenylation specificity factor complex subunit Pta1
SPBC1921.06c	pvg3	galactosylxylosylprotein 3-beta-galactosyltransferase Pvg3
SPAC2F3.13c	qtr2	queuine tRNA-ribosyltransferase Qtrtd1/Qtr2

Systematic ID	Name	Product description
SPAC1952.07	rad1	checkpoint clamp complex protein Rad1
SPBC3E7.08c	rad13	DNA repair nuclease Rad13
SPCC970.01	rad16	DNA repair endonuclease XPF
SPAC3C7.03c	rad55	RecA family ATPase Rad55/Rhp55
SPAC6F12.09	rdp1	RNA-directed RNA polymerase Rdp1
SPBC1198.11c	reb1	RNA polymerase I transcription termination factor/ RNA polymerase II transcription factor Reb1
SPAC637.09	rex1	3'-5'- exoribonuclease Rex1 (predicted)
SPBC887.19	rft1	Man5GlcNac2-PP-Dol translocation protein Rft1 (predicted)
SPBC3F6.05	rga1	RhoGAP, GTPase activating protein Rga1
SPCC553.08c	ria1	GTPase Ria1 (predicted)
SPAC688.09	rim2	mitochondrial pyrimidine nucleotide transmembrane transporter Rim2 (predicted)
SPCC4G3.18	rix1	Rix1 complex rRNA processing protein Rix1
SPAC664.06	rlp7	ribosomal protein L7-like Rlp7 involved in ribosome biogenesis (predicted)
SPAC26A3.03c	rmi1	RecQ mediated genome instability protein Rmi1 (predicted)
SPAC4F8.13c	rng2	IQGAP
SPBC17G9.10	rpl1102	60S ribosomal protein L11 (predicted)
SPAC26A3.04	rpl2002	60S ribosomal protein L20 (predicted)
SPBC365.03c	rpl2101	60S ribosomal protein L21 (predicted)
SPAC11E3.15	rpl22	60S ribosomal protein L22 (predicted)
SPBC4F6.04	rpl2502	60S ribosomal protein L25 (predicted)
SPBC776.11	rpl2801	60S ribosomal protein L27/L28
SPBC16C6.11	rpl3201	60S ribosomal protein L32

Systematic ID	Name	Product description
SPCC970.05	rpl3601	60S ribosomal protein L36
SPAC4G9.16c	rpl901	60S ribosomal protein L9
SPCC613.06	rpl902	60S ribosomal protein L9
SPAC3A12.04c	rpp1	RNase P and RNase MRP subunit p30, Rpp1
SPCP1E11.09c	rpp103	60S acidic ribosomal protein Rpp1-3
SPCC16C4.19	rpp21	RNase MRP subunit Rpp21
SPAC13G6.02c	rps101	40S ribosomal protein S3a
SPAC1071.07c	rps1502	40S ribosomal protein S15 (predicted)
SPAC23C11.02c	rps23	40S ribosomal protein S23 (predicted)
SPBC3D6.15	rps2501	40S ribosomal protein S25 (predicted)
SPBC21B10.10	rps402	40S ribosomal protein S4 (predicted)
SPAC2C4.16c	rps801	40S ribosomal protein S8 (predicted)
SPAC2G11.12	rqh1	RecQ type DNA helicase Rqh1
SPAC18G6.11c	rrn3	ribosomal DNA (rDNA) transcription factor Rrn3
SPBC119.12	rud3	Golgi matrix protein Rud3 (predicted)
SPAC3C7.01c	sac12	inositol polyphosphate phosphatase (predicted)
SPCC576.05	sac3	nuclear export factor Sac3 (predicted)
SPBC3B9.15c	scp1	Sre1 cleavage activating protein, Scap Scp1
SPAC12G12.01c	sea4	SEA complex subunit, ubiquitin-protein ligase E3, Sea4 (predicted)
SPCC306.04c	set1	histone lysine methyltransferase Set1
SPAC22E12.11c	set3	histone lysine methyltransferase Set3
SPCC297.04c	set7	histone lysine methyltransferase Set7 (predicted)
SPBC8D2.05c	sfi1	spindle pole body protein Sfi1
SPAC24C9.11	sgd1	ribosome small subunit biogenesis protein Sgd1 (predicted)
SPAC15A10.15	sgo2	inner centromere protein, shugoshin Sgo2

Systematic ID	Name	Product description
SPBC8D2.13	shq1	box H/ACA snoRNP assembly protein Shq1 (predicted)
SPAC12G12.15	sif3	mitochondrial protein, involved in mitochondrial gene expression (predicted)
SPAPB21F2.03	slx9	ribosome biogenesis protein Slx9 (predicted)
SPAC1250.01	snf21	ATP-dependent DNA helicase Snf21
SPAC2F7.08c	snf5	SWI/SNF complex subunit Snf5
SPCC16A11.04	snx12	sorting nexin Snx12 (predicted)
SPCC594.05c	spf1	Set1C PHD Finger protein Spf1
SPNCRNA.1644	spo4-antisense-1	antisense RNA (predicted)
SPAC23H4.17c	srb10	cyclin-dependent protein Srb mediator subunit kinase Srb10
SPCC1322.08	srk1	MAPK-activated protein kinase Srk1
SPAC17G6.10	ssr1	SWI/SNF and RSC complex subunit Ssr1
SPBC12C2.02c	ste20	Rictor homolog, Ste20
SPBC4F6.09	str1	siderophore-iron transmembrane transporter Str1
SPATRNASER.03	sup3	tRNA Serine
SPCTRNASER.11	sup9	tRNA Serine
SPBC354.03	swd3	WD repeat protein Swd3
SPBC16D10.10	tad2	tRNA specific adenosine deaminase subunit Tad2
SPBC15D4.14	taf73	transcription factor TFIID complex subunit Taf5-like
SPCC1223.06	tea1	cell end marker Tea1
SPAC6G10.02c	tea3	cell end marker Tea3
SPBC1706.01	tea4	tip elongation aberrant protein Tea4
SPBC106.12c	tho4	THO complex subunit Tho4 (predicted)
SPBC336.10c	tif512	translation elongation and termination factor eIF5A (predicted)
SPBC17A3.01c	tim50	TIM23 translocase complex subunit Tim50 (predicted)

Systematic ID	Name	Product description
SPAC343.15	tit1	tRNA isopentenyltransferase Tit1
SPBC30D10.10c	tor1	serine/threonine protein kinase Tor1
SPAC27D7.14c	tpr1	RNA polymerase II associated Paf1 complex subunit Tpr1
SPAC17G8.01c	trl1	tRNA ligase Trl1 (predicted)
SPCC285.14	trs130	TRAPP complex subunit Trs130 (predicted)
SPBC3D6.03c	trz2	mitochondrial 3'-tRNA processing endonuclease tRNAse Z, Trz2
SPAC11G7.04	ubi1	ribosomal-ubiquitin fusion protein Ubi1 (predicted)
SPCC1494.05c	ubp12	CSN-associated deubiquitinating enzyme Ubp12
SPAC19A8.08	upf2	nonsense-mediated decay protein Upf2
SPAC3A12.09c	ure4	urease accessory protein UreD
SPBC4B4.07c	usp102	U1 snRNP-associated protein Usp102
SPBC19C7.09c	uve1	endonuclease Uve1
SPBC21C3.01c	vps1301	chorein homolog Vps13a (predicted)
SPAC9E9.14	vps24	ESCRT III complex subunit Vps24
SPBC713.05	wdr83	WD repeat protein, human MAPK organizer 1 (MORG1) family Wdr83 (predicted)
SPCC11E10.05c	ynd1	nucleoside diphosphatase Ynd1
SPCC1795.09	yps1	aspartic protease, yapsin family, unknown specificity Yps1
SPAC11D3.06	-	MatE family transmembrane transporter (predicted)
SPAC11D3.18c	-	carboxylic acid transmembrane transporter (predicted)
SPAC11E3.02c	-	C2 domain protein, Munc family, implicated in exocytosis
SPAC144.05	-	ATP-dependent DNA helicase/ ubiquitin-protein ligase E3 (predicted)

Systematic ID	Name	Product description
SPAC17H9.08	-	mitochondrial coenzyme A transmembrane transporter (predicted)
SPAC186.06	-	phenazine biosynthesis PhzF protein family
SPAC22F8.05	-	alpha,alpha-trehalose-phosphate synthase (predicted)
SPAC22H10.08	-	DUF2009 family protein, conserved in yeast and apicomplexa
SPAC25A8.03c	-	mitochondrial arginine methyltransferase, human NDUFAF7 ortholog (predicted)
SPAC26F1.08c	-	malate transmembrane transporter (predicted)
SPAC2E1P3.05c	-	fungal cellulose binding domain protein
SPAC3H5.09c	-	mitochondrial conserved protein, human KIAA0100 ortholog (predicted)
SPAC6C3.06c	-	P-type ATPase, calcium transporting (predicted)
SPAC806.06c	-	nicotinamide mononucleotide (NMN) adenylyltransferase (predicted)
SPAC8E11.05c	-	DUF5102 family conserved fungal protein, associated with clathrin coated vesicles (predicted)
SPATRNAMET.01	-	tRNA Methionine
SPATRNAMET.03	-	tRNA Methionine
SPATRNASER.01	-	tRNA Serine
SPBC1348.09	-	short chain dehydrogenase (predicted)
SPBC1683.05	-	uricil/uridine permese transmembrane transporter family (predicted)
SPBC1773.16c	-	transcription factor, zf-fungal binuclear cluster type(predicted)
SPBC19C7.11	-	ClC chloride channel (predicted)

Systematic ID	Name	Product description
SPBC21C3.03	-	mitochondrial membrane ABC1 kinase family protein, unknown role, human ADCK2 ortholog
SPBC27B12.09c	-	mitochondrial FAD transmembrane transporter (predicted)
SPBC3B8.06	-	DUF2427 family conserved fungal protein
SPBC56F2.07c	-	ribosome biogenesis factor recycling AAA family ATPase (predicted)
SPBC947.06c	-	spermidine family transmembrane transporter (predicted)
SPBPJ4664.02	-	cell surface glycoprotein, flocculin, related to Gsf2
SPBTRNAASN.01	-	tRNA Asparagine
SPBTRNAMET.06	-	tRNA Methionine
SPCC11E10.09c	-	alpha-amylase homolog (predicted)
SPCC1529.01	-	transmembrane transporter (predicted)
SPCC1672.09	-	triglyceride lipase-cholesterol esterase (predicted)
SPCC1672.11c	-	P-type ATPase P5 type (predicted)
SPCC18.08	-	mitochondrial lysine-tRNA ligase (predicted)
SPCC1827.07c	-	SPX/EXS domain protein (predicted)
SPCC320.08	-	transmembrane transporter (predicted)
SPCC550.08	-	N-acetyltransferase (predicted)
SPCC594.04c	-	steroid oxidoreductase superfamily protein (predicted)
SPCC613.01	-	transmembrane transporter (predicted)
SPCP1E11.10	-	ankyrin repeat protein, unknown biological role
SPCTRNAMET.07	-	tRNA Methionine
SPNCRNA.1303	-	intergenic RNA (predicted)
SPNCRNA.865	-	antisense RNA (predicted)
SPNCRNA.953	-	intergenic RNA (predicted)

Systematic ID	Name	Product description
SPNCRNA.979	-	antisense RNA (predicted)

A.3.2 Gene Ontology enrichment

External ID	GeneSet Name	over represented/under represented	Corrected p-value
GO:0008150;	biological process	Enriched	8.1762799999999998E-13
GO:0009987;	cellular process	Enriched	1.19224E-9
GO:0043170;	macromolecule metabolic process	Enriched	4.9067999999999996E-9
GO:0044260;	cellular macromolecule metabolic process	Enriched	6.3501700000000005E-8
GO:0090304;	nucleic acid metabolic process	Enriched	1.23935E-7
GO:0006974;	cellular response to DNA damage stimulus	Enriched	8.0700700000000005E-7
GO:0071103;	DNA conformation change	Enriched	2.2914900000000001E-6
GO:0044238;	primary metabolic process	Enriched	2.89065E-6
GO:0006323;	DNA packaging	Enriched	3.8145500000000002E-6
GO:0006325;	chromatin organization	Enriched	6.7833300000000003E-6
GO:0071704;	organic substance metabolic process	Enriched	6.7833300000000003E-6
GO:0006139;	nucleobase-containing compound metabolic process	Enriched	1.4335700000000001E-5
GO:0031497;	chromatin assembly	Enriched	1.5452200000000001E-5
GO:0006333;	chromatin assembly or disassembly	Enriched	2.0701699999999999E-5
GO:0008152;	metabolic process	Enriched	2.0701699999999999E-5
GO:0051276;	chromosome organization	Enriched	2.0701699999999999E-5

External ID	GeneSet Name	over represented/under represented	Corrected p-value
GO:0043933;	macromolecular complex subunit organization	Enriched	3.00925E-5
GO:0010467;	gene expression	Enriched	3.9332500000000003E-5
GO:0016568;	chromatin modification	Enriched	5.6220600000000002E-5
GO:0006725;	cellular aromatic compound metabolic process	Enriched	9.9301099999999997E-5
GO:0071840;	cellular component organization or biogenesis	Enriched	1.15616E-4
GO:0033554;	cellular response to stress	Enriched	1.31564E-4
GO:0006950;	response to stress	Enriched	1.3453799999999999E-4
GO:1901360;	organic cyclic compound metabolic process	Enriched	1.3453799999999999E-4
GO:0046483;	heterocycle metabolic process	Enriched	1.41507E-4
GO:0006259;	DNA metabolic process	Enriched	1.5058399999999999E-4
GO:0006281;	DNA repair	Enriched	1.9199500000000001E-4
GO:0034641;	cellular nitrogen compound metabolic process	Enriched	2.4326799999999999E-4
GO:0044237;	cellular metabolic process	Enriched	3.5980400000000003E-4
GO:0044085;	cellular component biogenesis	Enriched	4.5406100000000002E-4
GO:0006807;	nitrogen compound metabolic process	Enriched	5.3895199999999999E-4
GO:0034645;	cellular macromolecule biosynthetic process	Enriched	6.27168E-4
GO:0009059;	macromolecule biosynthetic process	Enriched	6.3843299999999999E-4
GO:0044699;	single-organism process	Enriched	1.1922899999999999E-3
GO:0006338;	chromatin remodeling	Enriched	1.4951400000000001E-3

External ID	GeneSet Name	over represented/under represented	Corrected p-value
GO:0065004;	protein-DNA complex assembly	Enriched	1.7476099999999999E-3
GO:0071824;	protein-DNA complex subunit organization	Enriched	1.8114699999999999E-3
GO:0051716;	cellular response to stimulus	Enriched	2.04747E-3
GO:0044763;	single-organism cellular process	Enriched	2.0733800000000001E-3
GO:0050896;	response to stimulus	Enriched	2.1151799999999999E-3
GO:0070828;	heterochromatin organization	Enriched	2.53316E-3
GO:0031507;	heterochromatin assembly	Enriched	2.5760399999999999E-3
GO:0065003;	macromolecular complex assembly	Enriched	2.8392299999999999E-3
GO:0018205;	peptidyl-lysine modification	Enriched	2.8780899999999998E-3
GO:0016569;	covalent chromatin modification	Enriched	2.98563E-3
GO:0016570;	histone modification	Enriched	2.98563E-3
GO:0016070;	RNA metabolic process	Enriched	3.9333199999999997E-3
GO:0010558;	negative regulation of macromolecule biosynthetic process	Enriched	4.3495599999999997E-3
GO:2000113;	negative regulation of cellular macromolecule biosynthetic process	Enriched	4.3495599999999997E-3
GO:0016043;	cellular component organization	Enriched	6.0228599999999997E-3
GO:0045934;	negative regulation of nucleobase-containing compound metabolic process	Enriched	8.6910800000000003E-3
GO:0009890;	negative regulation of biosynthetic process	Enriched	8.8735299999999993E-3
GO:0031327;	negative regulation of cellular biosynthetic process	Enriched	8.8735299999999993E-3

External ID	GeneSet Name	over represented/under represented	Corrected p-value
GO:0051172;	negative regulation of nitrogen compound metabolic process	Enriched	8.8735299999999993E-3
GO:0019219;	regulation of nucleobase-containing compound metabolic process	Enriched	9.0120500000000006E-3
GO:0051171;	regulation of nitrogen compound metabolic process	Enriched	9.1160100000000008E-3

A.4 Negative non-scaling proteins

A.4.1 Gene list

Systematic ID	Name	Product description
SPBC56F2.09c	arg5	arginine specific carbamoyl-phosphate synthase subunit Arg5
SPBC106.17c	cys2	homoserine O-acetyltransferase (predicted)
SPAC23H4.06	gln1	glutamate-ammonia ligase Gln1
SPBC26H8.06	grx4	monothiol glutaredoxin Grx4
SPBC8D2.03c	hhf2	histone H4 h4.2
SPBC8D2.04	hht2	histone H3 h3.2
SPAC19G12.06c	hta2	histone H2A beta
SPCC622.09	htb1	histone H2B Htb1
SPBC1734.11	mas5	DNAJ domain protein Mas5 (predicted)
SPBC1D7.04	mlo3	RNA binding protein Mlo3
SPCC1906.01	mpg1	mannose-1-phosphate guanyltransferase Mpg1

Systematic ID	Name	Product description
SPBC725.02	mpr1	histidine-containing response regulator phosphotransferase Mpr1
SPBC23E6.10c	mri1	methylthioribose-1-phosphate isomerase Mri1 (predicted)
SPCC1223.02	nmt1	4-amino-5-hydroxymethyl-2-methylpyrimidine phosphate synthase Nmt1
SPBC30D10.13c	pdb1	pyruvate dehydrogenase e1 component beta subunit Pdb1
SPAC17H9.14c	pdi2	protein disulfide isomerase
SPCC830.07c	psi1	DNAJ domain protein Psi1
SPAC8F11.10c	pvg1	pyruvyltransferase Pvg1
SPBC17G9.10	rpl1102	60S ribosomal protein L11 (predicted)
SPCC364.03	rpl1702	60S ribosomal protein L17 (predicted)
SPBC365.03c	rpl2101	60S ribosomal protein L21 (predicted)
SPBP8B7.03c	rpl402	60S ribosomal protein L4 (predicted)
SPAC3H5.07	rpl702	60S ribosomal protein L7b involved in cytoplasmic translation
SPBC3B9.13c	rpp102	60S acidic ribosomal protein A3
SPCP1E11.09c	rpp103	60S acidic ribosomal protein Rpp1-3
SPBC16D10.11c	rps1801	40S ribosomal protein S18 (predicted)
SPAC25G10.06	rps2801	40S ribosomal protein S28 (predicted)
SPBC1685.09	rps29	40S ribosomal protein S29 (predicted)
SPBC29A3.16	rrs1	ribosome biogenesis protein Rrs1
SPAC24C9.12c	shm1	serine hydroxymethyltransferase Shm1 (predicted)
SPCC1739.13	ssa2	heat shock protein Ssa2
SPBC23G7.05	sui1	translation initiation factor eIF1
SPAC2F7.13c	wrs1	cytoplasmic tryptophan-tRNA ligase Wrs1 (predicted)
SPBC16D10.06	zrt1	ZIP zinc transmembrane transporter Zrt1
SPAC3G6.03c	-	Maf-like protein, human ASMTL ortholog

Systematic ID	Name	Product description
SPAC688.03c	-	human AMMECR1 homolog
SPBC8E4.03	-	agmatinase 2 (predicted)

A.4.2 Gene Ontology enrichment

External ID	GeneSet Name	over represented/under represented	Corrected p-value
GO:0002181;	cytoplasmic translation	Enriched	5.8507099999999996E-6
GO:0006412;	translation	Enriched	5.3881000000000003E-4

Bibliography

[1] C. E. Aitken and J. R. Lorsch. A mechanistic overview of translation initiation in eukaryotes. *Nature Structural & Molecular Biology*, 19(6):568–576, 6 2012.

[2] C. Allmang, P. Mitchell, E. Petfalski, and D. Tollervey. Degradation of ribosomal RNA precursors by the exosome. *Nucleic acids research*, 28(8):1684–91, 4 2000.

[3] R. C. Allshire and K. Ekwall. Epigenetics Regulation of Chromatin States in Schizosaccharomyces pombe. *Cold Spring Harb Perspect Biol.*, 7:1–25, 2015.

[4] A. A. Amodeo, D. Jukam, A. F. Straight, and J. M. Skotheim. Histone titration against the genome sets the DNA-to-cytoplasm threshold for the Xenopus midblastula transition. *Proceedings of the National Academy of Sciences of the United States of America*, 112(10):1086–95, 3 2015.

[5] A. A. Amodeo and J. M. Skotheim. Cell-Size Control. *Cold Spring Harbor perspectives in biology*, 8(4):a019083, 4 2016.

[6] Y. Aoi, S. a. Kawashima, V. Simanis, M. Yamamoto, and M. Sato. Optimization of the analogue-sensitive Cdc2/Cdk1 mutant by in vivo selection eliminates physiological limitations to its use in cell cycle analysis. *Open biology*, 4(7), 2014.

[7] J. E. Arenas and J. N. Abelson. Prp43: An RNA helicase-like factor involved in spliceosome disassembly. *Proceedings of the National Academy of Sciences of the United States of America*, 94(22):11798–802, 10 1997.

[8] M. J. Aryee, J. A. Gutiérrez-Pabello, I. Kramnik, T. Maiti, and J. Quackenbush. An improved empirical bayes approach to estimating differential gene expression in microarray

time-course data: BETR (Bayesian Estimation of Temporal Regulation). *BMC bioinformatics*, 10(1):409, 2009.

[9] S. R. Atkinson. *The Fission Yeast Non-Coding Transcriptome*. PhD thesis, 2014.

[10] T. L. Bailey, M. Boden, F. A. Buske, M. Frith, C. E. Grant, L. Clementi, J. Ren, W. W. Li, and W. S. Noble. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Research*, 37(Web Server):W202–W208, 7 2009.

[11] M. Basan, M. Zhu, X. Dai, M. Warren, D. Sévin, Y.-p. Wang, and T. Hwa. Inflating bacterial cells by increased protein synthesis. pages 1–7, 2015.

[12] S. Baumgärtner and I. M. Tolić-Nørrelykke. Growth pattern of single fission yeast cells is bilinear and depends on temperature and DNA synthesis. *Biophysical Journal*, 96(10):4336–4347, 2009.

[13] C. A. Beelman, A. Stevens, G. Caponigro, T. E. LaGrandeur, L. Hatfield, D. M. Fortner, and R. Parker. An essential component of the decapping enzyme required for normal rates of mRNA turnover. *Nature*, 382(6592):642–646, 8 1996.

[14] D. a. Bitton, C. Rallis, D. C. Jeffares, G. C. Smith, Y. Y. C. Chen, S. Codlin, S. Marguerat, and J. Bähler. LaSSO, a strategy for genome-wide mapping of intronic lariats and branch points using RNA-seq. *Genome Research*, 24(7):1169–1179, 2014.

[15] D. A. Bitton, F. Schubert, S. Dey, M. Okoniewski, G. C. Smith, S. Khadayate, V. Pancaldi, V. Wood, and J. Bähler. AnGeLi: A tool for the analysis of gene lists from fission yeast. *Frontiers in Genetics*, 6(NOV):1–9, 2015.

[16] M. A. Blanco, A. Sánchez-Díaz, J. M. de Prada, and S. Moreno. APC ste9/srw1 promotes degradation of mitotic cyclins in G 1 and is inhibited by cdc2 phosphorylation. *The EMBO Journal*, 19(15):3945–3955, 8 2000.

[17] R. Boeck, S. Tarun, M. Rieger, J. A. Deardorff, S. Müller-Auer, and A. B. Sachs. The yeast Pan2 protein is required for poly(A)-binding protein-stimulated poly(A)-nuclease activity. *The Journal of biological chemistry*, 271(1):432–8, 1 1996.

[18] G. T. Booth, I. X. Wang, V. G. Cheung, and J. T. Lis. Divergence of a conserved elongation factor and transcription regulation in budding and fission yeast. *Genome research*, 26(6):799–811, 2016.

[19] S. Buratowski. Progression through the RNA polymerase II CTD cycle. *Molecular cell*, 36(4):541–6, 11 2009.

[20] L. Cai, B. M. Sutter, B. Li, and B. P. Tu. Acetyl-CoA induces cell growth and proliferation by promoting the acetylation of histones at growth genes. *Molecular cell*, 42(4):426–37, 5 2011.

[21] G. Caponigro and R. Parker. Multiple functions for the poly(A)-binding protein in mRNA decapping and deadenylation in yeast. *Genes & development*, 9(19):2421–32, 10 1995.

[22] C. R. Carlson, B. Grallert, T. Stokke, and E. Boye. Regulation of the start of DNA replication in Schizosaccharomyces pombe. *Journal of cell science*, 112 (Pt 6:939–46, 3 1999.

[23] A. Carpy, K. Krug, S. Graf, A. Koch, S. Popic, S. Hauf, and B. Macek. Absolute proteome and phosphoproteome dynamics during the cell cycle of fission yeast. *Molecular & cellular proteomics : MCP*, 4 2014.

[24] E. Castonguay and Emilie. *Identification of novel inhibitors of heterochromatin integrity through a chemical screen in fission yeast*. PhD thesis, 6 2014.

[25] D. Chen, W. M. Toone, J. Mata, R. Lyne, G. Burns, K. Kivinen, A. Brazma, N. Jones, and J. Bähler. Global transcriptional responses of fission yeast to environmental stress. *Molecular biology of the cell*, 14(1):214–29, 1 2003.

[26] N. Chica, A. E. Rozalén, L. Pérez-Hidalgo, A. Rubio, B. Novak, and S. Moreno. Nutritional control of cell size by the greatwall-endosulfine-PP2A·B55 pathway. *Current Biology*, 26(3):319–330, 2 2016.

[27] R. Christiano, N. Nagaraj, F. Fröhlich, and T. C. Walther. Global Proteome Turnover Analyses of the Yeasts S. cerevisiae and S. pombe. *Cell Reports*, 9(5):1959–1965, 2014.

[28] T. R. Coleman and W. G. Dunphy. Cdc2 regulatory factors. *Current opinion in cell biology*, 6(6):877–82, 12 1994.

[29] C. Collart, G. E. Allen, C. R. Bradshaw, J. C. Smith, and P. Zegerman. Titration of Four Replication Factors Is Essential for the Xenopus laevis Midblastula Transition. *Science*, 341(6148), 2013.

[30] A. Conesa, M. J. Nueda, A. Ferrer, and M. Talón. maSigPro: A method to identify significantly differential expression profiles in time-course microarray experiments. *Bioinformatics*, 22(9):1096–1102, 2006.

[31] S. Cooper. Distinguishing between linear and exponential cell growth during the division cycle: single-cell studies, cell-culture studies, and the object of cell-cycle research. *Theoretical biology & medical modelling*, 3:10, 2006.

[32] D. Coudreuse and P. Nurse. Driving the cell cycle with a minimal CDK control network. *Nature*, 468(7327):1074–1079, 2010.

[33] E. Czeko, M. Seizl, C. Augsberger, T. Mielke, and P. Cramer. Iwr1 directs RNA polymerase II nuclear import. *Molecular cell*, 42(2):261–6, 4 2011.

[34] R. P. das Neves, N. S. Jones, L. Andreu, R. Gupta, T. Enver, and F. J. Iborra. Connecting variability in global transcription rate to mitochondrial variability. *PLoS biology*, 8(12):e1000560, 1 2010.

[35] C. M. DeGennaro, B. H. Alver, S. Marguerat, E. Stepanova, C. P. Davis, J. Bähler, P. J. Park, and F. Winston. Spt6 regulates intragenic and antisense transcription, nucleosome positioning, and histone modifications genome-wide in fission yeast. *Molecular and cellular biology*, 33(24):4779–4792, 12 2013.

[36] M. A. Dillies, A. Rau, J. Aubert, C. Hennequet-Antier, M. Jeanmougin, N. Servant, C. Keime, N. S. Marot, D. Castel, J. Estelle, G. Guernec, B. Jagla, L. Jouneau, D. Laloë, C. Le Gall, B. Schaëffer, S. Le Crom, M. Guedj, and F. Jaffrézic. A comprehensive evaluation of normalization methods for Illumina high-throughput RNA sequencing data analysis. *Briefings in Bioinformatics*, 14(6):671–683, 2013.

[37] S. Dischinger, A. Krapp, L. Xie, J. R. Paulson, and V. Simanis. Chemical genetic analysis of the regulatory role of Cdc2p in the S. pombe septation initiation network. *Journal of cell science*, 121(Pt 6):843–853, 2008.

[38] T. Dunckley and R. Parker. The DCP2 protein is required for mRNA decapping in Saccharomyces cerevisiae and contains a functional MutT motif. *The EMBO Journal*, 18(19):5411–5422, 10 1999.

[39] P. Eser, C. Demel, K. C. Maier, B. Schwalb, N. Pirkl, D. E. Martin, P. Cramer, and A. Tresch. Periodic mRNA synthesis and degradation co-operate during cell cycle gene expression. *Molecular systems biology*, 10:717, 1 2014.

[40] P. Eser, L. Wachutka, K. C. Maier, C. Demel, M. Boroni, S. Iyer, P. Cramer, and J. Gagneur. Determinants of RNA metabolism in the Schizosaccharomyces pombe genome. *Molecular systems biology*, 12(2):857, 2 2016.

[41] B. J. Fair and J. A. Pleiss. The power of fission: yeast as a tool for understanding complex splicing. *Current Genetics*, 63(3):375–380, 6 2017.

[42] D. Finley, H. D. Ulrich, T. Sommer, and P. Kaiser. The ubiquitin-proteasome system of Saccharomyces cerevisiae. *Genetics*, 192(2):319–60, 10 2012.

[43] D. L. Fisher and P. Nurse. A single fission yeast mitotic cyclin B p34cdc2 kinase promotes both S-phase and mitosis in the absence of G1 cyclins. *The EMBO journal*, 15(4):850–60, 2 1996.

[44] R. S. Fraser and P. Nurse. Altered patterns of ribonucleic acid synthesis during the cell cycle: a mechanism compensating for variation in gene concentration. *Journal of cell science*, 35:25–40, 2 1979.

[45] R. M. N. Friis and M. C. Schultz. Untargeted tail acetylation of histones in chromatin: lessons from yeast. *Biochemistry and cell biology = Biochimie et biologie cellulaire*, 87(1):107–16, 2 2009.

[46] T. Galitski, A. J. Saldanha, C. A. Styles, E. S. Lander, and G. R. Fink. Ploidy Regulation of Gene Expression. *Science*, 285(5425), 1999.

[47] R. C. Gentleman, V. J. Carey, D. M. Bates, B. Bolstad, M. Dettling, S. Dudoit, B. Ellis, L. Gautier, Y. Ge, J. Gentry, K. Hornik, T. Hothorn, W. Huber, S. Iacus, R. Irizarry, F. Leisch, C. Li, M. Maechler, A. J. Rossini, G. Sawitzki, C. Smith, G. Smyth, L. Tierney, J. Y. H. Yang, and J. Zhang. Bioconductor: open software development for computational biology and bioinformatics. *Genome biology*, 5(10):R80, 2004.

[48] S. Gonzalez and C. Rallis. The TOR Signaling Pathway in Spatial and Temporal Control of Cell Size and Growth. *Frontiers in Cell and Developmental Biology*, 5:61, 6 2017.

[49] T. Gross and N. F. Käufer. Cytoplasmic ribosomal protein genes of the fission yeast Schirosaccharomyces pombe display a unique promoter type: A suggestion for nomenclature of cytoplasmic ribosomal proteins in databases. *Nucleic Acids Research*, 26(14):3319–3322, 1998.

[50] F. Hahne, N. LeMeur, R. R. Brinkman, B. Ellis, P. Haaland, D. Sarkar, J. Spidlen, E. Strain, and R. Gentleman. flowCore: a Bioconductor package for high throughput flow cytometry. *BMC bioinformatics*, 10:106, 2009.

[51] C. G. Hansen, Y. L. D. Ng, W.-L. M. Lam, S. W. Plouffe, and K.-L. Guan. The Hippo pathway effectors YAP and TAZ promote cell growth by modulating amino acid signaling to mTORC1. *Cell Research*, 25(12):1299–1313, 12 2015.

[52] A. Hasan, C. Cotobal, C. D. S. Duncan, and J. Mata. Systematic Analysis of the Role of RNA-Binding Proteins in the Regulation of RNA Stability. *PLoS genetics*, 10(11):e1004684, 11 2014.

[53] A. Hergovich and B. A. Hemmings. Seminars in Cell & Developmental Biology Hippo signalling in the G2 / M cell cycle phase : Lessons learned from the yeast MEN and SIN pathways. *Seminars in Cell and Developmental Biology*, 23(7):794–802, 9 2012.

[54] C. S. Hoffman, V. Wood, and P. A. Fantes. An Ancient Yeast for Young Geneticists:

A Primer on the Schizosaccharomyces pombe Model System. *Genetics*, 201(2):403–423, 2015.

[55] C. L. Hsu and A. Stevens. Yeast cells lacking 5'–¿3' exoribonuclease 1 contain mRNA species that are poly(A) deficient and partially lack the 5' cap structure. *Molecular and cellular biology*, 13(8):4826–35, 8 1993.

[56] S. Hui, J. M. Silverman, S. S. Chen, D. W. Erickson, M. Basan, J. Wang, T. Hwa, and J. R. Williamson. Quantitative proteomic analysis reveals a simple strategy of global resource allocation in bacteria. *Molecular systems biology*, pages 1–15, 2015.

[57] R. Janke, A. E. Dodson, and J. Rine. Metabolism and Epigenetics. *Annual Review of Cell and Developmental Biology*, 31(1):473–496, 11 2015.

[58] A. Jansen and K. J. Verstrepen. Nucleosome positioning in Saccharomyces cerevisiae. *Microbiology and molecular biology reviews : MMBR*, 75(2):301–320, 2011.

[59] D. C. Jeffares, C. Rallis, A. Rieux, D. Speed, M. Převorovský, T. Mourier, F. X. Marsellach, Z. Iqbal, W. Lau, T. M. K. Cheng, R. Pracana, M. Mülleder, J. L. D. 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. E. U. Hellberg, L. V. Trigt, L. Jeffery, J.-j. Li, S. Atkinson, M. Thodberg, M. Febrer, K. Mclay, N. Drou, W. Brown, J. Hayles, R. E. C. Salas, M. Ralser, N. Maniatis, D. J. Balding, and F. Balloux. The genomic and phenotypic diversity of Schizosaccharomyces pombe. *Nature Publishing Group*, 47(January):235–241, 2015.

[60] I. Jonkers and J. T. Lis. Getting up to speed with transcription elongation by RNA polymerase II. *Nature Reviews Molecular Cell Biology*, 16(3):167–177, 2015.

[61] P. Jorgensen, J. L. Nishikawa, B.-J. Breitkreutz, and M. Tyers. Systematic identification of pathways that couple cell growth and division in yeast. *Science (New York, N.Y.)*, 297(5580):395–400, 7 2002.

[62] P. Jorgensen and M. Tyers. How cells coordinate growth and division. *Current biology : CB*, 14(23):1014–27, 12 2004.

[63] S. R. Joseph, M. Pálfy, L. Hilbert, M. Kumar, J. Karschau, V. Zaburdaev, A. Shevchenko, and N. L. Vastenhouw. Competition between histone and transcription factor binding regulates the onset of transcription in zebrafish embryos. *eLife*, 6:e23326, 4 2017.

[64] S. Jun and S. Taheri-Araghi. Cell-size maintenance: universal strategy revealed. *Trends in Microbiology*, 23(1):4–6, 1 2015.

[65] M. Kafri, E. Metzl-Raz, G. Jona, and N. Barkai. The Cost of Protein Production. *Cell Reports*, 14(1):22–31, 12 2016.

[66] R. Kafri, J. Levy, M. B. Ginzberg, S. Oh, G. Lahav, and M. W. Kirschner. Dynamics extracted from fixed cells reveal feedback linking cell growth to cell cycle. *Nature*, 494(7438):480–3, 2 2013.

[67] D. Keifenheim, X.-m. Sun, E. D'Souza, M. Ohira, M. Magner, M. Mayhew, S. Marguerat, and N. Rhind. Size-Dependent Accumulation of the Mitotic Activator Cdc25 as a Mechanism of Size Control in Fission Yeast. *Current Biology*, 27(10):1491–1497, 2017.

[68] C. Keller, R. Kulasegaran-Shylini, Y. Shimada, H.-R. Hotz, and M. Bühler. Noncoding RNAs prevent spreading of a repressive histone mark. *Nature structural & molecular biology*, 20(8):994–1000, 2013.

[69] D. Killander and A. Zetterberg. A quantitative cytochemical investigation of the relationship between cell mass and initiation of DNA synthesis in mouse fibroblasts in vitro. *Experimental cell research*, 40(1):12–20, 10 1965.

[70] M. Kim, N. J. Krogan, L. Vasiljeva, O. J. Rando, E. Nedea, J. F. Greenblatt, and S. Buratowski. The yeast Rat1 exonuclease promotes transcription termination by RNA polymerase II. *Nature*, 432(7016):517–522, 11 2004.

[71] S. Klumpp and T. Hwa. Bacterial growth: Global effects on gene expression, growth feedback and proteome partition. *Current Opinion in Biotechnology*, 28:96–102, 2014.

[72] S. Klumpp, M. Scott, S. Pedersen, and T. Hwa. Molecular crowding limits translation

and cell growth. *Proceedings of the National Academy of Sciences of the United States of America*, 110(42):16754–9, 2013.

[73] S. Klumpp, Z. Zhang, and T. Hwa. Growth rate-dependent global effects on gene expression in bacteria. *Cell*, 139(7):1366–75, 12 2009.

[74] R. Kolde. *pheatmap: Pretty Heatmaps*, 2015. R package version 1.0.8.

[75] T. Kouzarides. Chromatin modifications and their function. *Cell*, 128(4):693–705, 2 2007.

[76] G. Kudla, L. Lipinski, F. Caffin, A. Helwak, and M. Zylicz. High guanine and cytosine content increases mRNA levels in mammalian cells. *PLoS Biology*, 4(6):0933–0942, 2006.

[77] K. Kume, H. Cantwell, F. R. Neumann, A. W. Jones, A. P. Snijders, and P. Nurse. A systematic genomic screen implicates nucleocytoplasmic transport and membrane growth in nuclear size control. *PLOS Genetics*, 13(5):e1006767, 5 2017.

[78] A. B. Lantermann, T. Straub, A. Strålfors, G.-C. Yuan, K. Ekwall, and P. Korber. Schizosaccharomyces pombe genome-wide nucleosome mapping reveals positioning mechanisms distinct from those of Saccharomyces cerevisiae. *Nature structural & molecular biology*, 17(2):251–257, 2010.

[79] M. Lawrence, R. Gentleman, and V. Carey. rtracklayer: an r package for interfacing with genome browsers. *Bioinformatics*, 25:1841–1842, 2009.

[80] M. Lawrence, W. Huber, H. Pagès, P. Aboyoun, M. Carlson, R. Gentleman, M. T. Morgan, and V. J. Carey. Software for Computing and Annotating Genomic Ranges. *PLoS Computational Biology*, 9(8):1–10, 2013.

[81] H. O. Lee, J. M. Davidson, and R. J. Duronio. Endoreplication: polyploidy with purpose. *Genes & development*, 23(21):2461–77, 11 2009.

[82] H. Lempiäinen and D. Shore. Growth control and ribosome biogenesis. *Current Opinion in Cell Biology*, 21(6):855–863, 2009.

[83] M. Lenormand, F. Jabot, and G. Deffuant. Adaptive approximate Bayesian computation for complex models. *Computational Statistics*, 28(6):2777–2796, 2013.

[84] H. S. Leong, K. Dawson, C. Wirth, Y. Li, Y. Connolly, D. L. Smith, C. R. M. Wilkinson, and C. J. Miller. A global non-coding RNA system modulates fission yeast protein levels in response to stress. *Nature communications*, 5(May):3947, 2014.

[85] U. Leupold. Chapter 8 Genetical Methods for Schizosaccharomyces pombe. *Methods in Cell Biology*, 4:169–177, 1970.

[86] B. Li, M. Carey, and J. L. Workman. The role of chromatin during transcription. *Cell*, 128(4):707–19, 2 2007.

[87] D. Li, T. Wei, C. M. Abbott, and D. Harrich. The unexpected roles of eukaryotic translation elongation factors in RNA virus replication and pathogenesis. *Microbiology and molecular biology reviews : MMBR*, 77(2):253–66, 6 2013.

[88] H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, and R. Durbin. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 25(16):2078–2079, 2009.

[89] H. Li, J. Hou, L. Bai, C. Hu, P. Tong, Y. Kang, X. Zhao, and Z. Shao. Genome-wide analysis of core promoter structures in Schizosaccharomyces pombe with DeepCAGE. *RNA biology*, 12(5):525–37, 2015.

[90] J. Liepe, P. Kirk, S. Filippi, T. Toni, C. P. Barnes, and M. P. H. Stumpf. A framework for parameter estimation and model selection from experimental data in systems biology using approximate Bayesian computation. *Nature Protocols*, 9(2):439–456, 2014.

[91] L. López-Maury, S. Marguerat, and J. Bähler. Tuning gene expression to changing environments: from rapid responses to evolutionary adaptation. *Nature Reviews Genetics*, 9(8):583–593, 8 2008.

[92] M. I. Love, W. Huber, and S. Anders. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology*, pages 1–21, 2014.

[93] R. Lyne, G. Burns, J. Mata, C. J. Penkett, G. Rustici, D. Chen, C. Langford, D. Vetrie, and J. Bahler. Whole-genome microarrays of fission yeast: characteristics, accuracy, reproducibility, and processing of array data. *BMC Genomics*, 4(1):27, 2003.

[94] N. Maclean. Ribosome numbers in a fission yeast. *Nature*, 207(994):322–3, 7 1965.

[95] M. Malecki, D. A. Bitton, M. Rodríguez-López, C. Rallis, N. G. Calavia, G. C. Smith, and J. Bähler. Functional and Regulatory Profiling of Energy Metabolism in Fission Yeast. *Genome biology*, (October):1–25, 2016.

[96] S. Marguerat, A. Schmidt, S. Codlin, W. Chen, R. Aebersold, and J. Bähler. Quantitative analysis of fission yeast transcriptomes and proteomes in proliferating and quiescent cells. *Cell*, 151(3):671–683, 2012.

[97] J. M. Marin, P. Pudlo, C. P. Robert, and R. J. Ryder. Approximate Bayesian computational methods. *Statistics and Computing*, 22(6):1167–1180, 2012.

[98] A. Martin, S. Schneider, and B. Schwer. Prp43 Is an Essential RNA-dependent ATPase Required for Release of Lariat-Intron from the Spliceosome. *Journal of Biological Chemistry*, 277(20):17743–17750, 5 2002.

[99] S. G. Martin. Geometric control of the cell cycle. *Cell Cycle*, 8(22):3643–3647, 11 2009.

[100] C. Martín-Castellanos, M. A. Blanco, J. M. de Prada, and S. Moreno. The puc1 cyclin regulates the G1 phase of the fission yeast cell cycle in response to cell size. *Molecular biology of the cell*, 11(2):543–54, 2 2000.

[101] B. Martinez-Pastor, C. Cosentino, and R. Mostoslavsky. A tale of metabolites: the cross-talk between chromatin and energy metabolism. *Cancer discovery*, 3(5):497–501, 5 2013.

[102] J. Mata and J. Bähler. Global roles of Ste11p, cell type, and pheromone in the control of gene expression during early sexual differentiation in fission yeast. *Proceedings of the National Academy of Sciences of the United States of America*, 103(42):15517–22, 10 2006.

[103] D. J. McCarthy, Y. Chen, and G. K. Smyth. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Research*, 40(10):4288–4297, 2012.

[104] E. C. Merkhofer, P. Hu, and T. L. Johnson. Introduction to cotranscriptional RNA splicing. *Methods in molecular biology (Clifton, N.J.)*, 1126:83–96, 2014.

[105] E. Metzl-Raz, M. Kafri, G. Yaakov, I. Soifer, Y. Gurvich, and N. Barkai. Principles of cellular resource allocation revealed by condition-dependent proteome profiling. *eLife*, 6:e28034, 8 2017.

[106] T. P. Miettinen and M. Björklund. Mitochondrial Function and Cell Size: An Allometric Relationship. *Trends in Cell Biology*, xx:1–10, 2017.

[107] T. P. Miettinen and M. Bj??rklund. Cellular Allometry of Mitochondrial Functionality Establishes the Optimal Cell Size. *Developmental Cell*, 39(3):370–382, 2016.

[108] T. P. Miettinen, H. K. J. Pessa, M. J. Caldez, T. Fuhrer, M. K. Diril, and U. Sauer. Article Identification of Transcriptional and Metabolic Programs Related to Mammalian Cell Size. pages 598–608, 2014.

[109] T. Mizuguchi, G. Fudenberg, S. Mehta, J.-M. Belton, N. Taneja, H. D. Folco, P. FitzGerald, J. Dekker, L. Mirny, J. Barrowman, and S. I. S. Grewal. Cohesin-dependent globules and heterochromatin shape 3D genome architecture in S. pombe. *Nature*, 516(7531):432–5, 2014.

[110] O. Mondesert, C. H. McGowan, and P. Russell. Cig2, a B-type cyclin, promotes the onset of S in Schizosaccharomyces pombe. *Molecular and cellular biology*, 16(4):1527–33, 4 1996.

[111] S. Moreno and P. Nurse. Regulation of progression through the Gl phase of the cell cycle by the rum1 + gene. *Nature*, 367(6460):236–242, 1 1994.

[112] M. Morgan, H. Pagès, V. Obenchain, and N. Hayden. *Rsamtools: Binary alignment*

(BAM), FASTA, variant call (BCF), and tabix file import, 2016. R package version 1.24.0.

[113] J. B. Moseley, A. Mayeux, A. Paoletti, and P. Nurse. A spatial gradient coordinates cell size and mitotic entry in fission yeast. *Nature*, 459(7248):857–60, 6 2009.

[114] F. Mueller, A. Senecal, K. Tantale, H. Marie-Nelly, N. Ly, O. Collin, E. Basyuk, E. Bertrand, X. Darzacq, and C. Zimmer. FISH-quant: automatic counting of transcripts in 3D FISH images. *Nature Methods*, 10(4):277–278, 2013.

[115] F. J. Navarro, L. Weston, and P. Nurse. Global control of cell growth in fission yeast and its coordination with the cell cycle. *Current opinion in cell biology*, 24(6):833–7, 12 2012.

[116] F. R. Neumann and P. Nurse. Nuclear size control in fission yeast. *The Journal of cell biology*, 179(4):593–600, 11 2007.

[117] M. J. Nueda, S. Tarazona, and A. Conesa. Next maSigPro: updating maSigPro bioconductor package for RNA-seq time series. *Bioinformatics (Oxford, England)*, 6 2014.

[118] P. Nurse and Y. Bissett. Gene required in G1 for commitment to cell cycle and in G2 for control of mitosis in fission yeast. *Nature*, 292(5823):558–60, 8 1981.

[119] O. Padovan-Merhar, G. Nair, A. Biaesch, A. Mayer, S. Scarfone, S. Foley, A. Wu, L. Churchman, A. Singh, and A. Raj. Single Mammalian Cells Compensate for Differences in Cellular Volume and DNA Copy Number through Independent Global Transcriptional Mechanisms. *Molecular Cell*, pages 1–14, 2015.

[120] K. Z. Pan, T. E. Saunders, I. Flor-Parra, M. Howard, and F. Chang. Cortical regulation of cell size by a sizer cdr2p. *eLife*, 2014(3):1–24, 2014.

[121] O. Porrua and D. Libri. Transcription termination and the control of the transcriptome: why, where and how to stop. *Nat Rev Mol Cell Biol*, 16(3):190–202, 2015.

[122] M. Převorovský, M. Hálová, K. Abrhámová, J. Libus, and P. Folk. Workflow for Genome-Wide Determination of Pre-mRNA Splicing Efficiency from Yeast RNA-seq Data. *BioMed Research International*, 2016, 2016.

[123] R Core Team. R: A Language and Environment for Statistical Computing, 2015.

[124] C. Rallis, L. López-Maury, T. Georgescu, V. Pancaldi, and J. Bähler. Systematic screen for mutants resistant to TORC1 inhibition in fission yeast reveals genes involved in cellular ageing and growth. *Biology open*, 3(2):161–71, 1 2014.

[125] D. A. Rojas, S. Moreira-Ramos, S. Zock-Emmenthal, F. Urbina, J. Contreras-Levicoy, N. F. Käufer, and E. Maldonado. Rrn7 Protein, an RNA Polymerase I Transcription Factor, Is Required for RNA Polymerase II-dependent Transcription Directed by Core Promoters with a HomolD Box Sequence. *Journal of Biological Chemistry*, 286(30):26480–26486, 7 2011.

[126] S. Sainsbury, C. Bernecky, and P. Cramer. Structural basis of transcription initiation by RNA polymerase II. *Nature Reviews Molecular Cell Biology*, 2015.

[127] J. T. Sauls, D. Li, and S. Jun. Adder and a coarse-grained approach to cell size homeostasis in bacteria. *Current Opinion in Cell Biology*, 38:38–44, 2016.

[128] K. M. Schmoller and J. M. Skotheim. The Biosynthetic Basis of Cell Size Control. *Trends in Cell Biology*, 25(12):793–802, 12 2015.

[129] M. Scott, C. W. Gunderson, E. M. Mateescu, Z. Zhang, and T. Hwa. Interdependence of Cell Growth and Gene Expression: Origins and Consequences. *Science*, 330(6007), 2010.

[130] M. Scott and T. Hwa. Bacterial growth laws and their applications. *Current Opinion in Biotechnology*, 22(4):559–565, 2011.

[131] M. Scott, S. Klumpp, E. M. Mateescu, and T. Hwa. Emergence of robust growth laws from optimal regulation of ribosome synthesis. *Molecular systems biology*, 10(8):747, 1 2014.

[132] M. Sémon, D. Mouchiroud, and L. Duret. Relationship between gene expression and GC-content in mammals: Statistical significance and biological relevance. *Human Molecular Genetics*, 14(3):421–427, 2005.

[133] V. Shahrezaei and S. Marguerat. Connecting growth with gene expression: of noise and numbers. *Current Opinion in Microbiology*, 25:127–135, 6 2015.

[134] N. Slavov, S. Semrau, B. Budnik, N. Slavov, S. Semrau, E. Airoldi, B. Budnik, and A. V. Oudenaarden. Differential Stoichiometry among Core Ribosomal Report Differential Stoichiometry among Core Ribosomal Proteins. *CellReports*, 13(5):1–9, 2015.

[135] I. Soifer, L. Robert, and A. Amir. Single-Cell Analysis of Growth in Budding Yeast and Bacteria Reveals a Common Size Regulation Strategy. *Current Biology*, 26(3):356–361, 2 2016.

[136] I. Soriano, L. Quintales, and F. Antequera. Clustered regulatory elements at nucleosome-depleted regions punctuate a constant nucleosomal landscape in Schizosaccharomyces pombe. *BMC genomics*, 2013.

[137] K. Sugimoto-Shirasu and K. Roberts. Big it up: Endoreduplication and cell-size control in plants. *Current Opinion in Plant Biology*, 6(6):544–553, 12 2003.

[138] M. Sun, B. Schwalb, N. Pirkl, K. C. Maier, A. Schenk, H. Failmezger, A. Tresch, and P. Cramer. Global analysis of eukaryotic mRNA degradation reveals Xrn1-dependent buffering of transcript levels. *Molecular cell*, 52(1):52–62, 10 2013.

[139] M. Sun, B. Schwalb, D. Schulz, N. Pirkl, S. Etzold, L. Larivière, K. C. Maier, M. Seizl, A. Tresch, and P. Cramer. Comparative dynamic transcriptome analysis (cDTA) reveals mutual feedback between mRNA synthesis and degradation. *Genome research*, 22(7):1350–9, 7 2012.

[140] M. P. Swaffer, A. W. Jones, H. R. Flynn, A. P. Snijders, and P. Nurse. CDK Substrate Phosphorylation and Ordering the Cell Cycle. *Cell*, 167(7):1750–1761, 12 2016.

[141] S. Taheri-Araghi, S. Bradde, J. T. Sauls, N. S. Hill, P. A. Levin, J. Paulsson, M. Vergassola, and S. Jun. Cell-size control and homeostasis in bacteria. *Current Biology*, 25(3):385–391, 2 2015.

[142] H. Takahashi, J. M. McCaffery, R. a. Irizarry, and J. D. Boeke. Nucleocytosolic acetyl-coenzyme a synthetase is required for histone acetylation and global transcription. *Molecular cell*, 23(2):207–17, 7 2006.

[143] Y. Tanouchi, A. Pai, H. Park, S. Huang, R. Stamatov, N. E. Buchler, and L. You. A noisy linear map underlies oscillations in cell size and gene expression in bacteria. *Nature*, 2015.

[144] T. Toni, D. Welch, N. Strelkowa, A. Ipsen, and M. P. H. Stumpf. Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems. *J R Soc Interface*, 6(31):187–202, 2009.

[145] H. Tourrière, K. Chebli, and J. Tazi. mRNA degradation machines in eukaryotic cells. *Biochimie*, 84(8):821–37, 8 2002.

[146] C. Trapnell, A. Roberts, L. Goff, G. Pertea, D. Kim, D. R. Kelley, H. Pimentel, S. L. Salzberg, J. L. Rinn, and L. Pachter. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nature protocols*, 7(3):562–78, 3 2012.

[147] T. Trcek, J. A. Chao, D. R. Larson, H. Y. Park, D. Zenklusen, S. M. Shenoy, and R. H. Singer. Single-mRNA counting using fluorescent in situ hybridization in budding yeast. *Nature Protocols*, 7(2):408–419, feb 2012.

[148] M. Tucker, R. R. Staples, M. A. Valencia-Sanchez, D. Muhlrad, and R. Parker. Ccr4p is the catalytic subunit of a Ccr4p/Pop2p/Notp mRNA deadenylase complex in ¡i¿Saccharomyces cerevisiae¡/i¿. *The EMBO Journal*, 21(6):1427–1436, 3 2002.

[149] L. Vachon, J. Wood, E. J. Gina Kwon, A. Laderoute, K. Chatfield-Reed, J. Karagiannis, and G. Chua. Functional characterization of fission yeast transcription factors by overexpression analysis. *Genetics*, 194(4):873–884, 2013.

[150] G. Varsano, Y. Wang, M. Wu Correspondence, and M. Wu. Probing Mammalian Cell Size Homeostasis by Channel-Assisted Cell Reshaping. *Cell Reports*, 20:397–410, 2017.

[151] Y. Voichek, R. Bar-Ziv, and N. Barkai. A role for Rtt109 in buffering gene-dosage imbalance during DNA replication. *Nucleus*, 7(4):00–00, 2016.

[152] Y. Voichek, R. Bar-Ziv, and N. Barkai. Expression homeostasis during DNA replication. *Science*, 351(6277):1087–1090, 2016.

[153] M. C. Wahl, C. L. Will, and R. Lührmann. The Spliceosome: Design Principles of a Dynamic RNP Machine. *Cell*, 136(4):701–718, 2 2009.

[154] G. R. Warnes, B. Bolker, L. Bonebakker, R. Gentleman, W. H. A. Liaw, T. Lumley, M. Maechler, A. Magnusson, S. Moeller, M. Schwartz, and B. Venables. *gplots: Various R Programming Tools for Plotting Data*, 2016. R package version 3.0.1.

[155] S. C. Weber and C. P. Brangwynne. Inverse size scaling of the nucleolus by a concentration-dependent phase transition. *Current Biology*, 25(5):641–646, 2015.

[156] T. Wei and V. Simko. *corrplot: Visualization of a Correlation Matrix*, 2016. R package version 0.77.

[157] R. L. Weiss, J. R. Kukora, and J. Adams. The relationship between enzyme activity, cell geometry, and fitness in Saccharomyces cerevisiae. *Proceedings of the National Academy of Sciences of the United States of America*, 72(3):794–8, 3 1975.

[158] A. Y. Weiße, D. a. Oyarzún, V. Danos, and P. S. Swain. Mechanistic links between cellular trade-offs, gene expression, and growth. *Proceedings of the National Academy of Sciences*, page 201416533, 2015.

[159] S. West, N. Gromak, and N. J. Proudfoot. Human 5 3 exonuclease Xrn2 promotes transcription termination at co-transcriptional cleavage sites. *Nature*, 432(7016):522–525, 11 2004.

[160] S. Whitehall, P. Stacey, K. Dawson, and N. Jones. Cell cycle-regulated transcription in fission yeast: Cdc10-Res protein interactions during the cell cycle and domains required for regulated transcription. *Molecular biology of the cell*, 10(11):3705–3715, 1999.

[161] H. Wickham. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2009.

[162] H. Wickham. The Split-Apply-Combine Strategy for Data. *Journal of Statistical Software*, 40(1):1–29, 2011.

[163] H. Wickham and R. Francois. *dplyr: A Grammar of Data Manipulation*, 2016. R package version 0.5.0.

[164] E. Wood and P. Nurse. Pom1 and cell size homeostasis in fission yeast. *Cell Cycle*, 12(19):3228–3236, 2013.

[165] E. Wood and P. Nurse. Sizing Up to Divide: Mitotic Cell Size Control in Fission Yeast. 2014.

[166] 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*, 40(D1):695–699, 2012.

[167] C.-Y. Wu, P. A. Rolfe, D. K. Gifford, and G. R. Fink. Control of Transcription by Cell Size. *PLoS Biology*, 8(11):e1000523, 11 2010.

[168] H. Yamano, J. Gannon, and T. Hunt. The role of proteolysis in cell cycle progression in Schizosaccharomyces pombe. *The EMBO journal*, 15(19):5268–79, 10 1996.

[169] G. Yu, L.-G. Wang, Y. Han, and Q.-Y. He. clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. *OMICS: A Journal of Integrative Biology*, 16(5):284–287, 5 2012.

[170] A. Zeileis and G. Grothendieck. zoo: S3 Infrastructure for Regular and Irregular Time Series. *Journal of Statistical Software*, 30(6):1–27, 2005.

[171] M. Zhang, L. Galdieri, and A. Vancura. The Yeast AMPK Homolog SNF1 Regulates Acetyl Coenzyme A Homeostasis and Histone Acetylation. *Molecular and Cellular Biology*, 33(23):4701–4717, 12 2013.

[172] Z. Zhang, J. Fu, and D. S. Gilmour. CTD-dependent dismantling of the RNA polymerase II elongation complex by the pre-mRNA 3'-end processing factor, Pcf11. *Genes & Development*, 19(13):1572–1580, 7 2005.

[173] J. Zhurinsky, K. Leonhard, S. Watt, S. Marguerat, J. Bähler, and P. Nurse. A coordinated global control over cellular transcription. *Current Biology*, 20(22):2010–2015, 2010.