High-Throughput Chronological Lifespan Screening of the Fission Yeast Deletion Library Using Barcode Sequencing

Catalina-Andreea Romila

A dissertation submitted in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

of

University College London.

Department of Genetics, Evolution and Environment

Institute of Healthy Ageing

University College London

November 27, 2019

I, Catalina-Andreea Romila, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the work.

Abstract

Ageing is associated with the development of several chronic illnesses, including cardiovascular diseases, diabetes and cancer. To understand the genetic components driving cellular ageing in higher organisms, like ourselves, we study simple eukaryotic model systems which are more accessible and easier to manipulate than higher eukaryotes. This is possible due to the remarkably conserved ageing mechanisms that occurs between species. Here, we employ fission yeast one of the simplest eukaryotic model organisms to study cellular ageing. In this work, we decoded the fission yeast deletion collection using our in-house developed pipeline, developed an improved version of Bar-seq along with a custom-developed analysis pipeline, determined a method for high-quality RNA extraction and RNA-seq from long-term quiescent yeast cells, and finally, performed a high-throughput Bar-seq screen to profile the chronological lifespan of our decoded strains. We describe barcode decoding of 94% of the gene deletions; validation of our Bar-seq developed method; identification of ncRNAs as elements important for the cellular quiescence maintenance; Bar-seq screening of the competitively grown decoded strains which identified several long-lived and short-lived mutants following glucose-starvation and cellular culture re-growth; and also, validation of the top hits using isogenic cell cultures revealing eight novel gene deletions important for the early life maintenance, as well as ten novel gene deletion mutants with pro-ageing effects.

Overall, in addition to providing rich datasets, we describe several high-throughput methods that can be used for future genome-wide studies, whereby the complementarity of genomics and transcriptomics can be coupled together to further advance our understanding of the genetic factors underpinning cellular ageing in humans.

Impact Statement

This work describes new research methods and provides tools that can be used to obtain high-quality data in a timely and cost-effective manner. Our characterisation of the fission yeast deletion library strains is a valuable genetic resource allowing for the high-throughput study of gene deletions to uncover their function. The data obtained from using our developed tools has the potential to foster new research projects and/or support future grant applications. The work derived from this thesis is to be communicated internationally through a research publication as well as through the local deposition of the thesis on the central public library.

The data obtained from our work helped us uncover new gene deletions that impact yeast lifespan. Due to the translation potential from yeast to humans, our data has clinical relevance and may facilitate therapeutic developments to promote healthy ageing in humans. Because the gene deletions discovered from our work have been shown to affect lifespan, the results are also relevant for public health and as supporting material for policy making.

Finally, the work presented in this thesis can benefit not only individuals but entire enterprises dedicated to the development of anti-ageing therapies in humans, thus acting as a bridge fostering collaborations between academic and non-academic institutions.

Acknowledgements

First of all, I would like to express my gratitude to my supervisor Jürg Bähler for the opportunity to pursue my research interests and for his help and advice. I would also like to thank Michal Malecki for his help, invaluable discussions, answers to my questions and extremely valuable thesis feedback, StJohn Townsend and Stephan Kamrad for their collaboration on the development of tools used for data analysis and for their help in general. Many thanks to all the Bähler lab group members for making my PhD journey a great and fun experience with plenty of cake and laughter!

I would also like to extend my acknowledgements to people closer to my heart. My parents for the opportunity to receive my secondary education in England. My sister and my brothers for their help and support in life. Daniel for his unconditional support throughout my PhD endeavour and in everyday life; for his friendship, love and care and for reading my numerous thesis drafts and providing me with constructive and valuable feedback. I am also grateful to my dear friend Mavji Patel who has guided me from high-school through to university and thankfully to this day as well. He has also read my thesis and contributed with great feedback.

Last but not least, many thanks to Nadine Mogford, the UCL LIDo programme representative for organising many social events and to my LIDo friends for making my PhD a fun and enjoyable experience. I would also like to acknowledge BBSRC for funding my project.

Resources Derived From This Thesis

Barcode-Sequencing Screen of Fission Yeast Deletion Library for Altered Chronological Lifespan during Stationary Phase - manuscript in preparation.

The following tools and resources will be made publicly available upon publishing of our data:

Analytical tools

- BarSeqTools, the package used to decode the gene deletion strains;
- A gene browser developed to visualise the gene deletions;
- Barcount, the package used to analyse Bar-seq data.

Technical resources

- Improved Bar-seq method;
- Established RNA isolation method from aged yeast cells;

General resources

- List of all the decoded fission yeast deletion strains;
- List of the long-lived gene deletion mutants identified with Bar-seq.

List of Abbreviations

Bar-seq Barcode sequencing.

bp base pair.

C. elegans Caenorhabditis elegans.

CFUs Colony Forming Units.

CLS Chronological Lifespan.

D. melanogaster Drosophila melanogaster.

DGE Differential Gene Expression.

DNA Deoxyribonucleic Acid.

dntag downtag barcode.

DR Dietary Restriction.

e.g. exempli gratia, for example.

EMM Edinburgh Minimal Medium.

EtOH Ethanol.

FC Fold change.

FDR False Discovery Rate.

gDNA Genomic Deoxyribonucleic acid.

GO Gene Ontology.

GSEA Gene Set Enrichment Analysis.

IGF Insulin Growth Factor.

kb kilo base.

LiCl Lithium Chloride.

LTRs Long Terminal Repeats.

Mb Mega base.

min minute.

mRNA messenger RNA.

ncRNA non-coding Ribonucleic Acid.

nt nucleotide.

o/n overnight.

OD Optical Density.

ORF Open Reading Frame.

PCR Polymerase Chain Reaction.

qPCR quantitative PCR.

qRT-PCR quantitative Real Time PCR.

RIN Ribonucleic Acid Integrity Number.

RLS Replicative Lifespan.

RNA Ribonucleic Acid.

RNA-seq RNA sequencing.

RNAi RNA interference.

rpm rotations per minute.

rRNA ribosomal Ribonucleic Acid.

S. cerevisiae Schizosaccharomyces cerevisiae.

S. japonicus Schizosaccharomyces japonicus.

S. octosporus Schizosaccharomyces octosporus.

S. pombe Schizosaccharomyces pombe.

sec second.

snorRNA small nucleolar Ribonucleic Acid.

snRNA small nuclear Ribonucleic Acid.

TOR Target Of Rapamycin.

tRNA transfer Ribonucleic Acid.

UMIS Unique Molecular Identifiers.

uptag uptag barcode.

UTR Untranslated Region.

V Volts.

wt wild-type.

YES Yeast Extract + supplements.

Contents

Abstract **3**

Impact Statement **4**

Acknowledgements **5**

Resources Derived From This Thesis **6**

List of Abbreviations **9**

1 Introduction **20**

1.1 Ageing: A general perspective 20

1.2 A comparison of yeasts 23

1.3 Yeast: A model for human ageing 24

1.3.1 *S. pombe* biology 27

1.3.2 *S. pombe* genome 28

1.4 Premise of the project 30

2 Materials & Methods **33**

2.1 General Techniques 33

2.1.1 Cell media & culture 33

2.1.2 Colony Forming Units (CFUs) 34

2.1.3 Sample collection & storage 34

2.1.4 Genomic DNA extraction 34

2.1.5 Genomic DNA quantification & purification 35

2.1.6 qRT-PCR . . . 35
2.1.7 Gel electrophoresis . . . 36
2.1.8 BioAnalyzer . . . 36
2.2 Protocols . . . 36
2.2.1 Deletion library strains pool generation . . . 36
2.2.2 RNA isolation from quiescent cells . . . 38
2.2.3 Library preparation . . . 38
2.2.3.1 Barcode decoding of the deletion library strains . 38
2.2.3.2 Colony PCR to check for mutant strain gene deletions . . . 42
2.2.3.3 Bar-seq . . . 43
2.2.3.4 RNA-seq . . . 46
2.3 Data Analysis . . . 48
2.3.1 Barcode decoding of the deletion library strains . . . 48
2.3.2 Bar-seq strain fitness comparison to colony screen . . . 50
2.3.3 RNA-seq of cellular quiescent cells . . . 52
2.3.4 Chronological ageing of competitively grown decoded strains 52
2.3.5 Isogenic strain growth to validate CLS lifespan . . . 53
2.4 Data Repository . . . 56

3 Barcode decoding of the deletion library strains 57
3.0.1 Background . . . 57
3.0.2 Experimental design . . . 59
3.0.3 Decoded barcodes show a strong correlation . . . 60
3.0.4 We successfully decoded 94% of the library gene deletions . 65
3.0.5 Not all verified mutants showed successful gene deletions . 66
3.0.6 Our barcodes matched to previously decoded barcodes . . . 70

4 Bar-seq strain fitness comparison to colony screen 72
4.0.1 Background . . . 72
4.0.2 Experimental design . . . 74

4.0.3 Bar-seq detected 94% of the decoded genes 76
4.0.4 Bar-seq identified 112 mitochondrial protein-coding genes . 77
4.0.5 Bar-seq out-performed the colony screen 78

5 RNA-seq of cellular quiescent cells 82
5.0.1 Background 82
5.0.2 Experimental design 82
5.0.3 The new RNA isolation protocol preserves RNA 84
5.0.4 Quiescent cells gene enrichment shows ncRNA regulation . 85

6 Chronological ageing of competitively grown decoded strains 89
6.0.1 Background 89
6.0.2 Experimental design 90
6.0.3 Chronological ageing with Bar-seq requires culture re-growth 92
6.0.4 Re-growth data analysis identified more short-lived than long-lived mutants 97
6.0.5 Enrichment of differentially expressed genes suggests the identification of age-relevant genes 100

7 Isogenic strain growth to validate CLS lifespan 104
7.0.1 Background 104
7.0.2 Experimental design 105
7.0.3 Of the Bar-seq validated mutants, 90% showed the expected lifespan observed from the competitive pool growth 106
7.0.4 Only 25% of day 10 and day 12 long-lived mutants showed the expected lifespan observed from the competitive pool growth 108
7.0.5 Validation of the top 10 long-lived mutants distinguishes several novel longevity genes 108

8 Discussion 111
8.1 The fission yeast library decoded genes 111

8.1.1 Comparison of our decoded genes to previous versions shows that our characterisation was the most successful . . . 111
8.1.2 One-fifth of the undecoded genes represent genes with multiple copies and most are likely unsuccessful gene deletions 112
8.2 Bar-seq & RNA-seq . . . 113
8.2.1 Genome-wide analysis with Bar-seq distinguishes the method as a powerful genetic tool . . . 113
8.2.2 Transcriptomic analysis renders regulation of ncRNAs as important elements required for quiescence maintenance . . 114
8.2.3 Bar-seq whole-genome analyses & RNA-seq transcriptomic analyses are highly complementary . . . 114
8.3 Chronological lifespan of competitively grown decoded strains . . . 115
8.3.1 Comparison of long-lived mutants to published data . . . 115
8.3.2 Comparison of short-lived mutants to published data . . . 117
8.4 Validation of top mutants . . . 119
8.4.1 Validation of short-lived mutants uncovered eight novel genes important for the maintenance of early life . . . 119
8.4.2 Most of the validated long-lived mutants that saturated the re-growth cultures are likely cooperative cells . . . 120
8.4.3 The validation of the long-lived mutants uncovered ten novel pro-ageing genes . . . 121
8.5 Future Directions . . . 123

9 Concluding Remarks **124**

Appendices **125**

A Deletion library barcode strain decoding analysis pipeline **125**
A.1 Analysis . . . 125
A.2 Script . . . 131

B Bar-seq analysis pipeline **136**

B.1 Script 136

C CLS lifespan validation script **138**

C.1 Script 138

D List of the long-lived mutants **142**

References **153**

List of Figures

1.1 The nine hallmarks of ageing. 21
1.2 The conserved nutrient signaling pathway regulating longevity. . . . 22
1.3 Chronological lifespan in fission yeast. 26
1.4 *S. pombe* meiotic *vs* mitotic cycle. 28
1.5 *S. pombe* chromosomal architecture. 28

2.1 CFUs assay. 34
2.2 Deletion library strains pool preparation. 37
2.3 The new RNA isolation method. 38
2.4 Characterisation library preparation. 39
2.5 Barcode decoding reads structure. 41
2.6 Mutant gene deletion validation by colony PCR. 42
2.7 Bar-seq library preparation. 44
2.8 Bar-seq read assembly: a dntag example. 45
2.9 RNA library preparation. 46
2.10 Barcode decoding analysis overview. 48
2.11 Barcode decoding pipeline. 49
2.12 Barcount workflow. 50
2.13 Bar-seq analysis. 51
2.14 RNA-seq analysis. 52
2.15 Example of the maximum likelihood estimation of the number of viable cells of a mutant 55

3.1 Cassette construct used to delete the non-essential genes. 58

3.2 Barcode decoding experimental set-up. 59
3.3 Dntag samples show higher read loss than uptag samples. 61
3.4 The exclusion of sub-optimal gene-barcode pairs frequency. 62
3.5 Automatic high-confidence gene-barcode pairs selection. 63
3.6 Uptag and dntag barcode decoded genes show a strong correlation. . 64
3.7 Barcode decoding of 94% of the library gene deletions. 65
3.8 Decoded and undecoded genes show no fitness correlation. 66
3.9 Undecoded genes are neither telomeric nor LTR genes. 67
3.10 Possible mutant position bias for the undecoded but not for the duplicate genes. 68
3.11 PCR analysis of the validated deletion strains hints at tenuous deletions. 69
3.12 The barcode decoded sequences vary in length. 71

4.1 Bar-seq profiling of the decoded strains grown competitively in a pool. 73
4.2 Bar-seq proof-of-principle experimental set-up. 74
4.3 Uptag barcodes match to the reference barcodes with higher efficiency than the dntag barcodes. 76
4.4 Uptag and dntag gene counts show a good correlation. 77
4.5 Sample barcodes and gene counts show good correlations. 78
4.6 Bar-seq identified 112 mitochondrial protein-encoding genes. 79
4.7 Bar-seq and colony screens show good correlation. 79
4.8 Bar-seq vs colony screen of the mitochondrial protein-encoding gene enrichment. 80
4.9 Bar-seq was more efficient at identifying respiratory defective mutants than the colony screen. 81

5.1 Quiescence cells culture set-up. 83
5.2 The new RNA protocol preserves RNA integrity. 84

5.3 Samples treated with the new protocol show higher cDNA synthesis efficiency. 85
5.4 Differential gene expression of long-term quiescent cells. 86
5.5 Ribosome biogenesis and trans-membrane transport are the main RNA-seq GO-slim signatures. 87
5.6 Metabolic pathway genes seem to play a role in long-term cellular quiescence. 88

6.1 Chronological ageing experimental set-up. 90
6.2 Competitive strain pool growth shows reduced cell viability. 92
6.3 Uptag reads matched more efficiently to the reference barcode database than the dntag reads. 94
6.4 DNA persistence following cell death impacts the sequencing depth. 95
6.5 Over time the growth samples show no change in gene counts. . . . 96
6.6 Over time the re-growth samples show a decrease in gene counts. . . 96
6.7 Chronologically aged re-growth samples show strong correlations but only up to day 6. 97
6.8 The relative difference method was more suitable over the time course analysis. 99
6.9 Long-lived GO-term enrichment. 100
6.10 Short-lived genes GO-term enrichment. 101
6.11 Long-lived and short-lived genes KEGG enrichment. 101
6.12 The long-lived deleted genes have fewer KEGG annotations than the short-lived deleted genes. 103

7.1 Mutant CLS validation experimental set-up. 105
7.2 All validated control and Bar-seq short-lived mutants display the expected lifespan phenotype. 107
7.3 Validated Bar-seq top and unexplored long-lived mutants show 100% and 75% lifespan recapitulation to the competitively grown mutants. 107

7.4 Most of the validated day 10 and day 12 long-lived mutants show lifespans opposite to what was observed from the competitive pool growth. 108
7.5 Comparison of long-lived mutants with top lifespans revealed novel longevity genes. 109

A.1 Gene-browser example of a fully decoded gene. 129
A.2 Gene browser example of duplicate genes. 129
A.3 Gene browser example of dubious genes. 130
A.4 Gene browser example of genes with no identifiable barcodes. . . . 130

List of Tables

2.1 The list of primers used for decoding the deletion strains. 40
2.2 The list of primers used to verify mutant gene deletions. 43
2.3 Bar-seq primers. 44

3.1 Barcode decoding sample names and sequencing depth per barcode. 60

4.1 Bar-seq strain fitness comparison sample names and sequencing depth per barcode. 75

5.1 RNA-seq quiescent sample names and sequencing depth per barcode. 83

6.1 Chronological lifespan sample names and sequencing depth per barcode. 91
6.2 The number of cells used to inoculate the re-growth cultures. 98

8.1 Long-lived CLS gene overlap with the previously described Bioneer library screens. 115

D.1 Bar-seq long-lived gene deletion mutants and the top hits CLS validation. 143

Chapter 1

Introduction

1.1 Ageing: A general perspective

Ageing is a natural biological process which through its multi-factorial blueprint leads to cellular and molecular damage accumulation over time, ultimately increasing the risk of death (Gems and Partridge, 2013; Niccoli and Partridge, 2012). Consequently, ageing is the main risk factor for developing many age-related illnesses, including cardiovascular diseases, diabetes and cancer (Brooks and Mias, 2019; López-Otín et al., 2013). This intrinsic molecular clock intrigued humanity for many decades, especially the scientific community, since this seemingly natural, yet damaging process affects most living organisms including humans (Klass, 1983, 1977). To date several universal hallmarks of ageing have been described including, altered intercellular communication, genomic instability, telomere attrition, epigenetic alterations, loss of proteostasis, deregulated nutrient sensing, mitochondrial dysfunction, cellular senescence and stem cell exhaustion (López-Otín et al., 2013), also depicted in (**Figure 1.1**). The expected rise in life expectancy throughout the world (Oeppen and Vaupel, 2002) led to the extensive characterisation of some of these ageing hallmarks (Barbosa et al., 2019; Rebelo-Marques et al., 2018). In an effort to alleviate some of the expected socio-economic challenges posed by an ever-increasing elderly population, ageing research laboratories devoted to developing therapeutics aimed at improving the life quality at old age are on the rise (McCurry, 2015; Stratton, 2013).

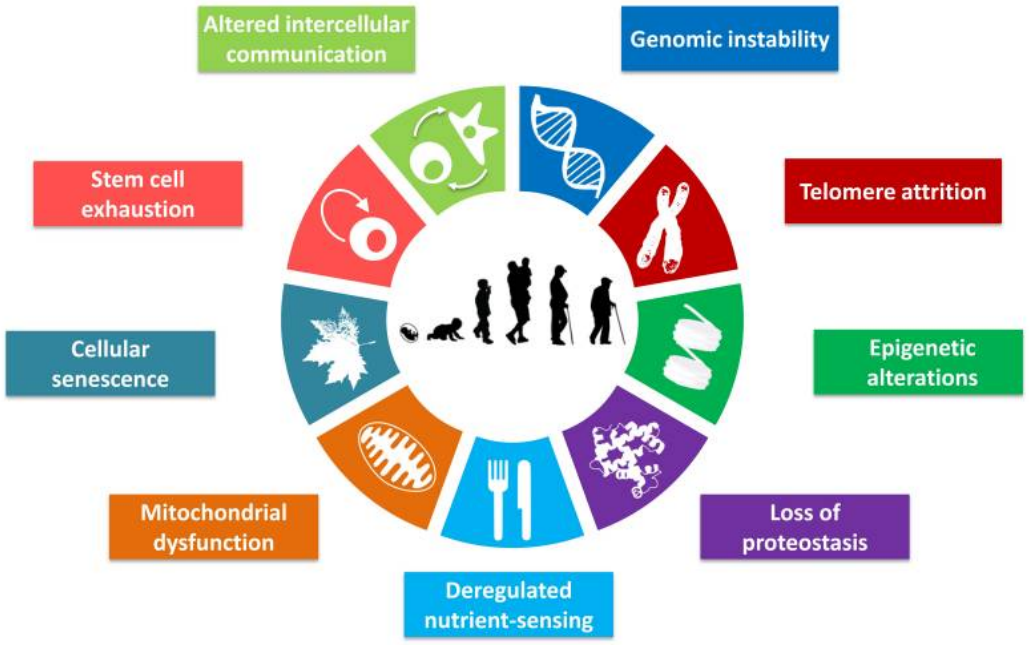

Figure 1.1: The nine hallmarks of ageing.
Summary of the current hallmarks of ageing; altered intercellular communication, genomic instability, telomere attrition, epigenetic alterations, loss of proteostasis, deregulated nutrient sensing, mitochondrial dysfunction, cellular senescence and stem cell exhaustion. Figure adapted from (López-Otín et al., 2013).

Nevertheless, several studies using established laboratory model organisms, such as yeast, nematodes, flies, mice and monkeys corroborated out current understanding of ageing being plastic (Gems and Partridge, 2013; de Magalhães et al., 2012; Fontana et al., 2010). The malleability of ageing allows for human interventions to help promote healthy ageing. In fact, simple genetic interventions, such as gene deletion/over-expression and environmental manipulations, including dietary restriction (DR) without malnutrition have shown to greatly influence both lifespan and healthspan by delaying the onset of the age-associated pathologies, thus compacting morbidity (Gems and Partridge, 2013; Fontana et al., 2010; Kenyon, 2010). DR via the nutrient sensing pathway is one of the most studied mechanisms of longevity, and highly conserved across a diverse range of species, including yeast, worms, flies and mammals (**Figure 1.2**). Other than yeast which use the serine/threonine kinases Tor, Sch9 and PKA to sense nutrients (Kaeberlein et al., 2005), nutrient sensing via the Insulin Growth Factor (IGF)-1like receptor pathway is common

to *Caenorhabditis elegans* (*C. elegans*), the fruit fly *Drosophila melanogaster* (*D. melanogaster*), rodents and primates, collectively referred to as mammals (Fontana et al., 2010). Target of Rapamycin (TOR), however, is universal across all of these highly conserved nutrient sensing pathways, thus inhibition of TOR as a key regulator of longevity is a means for human intervention to promote healthy ageing (Smith et al., 2016; Kenyon, 2010; Partridge, 2010). Pharmacological interventions studies on yeast identified three TOR inhibiting drugs; resveratrol, rapamycin and spermidine (Kaeberlein, 2010). The efficacy of these drugs toward improved lifespan and healthspan has also been shown in mammalian systems (Madeo et al., 2018; Bhullar and Hubbard, 2015; Neff et al., 2013). However, their beneficial effects in humans remains elusive. Nonetheless, the use of yeasts as simple model organisms to help uncover common drug targets that may one day work in humans should not be dismissed but encouraged.

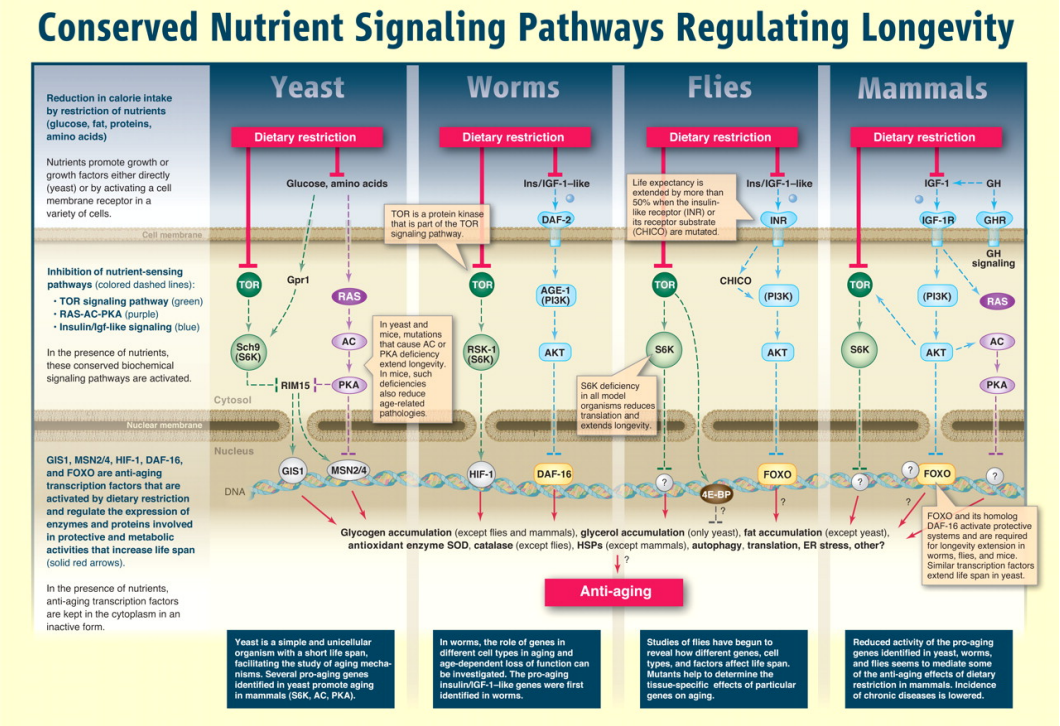

Figure 1.2: The conserved nutrient signaling pathway regulating longevity.
The nutrient signaling pathways known to promote longevity from yeast to mammals. Figure adapted from (Fontana et al., 2010).

1.2 A comparison of yeasts

Yeasts are single-celled eukaryotic fungi that grow either by budding or fission without forming fruiting bodies (Buzzini et al., 2017). Yeasts can grow on a wide range of carbon sources though glucose is their preferred source where cells grow by fermentation (Malecki et al., 2016; Turcotte et al., 2010). Due to their ephemeral habitat, yeasts can be found in rotten fruit, plant roots and exudates, nectar, sugar cane etc (Jeffares, 2018; Buzzini et al., 2017). Generally, yeasts form symbiotic relationships with their host but are equally capable of living on their own as well (Bähler and Wood, 2006; Yanagida, 2002). To date there are over 500 known yeast species that are grouped into two taxa: *ascomycetes* and *basidiomycetes* (Sipiczki, 2000). *Ascomycetes* is by far the largest taxa encompassing, *archaeascomycetes* (fission yeasts), *hemiascomycetes* (budding yeasts) and *euascomycetes* (filamentous yeasts) (Forsburg and Rhind, 2006). Owed to their unique characteristics, yeasts can be distinguished from other eukaryotes relatively easily. Fission yeasts, for example, can be identified by several distinct features, such as closed mitosis, cell wall and spore formation. The genomic diversity with classes of protoplast fusion (Sipiczki et al., 1982), phenotypic features (Bridge and May, 1984), DNA re-association and taxonomic characteristics (Martini, 1991) revealed three fission yeasts in total: *Schizosaccharomyces pombe (S. pombe), Schizosaccharomyces japonicus (S. japonicus)* and *Schizosaccharomyces octosporus (S. octosporus)*. However, *S. pombe* is by far the most characterised fission yeast (Sipiczki, 2000).

In contrast to the budding yeast, *Saccharomyces cerevisiae* (*S. cerevisiae*) also known as baker's yeast which was first introduced to science in the late 1800s (Duina et al., 2014), *S. pombe* research began in the late 1940s (Fantes and Hoffman, 2016). Their genome assembly distinguishes *S. cerevisiae* with approximately 6,000 genes (Goffeau et al., 1996) and *S. pombe* with over 5,000 protein-coding genes (Wood et al., 2002). Though despite the similarity in the number of genes, *S. pombe* diverged from budding yeast over 350 million years ago and shares more features with higher eukaryotes than its counterpart (Hoffman et al., 2015). Moreover, compared to *S. cerevisiae* which shows gene evolution and some degree of retained similarity

to mammalian systems (Botstein et al., 1997), approximately 67% of the *S. pombe* genes have human orthologs, thus *S. pombe* is a more advanced yeast model. In addition, most of its molecular mechanisms resemble that of humans (Lock et al., 2019; Wood et al., 2012), including, the mode of cell division and the regulation of the cell cycle (Nurse, 1990), mRNA splicing machinery (Aravind et al., 2000), RNA-interference proteins (Bühler et al., 2008), centromere proteins (Irelan et al., 2001), mitochondria inheritance (Chiron et al., 2007) as well as stress response pathways (Vivancos et al., 2006). Therefore, as an ageing model *S. pombe* not only complements budding yeast research but provides unique insights into yeast cellular ageing (Fruhmann et al., 2017).

1.3 Yeast: A model for human ageing

Much of our current understanding of ageing is owed to the contributions made from laboratory studies using simple eukaryotic model organisms (Fontana et al., 2010). While having distinct features, the age-associated mechanisms, pathways and diseases are remarkably conserved across the animal kingdom (López-Otín et al., 2013; Fontana et al., 2010; Kenyon, 2010; Tissenbaum and Guarente, 2002). For example, yeasts are simple eukaryotic model organisms and due to their short lifespan, typically of a few days, make an excellent model to study cellular ageing. Also, the conserved gene function between yeast and humans enables for translation potential from yeast ageing research to humans.

Yeasts are simple, yet powerful models, the genetic tractability of which contributed to establishing the organism as a reliable ageing system (Zimmermann et al., 2018; Janssens and Veenhoff, 2016; Longo et al., 2012; Roux et al., 2010; Kaeberlein et al., 2007; Fabrizio and Longo, 2003; Longo and Finch, 2003). Ageing in yeast can be performed either via replicative lifespan (RLS) studies, or chronological lifespan (CLS) studies (Longo et al., 2012). While RLS is a measure of the number of cellular mitotic divisions, and thus a model for dividing cells (Longo et al., 2012; Roux et al., 2010; Steinkraus et al., 2008; Barker and Walmsley, 1999), CLS measures the viability of the non-dividing cells in the stationary phase once nu-

trients have been depleted, and acts a post-mitotic model to study the ageing of non-proliferating cells, such as fibroblasts (Longo et al., 2012; Fabrizio and Longo, 2003). Although until recently, the preferred yeast for ageing studies was *S. cerevisiae*, *S. pombe* is fast emerging as a complementary and reliable system (Lin and Austriaco, 2014; Roux et al., 2010, 2009, 2006). For example, due to the asymmetric cell division RLS studies in *S. cerevisiae* have been well-established (Jo et al., 2015; Wasko and Kaeberlein, 2014; Henderson and Gottschling, 2008). On the other hand, RLS studies in organisms such as fission yeast due to symmetric division (Longo et al., 2012) are more difficult to undertake. As a result, whether *S. pombe* undergoes replicative ageing is currently a topic of controversy. While RLS in fission yeast has been evidenced in the late 90s (Barker and Walmsley, 1999), two recent studies concluded the lack of RLS in fission yeast (Spivey et al., 2016; Nakaoka and Wakamoto, 2017), despite several publications showing that the two *S. pombe* symmetrical daughter cells do age (Roux et al., 2010; Minois et al., 2006) as a result of asymmetrical cytokinesis which leads to unequal partitioning of the maternal damaged proteins (Erjavec et al., 2008). On the contrary, CLS studies in *S. pombe* have always prevailed (Roux et al., 2009). The studies are typically performed by growing the cell population cultures to the stationary phase and once the cells have been depleted of nutrients, cell viability is measured using the standard colony forming units (CFUs) method (see **Subsection 2.1.2**). An example of yeast cellular growth followed by chronological lifespan measurement is depicted in (**Figure 1.3**).

The non-dividing stationary-phase model can be used to study two distinct cell-cycle arrest states, both of which are involved in longevity, hence are important and relevant to ageing research (Masuda et al., 2016; Yao, 2014; Aranda-Anzaldo, 2012; Yanagida, 2009; Su et al., 1996). One cell-cycle arrest state is the differentiated G0-like state where cells, such as the fibroblasts resemble quiescence, a state with minimal metabolic activity but retained proliferative ability (Roche et al., 2017; Yao, 2014; Aranda-Anzaldo, 2012). The other cell-cycle arrest is the G2-phase where cells age in a manner analogous to post-mitotic cells, including neu-

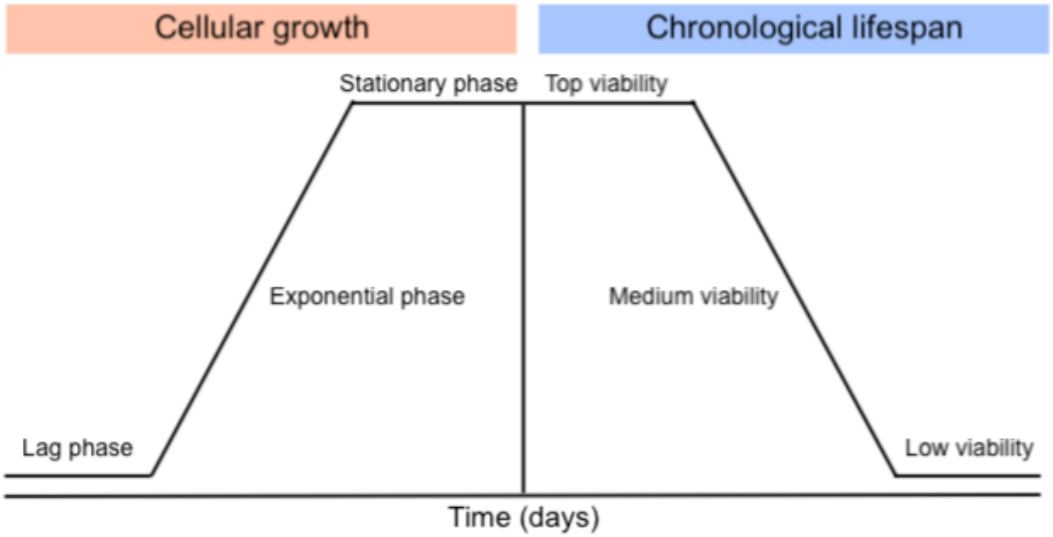

Figure 1.3: Chronological lifespan in fission yeast.
During cellular growth (on the left) cells progress through the lag phase, exponential and stationary phase. Chronological lifespan (on the right) measurements start once cellular growth ceases, concurrent with the cells being maintained in the stationary phase. Viability is measured from colony forming units (CFUs) where the initial time point with 100% viability is the reference time point. The viability of other time points is calculated as the relative difference to the reference time point.

rons (Aranda-Anzaldo, 2012) and quiescent stem cells (Roche et al., 2017). Both of these states can be induced in a manner dependent on the environmental cues. For example, the G0-like cell-cycle arrest can be induced by withdrawing the nitrogen from the media (Yanagida, 2009; Su et al., 1996), while glucose removal from the media triggers cellular arrest at the G2-phase of the cell-cycle (Masuda et al., 2016; Aranda-Anzaldo, 2012).

However, in contrast to the glucose-starvation model where cells typically survive for one week (Roux et al., 2006, 2009), quiescent cells can survive for several months and the process is also reversible, as upon restoring the nitrogen level, the cells revert from quiescence and re-enter the cell cycle (Marguerat et al., 2012; Takeda et al., 2010; Yanagida, 2009; Mochida and Yanagida, 2005). Therefore, beyond simply being a well-established model organism, *S. pombe* also allows for a plethora of ageing studies to be performed, providing a relatively inexpensive way to study the interplay between genes and environment, and the effect on lifespan.

1.3.1 *S. pombe* biology

The scientific history of fission yeast dates back to 1893 when the first strain was isolated from the East African millet beer (Paul, 1893). Because of this, *S. pombe* was commonly believed to have also originated from Africa though a recent study found no evidence to support this (Jeffares, 2018). However, *S. pombe* taxonomy is more clear as within the genus name '*Schizosaccharomyces*', '*Schizo...*' captures the organism mode of division, '*...saccharomyces*' identifies it as a yeast and the species name identifies the organism isolation origin as '*pombe*' translates into the Swahili word for beer (Sipiczki, 2000).

Morphologically, the cells are cylindrical in shape, typically measuring between 3-4 μm in diameter and between 7-15 μm in length (Mitchison and Nurse, 1985). Depending on the mating-type genes inherited from the mat cassette (Pc and Mc), three mating cell types can be distinguished: h^+, h^- and h^{90} cells (Forsburg and Rhind, 2006). Cells that carry both genes are known as homothallic (h^{90}) cells, whereas cells that carry only one gene type, either Pc or Mc, h^+ and h^- respectively, are known as heterothallic cells (Nielsen, 2008; Forsburg and Rhind, 2006; Mochida and Yanagida, 2005; Gómez and Forsburg, 2004). The cells can also be distinguished by the DNA copy number (Forsburg and Rhind, 2006), where haploid cells (e.g. one DNA copy) are more common than diploid cells (e.g. two DNA copies) (Nielsen, 2008; Forsburg and Rhind, 2006).

Cellular division by binary fission which commonly occurs every 2-4 hours, whereby the cells grow at the cell tips and split by medial fission to generate two symmetrical daughter cells (Forsburg and Rhind, 2006; Mitchison and Nurse, 1985), can easily be influenced by the cell mating-type, media type and ploidy number. The life cycle and thus the mode of division of *S. pombe* can occur either by mitosis or meiosis as shown in (**Figure 1.4**).

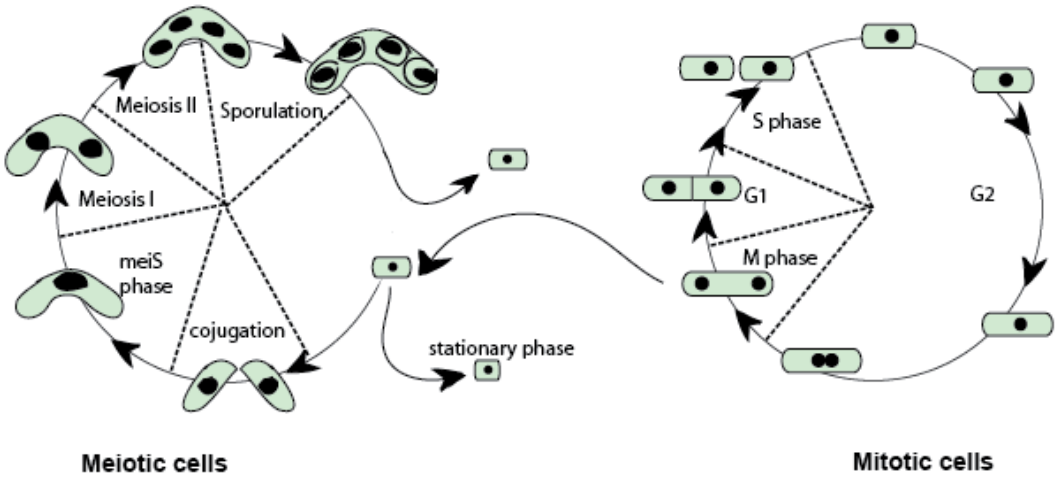

Figure 1.4: S. pombe meiotic vs mitotic cycle.
Diploid or meiotic cells undergo sexual conjugation producing four spores. The spores germinate forming new diploid cells that can re-enter and re-set the cycle. Haploid or mitotic cells grow in a vegetative manner until growth ceases, re-setting the cycle. Figure adapted from (Forsburg, 2011a).

The mitotic cycle starts with haploid cells which grow vegetatively during the normal cell cycle phase until growth ceases causing the mitotic cycle to re-set (Piel and Tran, 2009). In contrast, the meiotic cycle is triggered by nitrogen depletion in the media, where in the presence of different mating cell types, diploid cells undergo sexual differentiation by conjugation. Upon successful conjugation, a zygote is produced that splits into four genetically distinct spores, which in turn can sporulate, hence re-setting the meiotic cycle (Hoffman et al., 2015; Nielsen, 2008; Forsburg and Rhind, 2006).

1.3.2 *S. pombe* genome

Excluding the mitochondrial chromosome of 20 kilo bases (kb), the genome comprises of approximately 12.6 mega bases (Mb) compacted into three distinct chromosomes (Wood et al., 2002; Lang et al., 1987; Smith et al., 1987) as shown in (**Figure 1.5**).

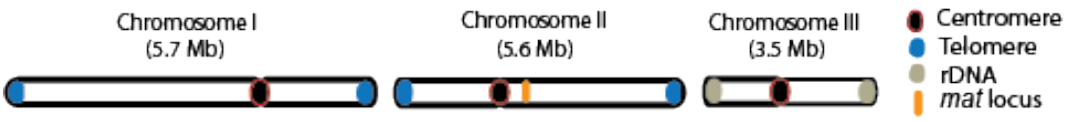

Figure 1.5: S. pombe chromosomal architecture.
Figure adapted from (Mizuguchi et al., 2015).

Genome assembly revealed 5,122 protein-coding genes, 1,522 non-coding RNAs, short introns of around 100 bp, only present for 46% of the genes, an average gene density of 1 gene for every 2,528 bp, and centromeric and telomeric sequence gaps of 20 kb and 40-100 kb, respectively (Wood et al., 2012, 2002). In contrast to the spontaneous mutation rates across the entire genome of 2.0×10^{-10} per site per generation, the spontaneous mutation rates of 1.7×10^{-10} per base in CpG rich-regions are slightly lower (Behringer and Hall, 2016; Farlow et al., 2015). It emerged, however, that the non-protein coding genes were biased for small insertions and deletions, an observation concordant with *S. pombe* retained genetic variation previously described (Jeffares et al., 2015; Wood et al., 2002).

Accordingly, the genome has a minimal degree of gene redundancy where only 41% of the protein-coding genes show gene duplication (Kim et al., 2010; Wood et al., 2002). Recent work applying the Markov model to analyse transposon insertions across the non-coding/unannotated genome defined by UTRs, ncRNAs, and nonessential coding regions showed that these regions incur depleted transposon insertion, suggesting that approximately 91% of the non-encoding genome comprises of functional elements (Grech et al., 2019). Retained gene function as a result of little gene evolution led to over three-quarters of the protein-coding genes as non-paralog genes, and roughly two-thirds of them have human orthologs (Lock et al., 2019). As a result yeast chromosomal organisation, including large replication origins and conserved telomeric proteins (Wood et al., 2012, 2002) as well as a large number of molecular mechanisms (Fruhmann et al., 2017) are highly resembled to that of higher eukaryotes.

1.4 Premise of the project

Although functional genomic techniques have been made available for a diverse range of organisms, including gene deletion libraries in yeast (Giaever and Nislow, 2014), CRISPR-dCas9 in bacteria (Rousset et al., 2018) and RNA interference (RNAi) libraries in worms, flies and mammals (Paddison et al., 2004), genome-wide screen studies on longevity remain elusive. Of the few existing RNAi based screens in *C. elegans*, *D. melanogaster* and mammals most focus on neural disease (Buckingham et al., 2004; Paddison et al., 2004) with the exception for *C. elegans* where a handful of longevity RNAi based genome-wide screens are available (Sinha and Rae, 2016; E. Yanos et al., 2012; Lee et al., 2003). Notably, high-throughput genome-wide longevity based screens in yeast are by far the most abundant, mainly due to the emergence of the systematic non-essential gene deletion libraries, first developed in *S. cerevisiae* (Giaever et al., 2002) and shortly followed in the fission yeast, *S. pombe* (Kim et al., 2010). The availability of the gene deletion library allowed for several functional genome-wide studies to be performed (Yang et al., 2018; García et al., 2016; Zhang et al., 2015; Li et al., 2014; Ucisik-Akkaya et al., 2014; Li et al., 2013; Fang et al., 2012; Pan et al., 2012; Takeda et al., 2011; Calvo et al., 2009; Deshpande et al., 2009; Kennedy et al., 2008). These, however, were limited by performing solid media screens (Lie et al., 2018; Malecki et al., 2016; Doi et al., 2015; Rallis et al., 2014, 2013), or culturing individual mutants in 96-well plates (Sideri et al., 2015; Garay et al., 2014; Kim et al., 2014; Fabrizio et al., 2010; Matecic et al., 2010; Powers et al., 2006).

However, this limitation was addressed by developing Bar-seq, an elegant method which identifies individual mutants by the deep sequencing of their unique molecular barcodes, herein referred to as uptag and dntag (Robinson et al., 2014; Smith et al., 2009). Genome-wide Bar-seq screens in fission yeasts are limiting (Kim et al., 2016; Robinson et al., 2014; Ucisik-Akkaya et al., 2014; Delneri, 2010; Han et al., 2010) and deletion library screens for genes with pro-ageing effects are even more lacking, as to date only one genome-wide Bar-seq screen in *S. pombe* exists (Sideri et al., 2015). The bottleneck to performing Bar-seq despite several developed tools

(Simpkins et al., 2019; Zhao et al., 2018; Lee et al., 2017; Mun et al., 2016) due to specific experimental designs, is that currently no standard tool is available for analysing Bar-seq data. Additionally, another layer of complexity is that the gene deletion mutants must first be decoded prior to Bar-seq. Thus, to overcome these limitations, in this work using insights from previously published work (Grech et al., 2019; Sideri et al., 2015; Han et al., 2010; Smith et al., 2009) we develop methods to first decode the mutant strains from the latest fission yeast deletion library collection and validate our decoded strains by applying an improved version of Bar-seq.

Our developed method to decode the deletion library mirrored the previously published method used to characterise the budding yeast library (Smith et al., 2009) where other than customising our own linker sequences with insights from (Grech et al., 2019) the steps remained roughly the same, including DNA shearing, adaptor ligation, gDNA and barcode-specific amplification using custom-designed primers and library amplification to add the Illumina adaptors required for sequencing. Our improved version of Bar-seq builds onto the previously described method (Han et al., 2010; Sideri et al., 2015). The changes were initially implemented in the uptag and dntag primer specific sequences which we custom-developed to include ”ATCG” and ”GTCA” for the forward and reverse sequences as standard bases to help identify the start/end of the reads, followed by part of the Illumina adaptor sequence, four Ns added as unique molecular identifier sequences and the U1/U2 and D2/D1 sequences flanking the uptag and the dntag barcodes, respectively. Additionally, we also optimised the first PCR reaction used to amplify the barcodes to help reduce PCR amplification bias, followed by product purification and library amplification to add the Illumina adaptors required for sequencing.

Overall project aim

Our project aim was the development of high-throughput methods to enable genome-wide screening of the fission yeast genome to delineate the genetic factors underlying cellular ageing in yeast. Our main focus was on identifying novel genes with pro-ageing effects.

Specific objectives

- Barcode decoding of the fission yeast deletion library strains;
- Develop an improved Bar-seq method, including a robust analysis pipeline;
- Develop and implement a method for RNA-seq from long-term quiescence cells;
- High-throughput chronological lifespan screening of the decoded strains grown competitively in a pool;
- Independent chronological lifespan validation of the top hits.

Chapter 2

Materials & Methods

2.1 General Techniques

2.1.1 Cell media & culture

Fission yeast cells grow in liquid or solid, and on both rich Yeast Extract + supplements (YES) media and minimal media or Edinburgh Minimal Medium (EMM), *(Formedium, Norfolk, UK)*. While the YES media used consisted of 0.5% w/v yeast extract, 3.0% w/v glucose, plus 225 mg/l adenine, histidine, leucine, uracil and lysine hydrochloride as supplements, the EMM media used was composed of 14.7 mM potassium hydrogen phthallate, 15.5 mM Na_2HPO_4, 93.5 mM NH_4Cl, 2% w/v glucose, 20 ml/l salts, 1 ml/l vitamins and 0.1 ml/l minerals as recommended (Forsburg, 2011b). The heterothallic 972 h^- strain was the wild-type (wt) used throughout. Fresh cell colonies were obtained by thawing the required stock on ice and transferring a small quantity of biomass onto a solid YES agar plate using a sterile inoculation loop. The plate was then incubated at 32°C for 2 days and fresh colonies were formed. Pre-cultures were prepared from fresh cell colonies and unless stated otherwise, grown overnight (o/n) at 32°C with shaking at 170 rotations per minute (rpm). Pre-cultures were used to inoculate the main cultures at the required optical density (OD), typically 0.15-0.20 OD_{600nm} using a spectrophotometer *(Fisher Scientific, Leicestershire, UK)* following the manufacturer's instructions. The biochrome absorbs the scattered light emitted by cells at a wavelength of approximately 600 nm, thus estimating cell density.

2.1.2 Colony Forming Units (CFUs)

Colony Forming Units (CFUs) were used to estimate cell viability within a culture population. CFUs were prepared from cultures that reached saturation density, the point at which the first measurements were taken and where 100% viability was assumed. Saturation density is typically attained 2 days post-inoculation and can be detected as no further changes in the OD of the culture. CFUs were prepared by serially diluting the cells collected as a function of time and plating the dilutions on solid media. The plates containing an even lawn of cells were incubated at 32°C until colonies appeared. The number of colonies was used as an indirect measure of cell survival as depicted in (**Figure 2.1**).

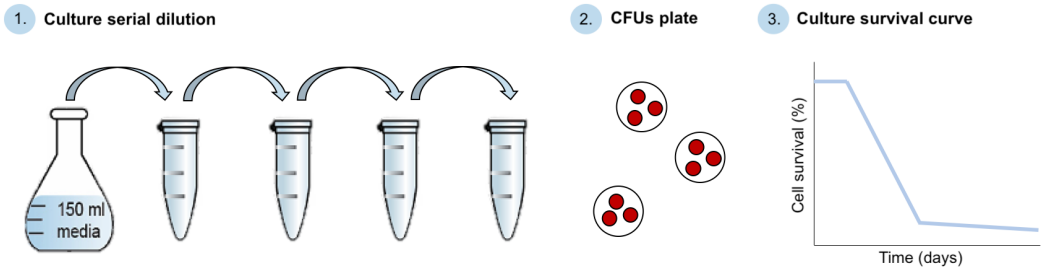

Figure 2.1: CFUs assay.
Culture aliquots of 20 µl were serially diluted ten-fold in 80 µl fresh media ***(1)****. The dilutions were chosen appropriately to obtain between 50-150 colonies per dilution. Serial dilutions were plated in volumes of 50 µl onto solid plates containing 2 mm (diameter) sterile glass beads for an even cell spreading and incubated at 32°C until colonies appeared* ***(2)****. Colonies were counted to generate a viability curve which was used to estimate cell survival within a culture population over time* ***(3)****.*

2.1.3 Sample collection & storage

Collected samples were treated similarly before being stored at -80°C. Briefly, the samples were centrifuged *(Eppendorf, centrifuge model 5810R)* for 5 min at 3,000 rpm. Cell pellets were re-suspended in 1 ml ice-cold 1 × *PBS* (137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM K_2PO_4, pH 7.4) and transferred into 2 ml Sarstedt microtubes *(Nümbrecht, Germany)*. Cells were centrifuged again with cell pellets stored at -80°C until required for further processing.

2.1.4 Genomic DNA extraction

Cell pellets were thawed on ice and processed for DNA extraction using the MasterPure Yeast DNA Purification Kit *(Cambio, Cambridge, UK)* and following the

manufacturer's instructions. Briefly, the pellets were re-suspended in 300 µl Lysis buffer with 3-5 µl of 20 µg/µl RNase A *(Sigma-Aldrich, Haverhill, UK)*. The cells were mechanically lysed by beating with 0.5 mm (diameter) glass beads *(BSP, Stratech Scientific, Ely, UK)* at 6.5 MS for 40 s *(QUICK PREP)* in a FastPrep-24 Instrument *(MP Biomedicals, Leicester, UK)*. The lysis process was repeated twice with 5 min ice incubation intervals. The samples were further incubated for 1 h at 65°C and 150 µl MPC Protein Precipitation Reagent was added before centrifuging *(Eppendorf, centrifuge model 5427R)* the cells for 10 min at 13,000 rpm. The supernatant was transferred into 1.5 ml Sarstedt microtubes (Nümbrecht, Germany) and processed with 500 µl isopropanol *(Sigma-Aldrich, Haverhill, UK)*, centrifuged for another 10 min and washed with 70% ethanol *(Sigma-Aldrich, Haverhill, UK)*. DNA was re-suspended in nuclease-free water *(Qiagen, Manchester, UK)* and stored at 4°C until required for further processing.

2.1.5 Genomic DNA quantification & purification

Genomic DNA (gDNA) was quantified with an Invitrogen Qubit Fluorometer *(ThermoFisher Scientific, Rochford, UK)* and purified using the QIAquick PCR purification kit *(Qiagen, Manchester, UK)* following the manufacturer's instructions.

2.1.6 qRT-PCR

RNA was quantified using NanoDrop™ (*ThermoFisher, Rochford, UK*) before being processed for DNA digestion and RNA column clean-up with TURBO™ DNase (*Invitrogen, UK*) and PureLink RNA Mini Kit (*ambion, UK*) following the manufacturer's instructions. The samples were diluted and processed for cDNA synthesis using the Invitrogen™ SuperScript™ II Reverse Transcriptase (*Invitrogen, UK*) following the manufacturer's instructions. qPCR was set-up with 20 ng/µl cDNA template and SYBR™ Green quantitative Real Time-PCR (qRT-PCR) Kit (*Sigma-Aldrich, UK*) as per the manufacturer's instructions. Samples were made in triplicate with the relevant primers and negative controls. qRT-PCR was run using the QuantStudio 6 Flex Real-Time PCR System *(ThermoFisher Scientific, UK)*.

2.1.7 Gel electrophoresis

UltraPure™ agarose *(Invitrogen™, Rochford, UK)* was used to prepare agarose gels at the required concentrations by following the manufacturer's instructions. Ethidium bromide (0.625 mg/ml) *(ThermoFisher Scientific, Rochford, UK)* drops were added prior to the gel setting to allow for band visualisation. Before running the samples on the gel, the samples were mixed with the loading buffer *(BioLine, London, UK)* as $1/2\times$ sample volume prior to being loaded into the wells with the relevant Hyperladder I-V marker *(Bioline, London, UK)*. The gel was run for 30-60 min at 80-120 volts (V) or until bands separated clearly. The expected product size was observed against the pre-loaded marker using the MultiDoc-It imaging system (*UVP*) under UV-light. Pictures of the gels were attained with a Canon camera (*model PC1305*) inbuilt in the gel doc station.

2.1.8 BioAnalyzer

The BioAnalyzer instrument *(2100 BioAnalyzer, Agilent, UK)* was used to check for the library size and quality. Samples were prepared and run on the BioAnalyzer as per the manufacturer's instructions.

2.2 Protocols

2.2.1 Deletion library strains pool generation

Three library pools of the latest fission yeast deletion collection obtained from Bioneer (`http://www.bioneer.com/`) were independently generated in both auxotroph and prototroph background. The prototroph deletion library strains were constructed from the auxotroph collection using the method described previously (Malecki and Bähler, 2016; Sideri et al., 2015). Independent copies of the auxotroph and the prototroph deletion library pools were generated using the method described in (**Figure 2.2**).

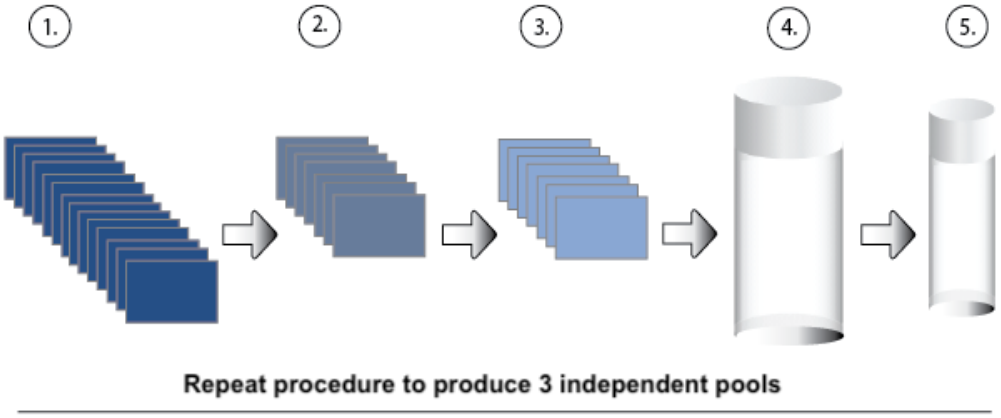

Figure 2.2: Deletion library strains pool preparation.
The deletion collection was thawed at room temperature ***(1)*** *and arrayed onto new plates containing solid YES media using an automated handling robot* ***(2)****. The plates were incubated at 32°C for 2 days to allow the colonies to grow. The array procedure was repeated to transfer some biomass of the grown colonies onto selective plates* ***(3)****. The plates were incubated for 2 more days and washed with liquid YES media to pool the cells together. The pooled cells were collected into a falcon tube* ***(4)*** *and aliquots of 500 µl were prepared and stored at -80°C* ***(5)****.*

The deletion library collection was thawed at room temperature and compacted onto 9×PlusPlates *(Singer Instruments, Watchet, UK)* in a format of 384 colonies per plate using the RoToR robot *(Singer Instruments, Watchet, UK)*. The strains were grown on YES plates for 2 days at 32°C. Colonies were re-arrayed onto selective YES plates supplemented with kanamycin (0.1 mg/ml) and grown for another 2 days at 32°C. Except for an auxotroph pool where the strains were independently pooled together from each plate (used for a better representation of the mutants in the pool during the barcode strain decoding), all the pools were generated by pooling together the strains from all the plates. Colonies were pooled together by washing the plates with 2×1 ml liquid YES medium and collecting the cells into 15 ml Greiner Bio-One^{TM} falcon tube *(CELLSTAR^{TM}, Gloucestershire, UK)* to which 50% (v/v) glycerol stock was added to a final concentration of 20% (v/v). Aliquots of 500 µl in 1.5 ml Sarstedt microtubes (Nümbrecht, Germany) were made and stored at -80°C until required for later use.

2.2.2 RNA isolation from quiescent cells

The new RNA extraction protocol adapted from (Chan et al., 2007) included cell pellet re-suspension in 500 µl extraction buffer (0.25 M NaCl, 0.05 M Tris-HCl (pH 7.5), 20 mM EDTA, 1% (w/v) SDS and 1% (w/v) PVP), mechanical rupture of the cell wall using a FastPrep-24 Instrument^{TM} *(MP Biomedicals^{TM}, Leicester, UK)*, phenol:chloroform isoamyl alcohol (1:1 v/v) *(Sigma-Aldrich, UK)* extraction, cold 70 % (v/v) ethanol (EtOH) precipitation followed by 10 mM lithium chloride (LiCl) precipitation, and finally total RNA re-suspension in nuclease-free water *(Qiagen, Manchester, UK)*. These steps are also depicted in (**Figure 2.3**).

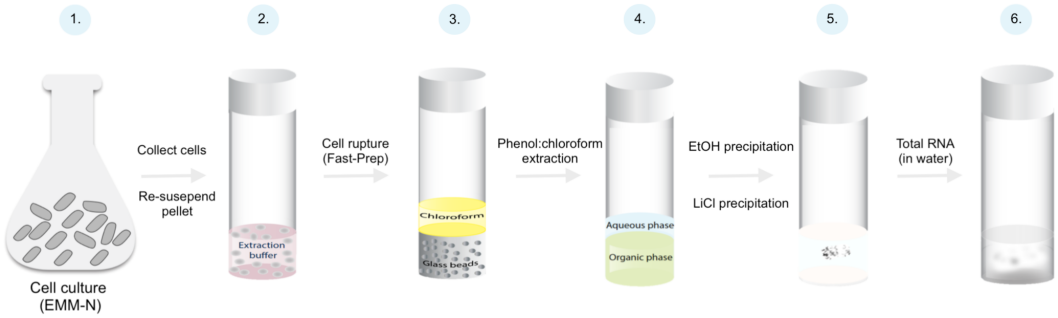

Figure 2.3: The new RNA isolation method.
*Wild-type cells were cultured in 500 ml EMM-N media to saturation density where 2×50 ml samples were collected **(1)**. The pellets were re-suspended in 600 µl extraction buffer and transferred into 2 ml safe-screw tubes (Nümbrecht, Germany) containing 0.4 mm glass beads (BSP, Stratech Scientific, UK) and 500 µl chloroform:isomyl (24:1) (Sigma-Aldrich, UK), **(2)**. Cell wall rupture was performed with a FastPrep-24^{TM} (MP Biomedicals^{TM}, Leicester, UK) using the "QUICKPREP" program as 6.0 MS for 20 s **(3)**. Cell suspension underwent two successive 500 µl phenol:chlorophorm extractions followed by one EtOH and one LiCl precipitation **(4 & 5)**. Total RNA was re-suspended in nuclease-free water and stored at 20°C until required for further processing **(6)**.*

2.2.3 Library preparation

2.2.3.1 Barcode decoding of the deletion library strains

The protocol to decode the mutant strain barcodes was built with insights obtained from previously published work (Grech et al., 2019; Sideri et al., 2015; Han et al., 2010; Smith et al., 2009). Though, the basis of the procedure remained roughly the same, several changes were made. These changes included determining the sonication parameters required for the DNA shearing, ligating adaptors to the end-repaired DNA fragments using custom linkers, amplification and library preparation using custom primers designed to enrich for the uptag and the dntag barcodes separately.

We also increased the complexity of the libraries and reduced PCR amplification bias. The former was achieved by performing two independent PCR reactions per plate and the latter was performed using 5 PCR amplification cycles as these were optimal to obtain enough amplicons while in retrospect minimising the PCR amplification bias. The barcodes were amplified from 1-2 µg of template DNA. The library preparation steps are summarised below and depicted in (**Figure 2.4**).

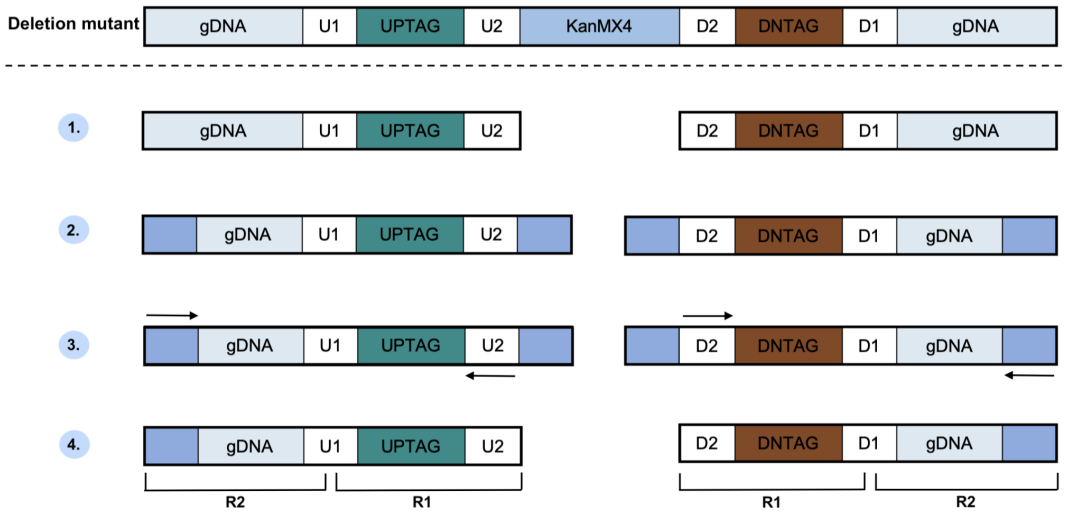

Figure 2.4: Characterisation library preparation.
Extracted DNA was sheared to approximately 400 bp ***(1)****. Adaptors were ligated to the end-repaired DNA fragments containing the barcode tag construct* ***(2)****. The barcode tags were enriched using custom designed primers* ***(3)*** *and libraries were generated upon the addition of the Illumina primers* ***(4)****. The R1 and R2 reads were paired-end sequenced using a MiSeq sequencing instrument. Note, while R2 contains gDNA required for alignment to the genome, R1 contains the amplified barcode.*

Step 1: DNA extraction, purification & quantification

Pellets were removed from -80°C and thawed at room temperature prior to being processed for DNA extraction using the method previously described (see **Subsection 2.1.4**). The extracted DNA was purified and quantified prior to being run on 1% agarose gel using the methods previously described (see **Subsection 2.1.5** and **Subsection 2.1.7**).

Step 2: DNA shearing

gDNA was diluted to approximately 25 ng/µl in 100 µl of nuclease-free water *(Qiagen, Manchester, UK)* in 1.5 ml Bioruptor® microtubes *(ATG Scientific, Oxford, UK)* and sheared to an average size of 400 base pairs (bp) using the Diagenode

Bioruptor® *(ATG Scientific, Oxford, UK)* as per the manufacturer's instructions. Shearing consisted of 7 cycles with intermittent 30 s shearing and 30 s rest. The sheared samples were diluted four-fold and then BioAnalyzed as previously described (see **Subsection 2.1.8**).

Step 3: End repair & linker ligation

Sheared DNA samples were split to treat each barcode reaction independently. End-repair was performed using the NEBNext® End Repair Module *(NEB, Hitchin, UK)*. Linker1 and linker2 (see **Table 2.1**) were ligated using the NEBNext® Quick Ligation Module *(NEB, Hitchin, UK)*. The NEBNext® modules were used as per the manufacturer's instructions. Following end-repair and linker ligation, the DNA was purified and quantified with Qubit as previously described (see **Subsection 2.1.5**).

Step 4: Barcode & genomic enrichment

The barcode tags were enriched with the Phusion® High-Fidelity DNA polymerase *(NEB, Hitchin, UK)* using a primer complementary to linker1 (**Linker1, Table 2.1)** and a primer complementary to U2 (**Uptag-Reverse, Table 2.1)** and D2 (**Dntag-Reverse, Table 2.1)** for the uptag and the dntag barcode tags respectively. The thermocycler parameters used were 15 cycles of 10 s at 98°C, 45 s at 65°C, and 30 s at 72°C. PCR products were purified and quantified with Qubit using the methods previously described (see **Subsection 2.1.5**).

Table 2.1: The list of primers used for decoding the deletion strains.

Oligo	Sequence
Linker1	5'- TTCAGACGTGTGCTCTTCCGATCT**NNNNNNNNN** CAGGCTACTCCGCTTAAGGGAC -3'
Linker2	5- GTCCCTTAAGCGGAGTAGCCTG/3AmMO/ -3
Uptag-Reverse	5'- CACGACGCTCTTCCGATCTAGTA**NNNN** GGGGACGAGGCAAGCTAAGATATC -3'
Dntag-Reverse	5'- CACGACGCTCTTCCGATCTAGTA**NNNN** CGCCATCCAGTGTCGAAAAGTATC -3'

*The primers were custom designed and included part of the Illumina adaptor sequence (red), constant bases (blue), N's acting as unique molecular identifiers (UMIs) (**bold**) and the corresponding U2/D2, uptag and dntag sequences, respectively.*

Step 5: Library generation

The PCR products were diluted ten-fold and used as a template for the second PCR during which Illumina adaptors were added using the NEBNext® Multiplex Oligos Illumina dual index kit *(NEB, Hitchin, UK)*. The thermocycler parameters used were 10 cycles of 10 s at 98°C, 45 s at 65°C and 30 s at 72°C. The products were dual size selected using the AMPure® XP beads *(Beckman Coulter, UK)* as per the manufacturer's instructions. The dual size selection included 0.5× sample volume followed by 0.7× sample volume to remove fragments larger than 700 bp and lower than 200 bp, respectively. The samples were quantified with Qubit, diluted accordingly and run on the BioAnalyzer using the methods previously described (see **Subsection 2.1.5** and **Subsection 2.1.8**).

Step 6: Sequencing

Libraries were multiplexed and sequenced on a MiSeq Illumina sequencing instrument as 75 bp paired-end reads with 165 cycles generating 30 million reads. The architecture of the reads is depicted in (**Figure 2.5**).

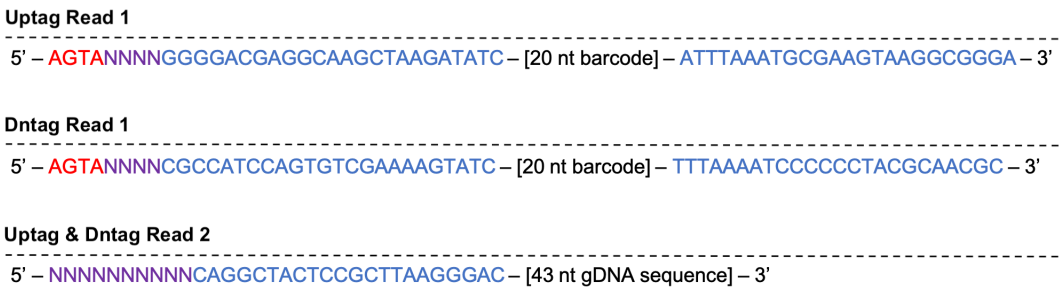

Figure 2.5: Barcode decoding reads structure.
R1 for both uptag and dntag barcode tags share the same architecture: [AGTA]-[NNNN]-[24 bp primer]-[20 bp barcode]-[23 bp primer]. However, the primer sequence (blue) is uptag and dntag barcode specific. The constant U1/U2 and D2/D1 sequences that surround the uptag and the dntag barcode are used to independently amplify the barcodes. R2 architecture [NNNNNNNNN]-[22 bp primer]-[43 bp gDNA] is the same for both uptag and dntag barcodes. Note that the constant AGTA sequence was inbuilt into the primer sequence to easily identify the starting read site with the 'Ns' acting as UMIs.

2.2.3.2 Colony PCR to check for mutant strain gene deletions

Barcode decoding of the deletion library strains revealed 178 mutants for which no barcode tags were found to be associated to their gene deletions. To determine if these genes were successfully deleted, we selected ten mutants at random to verify their deletions. Mutants were selected as one strain from each plate except for the last plate (e.g. plate nine) from which two mutants were selected as this plate corresponded to the largest number of undecoded mutants. The mutants were manually selected from solid YES agar plates and independent colony PCRs were performed as depicted in (**Figure 2.6**).

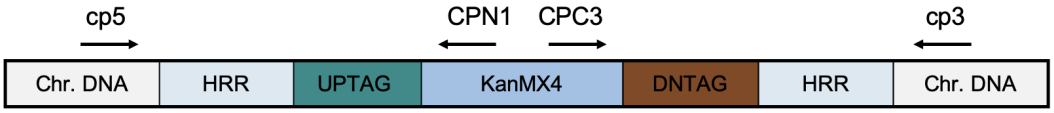

Figure 2.6: Mutant gene deletion validation by colony PCR.
Primers specific for each gene were designed upstream (cp5) and downstream (cp3) of the UTR regions. Each side of the gene deletion was amplified by combining the gene specific primers with the KanMX4 primers: cp5 with CPN1, and cp3 with CPC3. Amplifications spanning the entire genes were also performed using the cp5 and cp3 primers.

Colony PCR for all the genes was performed with the TopTaq DNA polymerase *(Qiagen, Manchester, UK)* using 35 cycles of 30 s at 94°C, 30 s at 52°C and 2 min at 72°C. The products were purified using the QIAquick PCR purification kit *(Qiagen, Manchester, UK)* before being run on 1% agarose gel. The PCR products obtained with the gene-specific primers (cp5 and cp3) were prepared as recommended for the Sanger Lightrun and sent for Sanger sequencing using the CPN1 and CPC3 primers (*Eurofins, Genomics, UK*). The primers used to verify for each gene deletion are shown in (**Table 2.2**).

Table 2.2: The list of primers used to verify mutant gene deletions.

Gene	Plate	Primer
SPAC23C4.16c	1	cp5: 5'-CCACGCTATGATGCGTAACT-3' cp3: 5'-GCTTAAATTTTTGGATTAGA-3'
SPBC146.09c	2	cp5: 5'-TATTTTCTGTCACCATTAGT-3' cp3: 5'-ACAAGATGATATGCATTTCA-3'
SPCC188.13c	3	cp5: 5'-ATAATAATATGTATATGCAA-3' cp3: 5'-TTTGAAACAAACCTGAACCT-3'
SPAC24B11.06c	4	cp5: 5'-AAGCAAACACCACAATCTGT-3' cp3: 5'-ACAGATTAGCATTATAAACC-3'
SPBC1D7.04	5	cp5: 5'-GTTCAACATCGCTAAATATA-3' cp3: 5'-ATATACTAACAAGAACTGGT-3'
SPBC14F5.12c	6	cp5: 5'-TATGGCATTCAAATTCTTTT-3' cp3: 5'-ATTATTTTATCAATATATCT-3'
SPCC1393.03	7	cp5: 5'-AAGCGATTGTGCAACGTAAC-3' cp3: 5'-CTCTGTGTTACACTTCCATG-3'
SPAC15E1.04	8	cp5: 5'-AATGAGTTTATAAAAGTTTC-3' cp3: 5'-ACGTCGGCAGCATTATGAAG-3'
SPBC3H7.07c	9.1	cp5: 5'-AGTACTTGTTGTCGTTTAAC-3' cp3: 5'-TGAATACCTCTGAATATATA-3'
SPAC4F10.05c	9.2	cp5: 5'-TAACCTTAATGTTATTTAAG-3' cp3: 5'-GAAAGCCCGCAAATTTACTG-3'
CPN1: 5'-CGTCTGTGAGGGGAGCGTTT-3' CPC3: 5'-GGCTGGCCTGTTGAACAAGTCTGGA-3'		

The 5' and 3' gene specific primers are denoted as cp5 and cp3. The KanMX4 oligo sequences are denoted as CPN1 and CPC3, respectively.

2.2.3.3 Bar-seq

Genomic DNA was extracted using the method previously described (see **Subsection 2.1.4**) and processed into DNA libraries. The library generation procedure is detailed below and depicted in (**Figure 2.7**).

Step 1: DNA extraction, purification & quantification

Samples were processed using the DNA extraction protocol described previously (see **Subsection 2.1.4**). The DNA was purified, quantified and run on 1% agarose gel using the methods previously described (see **Subsection 2.1.5** and **Subsection 2.1.7**).

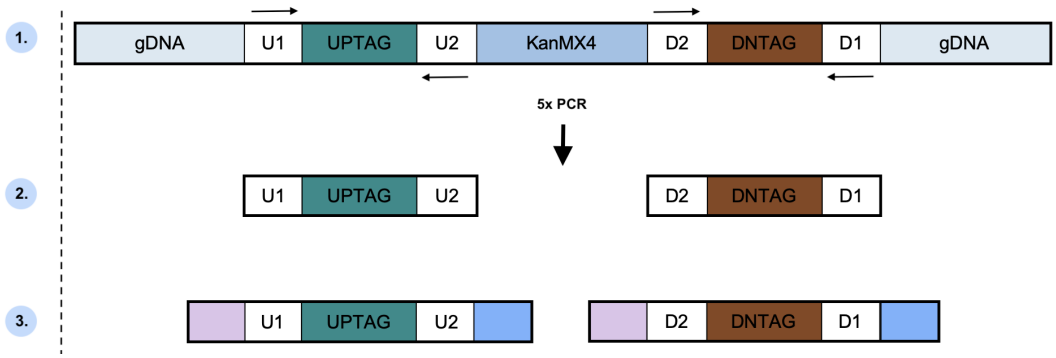

Figure 2.7: Bar-seq library preparation.
*The uptag and the dntag barcode tags were amplified using custom primers with 5 PCR cycles to reduce PCR bias **(1)**. The products were purified, diluted ten-fold and used as a template for the second PCR **(2)**. The libraries were generated upon addition of the Illumina adaptors (purple and blue blocks) using 10 PCR cycles **(3)**.*

Step 2: Barcode enrichment

The uptag and the dntag barcodes were enriched from 125 ng of gDNA with barcode specific primers (see **Table 2.3**) and using the Phusion® High-Fidelity DNA polymerase *(NEB, Hitchin, UK)* following the manufacturer's instructions. The thermocycler parameters used were 5 cycles of 10 s at 98°C, 30 s at 60°C, and 30 s at 72°C.

Table 2.3: Bar-seq primers.

Primer	Sequence
Forward.uptag	5'- TTCAGACGTGTGCTCTTCCGATCTGTCA**NNNN** CGCTCCCGCCTTACTTCGCATTTAAA -3'
Reverse.uptag	5- CACGACGCTCTTCCGATCTAGTA**NNNN** GGGGACGAGGCAAGCTAAGATATC -3
Forward.dntag	5'- CACGACGCTCTTCCGATCTAGTA**NNNN** CGCCATCCAGTGTCGAAAAGTATC -3'
Reverse.dntag	5'- TTCAGACGTGTGCTCTTCCGATCTGTCA**NNNN** TTGCGTTGCGTAGGGGGGATTTAAA -3'

*The primers were custom-designed and included part of the Illumina adaptor sequence (red), 4 constant bases (blue), 4 N's acting as UMIs (**bold**) and the corresponding U1/U2 and D2/D1, for the uptag and the dntag barcodes, respectively.*

Step 3: Library generation

The products were purified and quantified using the method previously described (see **Subsection 2.1.5**). These were used as a template (0.8-3 ng) for the second

PCR where libraries were prepared with the Illumina NEBNext® Multiplex Oligos kit *(NEB, Hitchin, UK)* using 10 cycles of 10 s at 98°C, 30 s at 65°C, and 30 s at 72°C. The products were size selected using 1.4× AMPure® XP beads *(Beckman Coulter, UK)* to remove fragments less than 150 bp. Samples were quantified and BioAnalyzed using the methods previously described (see **Subsection 2.1.5** and **Subsection 2.1.8**).

Step 4: Sequencing

Libraries were multiplexed and sequenced on a MiSeq Illumina instrument as paired-end reads of 75 bp with 165 cycles generating approximately 20 million reads. The reads architecture was designed to aid with the data analysis. R1 and R2 read architecture [XXXX]-[NNNN]-[primer]-[barcode]-[primer] is the same for both uptag and dntag barcodes. Within the read architecture 'XXXX' refers to 4 constant bases (e.g. AGTA and GTCA) inbuilt into the primer sequence to easily detect the start of the read and 'NNNN' refers to 4 bases introduced as UMIs to help remove PCR duplicates. However, while the primer sequences U1/U2 for uptag and D2/D1 for dntag are constant, the barcode sequences vary from mutant to mutant. This similarity in read architecture allows for the full R1/R2 read overlap. This overlap is similar for both barcodes. An example of a dntag read assembly is provided in (**Figure 2.8**).

Figure 2.8: Bar-seq read assembly: a dntag example.
*The 75 bp paired-end reads are assembled using the full R1 and R2 read overlap **(1 & 2)**. Only the fully assembled reads of 86 bp are used for further processing. **(3)**.*

2.2.3.4 RNA-seq

Strand specific mRNA libraries were prepared using a custom version of the Illumina TruSeq Small RNA Sample Prep Kit previously described (Atkinson et al., 2018). The procedure is detailed below and depicted in (**Figure 2.9**).

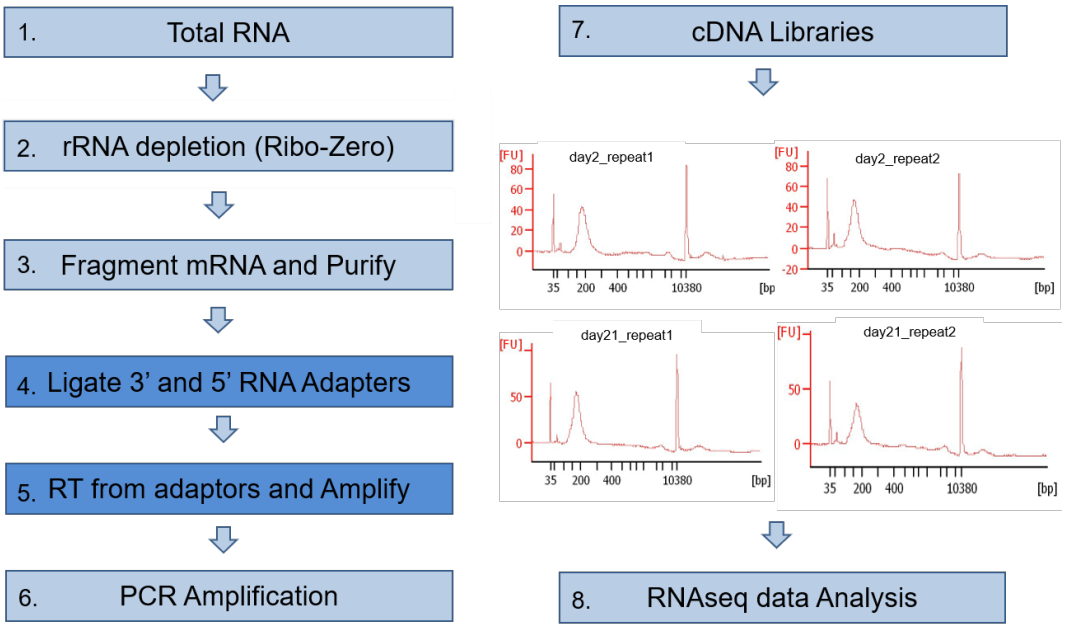

Figure 2.9: RNA library preparation.
Flowchart depicting the mRNA sample library preparation steps. Sample electropherograms were also included.

Step 1: Total RNA & rRNA depletion

The cDNA samples were rRNA depleted using the Ribo-Zero rRNA Removal Kit (*Illumina, UK*) and purified using the yeast RNeasy MiniElute kit (*QIAGEN, UK*) following the manufacturer's instructions.

Step 2: mRNA fragmentation

mRNA fragmentation was carried out using the NEB mRNA Fragmentation protocol (*NEB, England, UK*). Samples were purified using the RNeasy MiniElute kit (*QIAGEN, UK*) following the manufacturer's instructions.

Step 3: Phosphatase treatment

The mRNA fragments were treated with Antarctic Phosphatase (*NEB, England, UK*) following the manufacturer's instructions before purifying the samples with the RNeasy MiniElute kit (*QIAGEN, UK*).

Step 4: Ligation of the 5' and 3' adaptors

Universal RNA 3' and 5' adaptors were ligated to the mRNA fragments using the NEB next kit (*NEB, England, UK*) and following the manufacturer's instructions. The ligated fragments were purified using the RNeasy MiniElute kit (*QIAGEN, UK*).

Step 5: RT-PCR & amplification

RT-PCR was carried out using the Truseq Small RNA Library Prep Kit (*Illumina, England, UK*) with SuperScript^{TM} II Reverse Transcriptase (*Invitrogen, UK*) following the manufacturer's instructions. PCR amplification was carried out using 15 cycles of 10 s at 98°C, 30 s at 60°C and 15 s at 72°C.

Step 6: cDNA library purification & sequencing

The cDNA libraries were purified using the AMPure® XP beads *(Beckman Coulter, UK)* following the manufacturer's instructions. The products were quantified with Qubit and BioAnalyzed using the methods previously described (see **Subsection 2.1.5** and **Subsection 2.1.7**). Libraries were sequenced using the MiSeq Illumina instrument as paired-end reads of 75 bp with 165 cycles generating 28 million reads.

2.3 Data Analysis

2.3.1 Barcode decoding of the deletion library strains

BarSeqTools, our custom analysis pipeline was developed to process the sequence data. The pipeline works by first recognising and extracting the read barcodes and then annotating the extracted barcode sequences to their corresponding genes by matching them to the reference database. The procedure is depicted in (**Figure 2.10**).

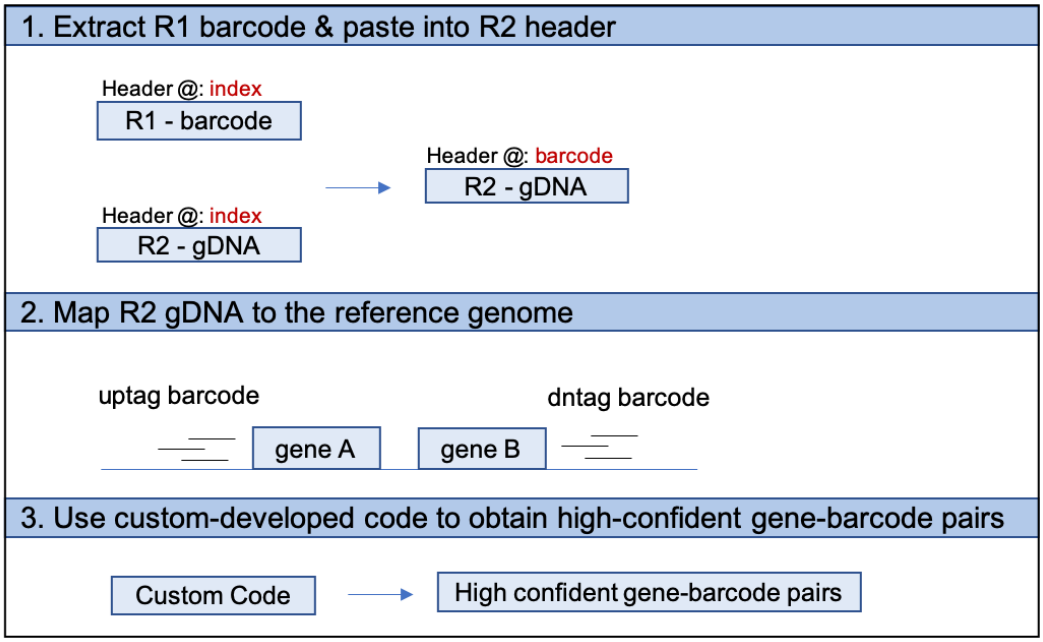

Figure 2.10: Barcode decoding analysis overview.
R1 reads were processed to extract the uptag and dntag barcode sequences. The extracted barcode sequences were added into the R2 header to link the gDNA to its corresponding barcode sequence ***(1)****. gDNA was mapped to the reference genome* ***(2)****. The uniquely mapped reads containing the barcode sequences were assigned to highly confident gene-barcode pairs* ***(3)****.*

An overview of how the pipeline works is shown in (**Figure 2.11**). For a detailed step by step explanation of the analysis (see **Subsection A.1**).

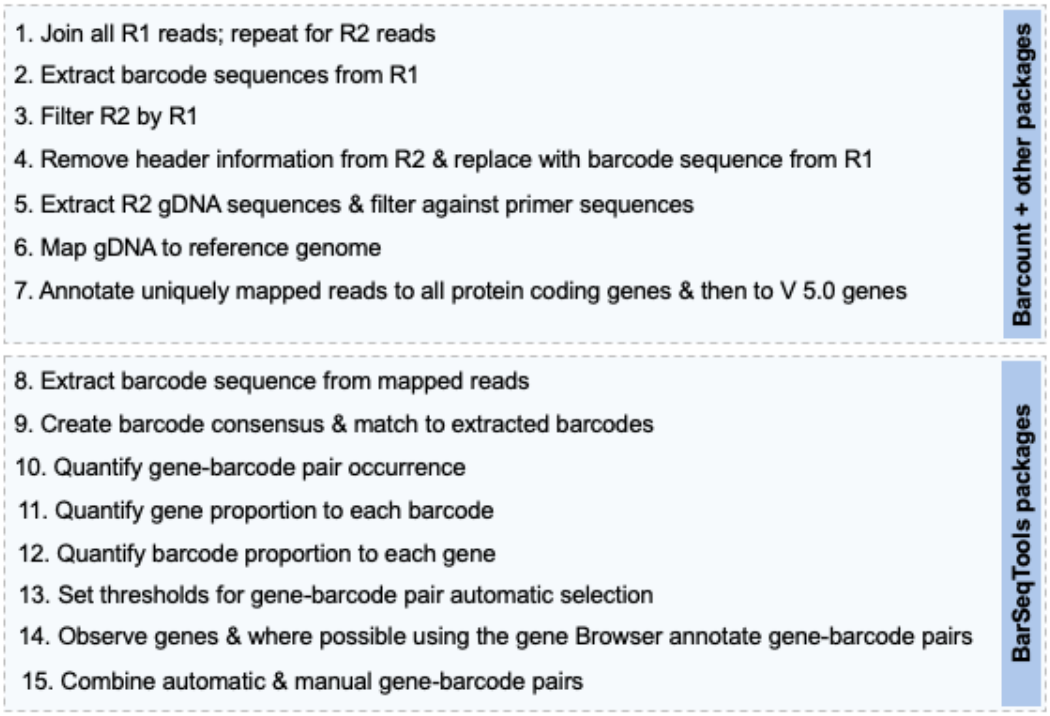

Figure 2.11: Barcode decoding pipeline.
The corresponding uptag and the dntag paired-end fastq files were separately merged into one large R1 and one large R2 fastq file ***(1)****. R1 containing barcodes were extracted with Barcount using the constant U1/U2 and D2/D1, uptag and dntag flanks, respectively* ***(2)****. R1 reads now containing only the barcode sequences were used to filter the R2 reads while replacing the R2 header information with the extracted barcode sequences* ***(3 & 4)****. These genomic sequences were then extracted and filtered against the U2/D2 primer sequences* ***(5)****. gDNA was mapped to the reference genome and only uniquely mapped reads were annotated to the S. pombe's protein-coding genes* ***(6)****. Merging of the annotated S. pombe genes to the version 5.0 genes, linked the gDNA to the deleted gene co-ordinates* ***(7)****. The barcode tag sequences were extracted from the R2 header using BarSeqTools* ***(8)****. Barcode-consensuses from the extracted barcode sequences were generated, matched to the extracted barcode sequences and only barcodes with less than two bp sequence mismatches were used for further analysis* ***(9)****. The frequency occurrence for each unique gene-barcode pair was quantified* ***(10)****. Similarly, the proportion each gene was associated with a barcode within a unique gene-barcode pair was calculated* ***(11)****. This process was also repeated from a barcode perspective* ***(12)****. Combining the frequency occurrence threshold of 10 with the calculated gene and barcode proportion of 80% each, generated our list of high-confidence gene-barcode pairs* ***(13)****. The gene browser was used not only to validate these gene-barcode pairs but to visualise the remaining un-decoded genes and where possible curate them manually* ***(14)****. Automatically and manually curated genes were combined, thus constructing the list of barcode decoded mutant strains* ***(15)****.*

2.3.2 Bar-seq strain fitness comparison to colony screen

The data was analysed using Barcount, our in-house developed tool, the workflow of which is presented below in (**Figure 2.12**).

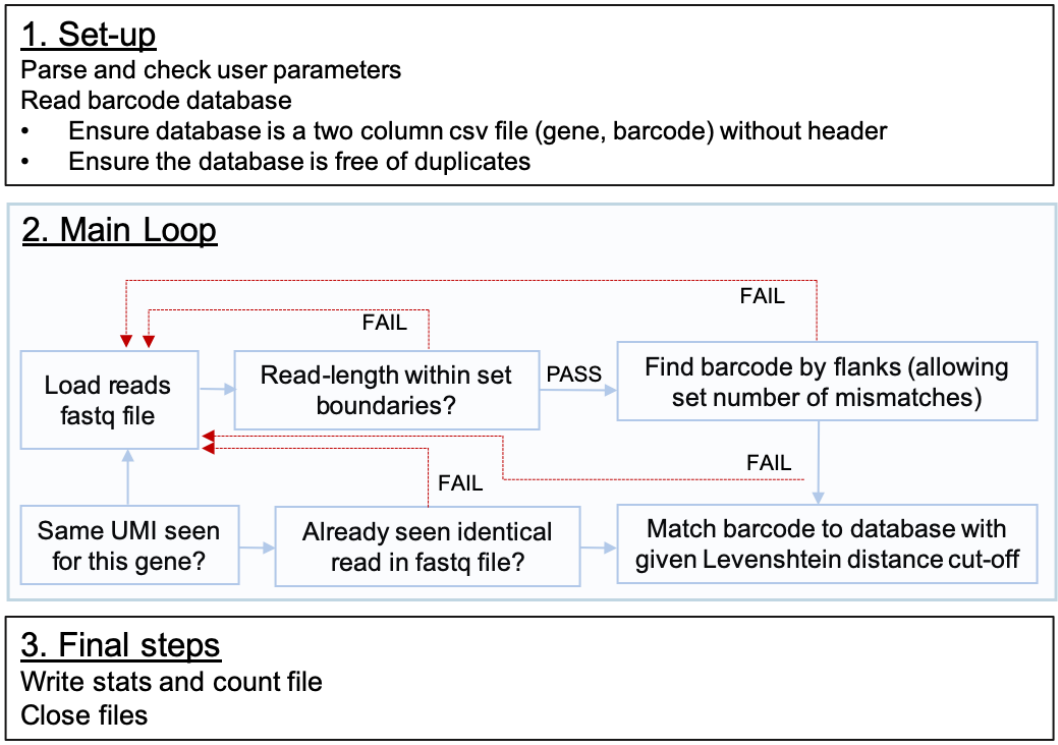

Figure 2.12: Barcount workflow.
Barcount runs with a reference database and a read file ***(1)****. Reads in the sequence file are loaded using the BioPython SeqIO module. The user can specify the read file format which will be passed onto the 'SeqIOs.parse' function with the default being fastq file format. Barcount then checks if the read-length is within the upper and lower limits set by the user (this does not necessarily have to be used simultaneously). Barcode sequences are found using the U1/U2 and D2/D1 flank sequences (the user should provide flanks long enough not to appear at random). The algorithm first searches for exact matches in the flank sequence, then allowing one mismatch, followed by one insertion and one deletion. If both flanks are found, the sequence in-between is extracted as the barcode sequence and compared to the reference database. If the database contains the exact barcode, this is immediately assigned to the gene that matches that barcode. Otherwise, the Levenshtein distance of the extracted barcode to all database barcodes is computed. The best match is selected and the corresponding gene is assigned only if the matching distance is within the user-defined cut-off. If there is a tie between two database entries, no barcode is assigned. If the 'rmdup' filter is set, a hash of the sequence is computed using the function 'Bio.SeqUtils.CheckSum.seguid' which is checked against a cache of previously seen hashes. However, this can increase the run-time of Barcount so if required it might be worth considering alternative tools for this step. Finally, the UMIs are identified by position. If two UMIs are present (–umiA_position and –umiB_position set), they are concatenated and checked against a cache of previously seen UMIs. The cache is specific to each database entry which means the maximum signal per gene is at* $4^{length(UMI)}$ *any number below this shows significant saturation* ***(2)****. The main output generated are the stats and the raw count files* ***(3)****.*

Bar-seq sequence analysis with Barcount included the following steps. R1 and R2 read assembly with PEAR (`https://sco.h-its.org/exelixis/web/software/pear/doc.html`). The assembled reads were then processed independently with Barcount since the flanking sequences to extract the barcodes were based on the U1/U2 and D2/D1 sequences that surround the uptag and dntag barcodes. Therefore, a reference table for each barcode tag exists. Essentially, the reference tables are the decoded uptag and dntag deleted genes. The minimum number of files that Barcount can produce are the count and the stats file. Though individual files for the extracted barcodes, the barcodes matched and unmatched to the reference can also be generated. However, the stats and the count files are the most important files, as the counts file contains the counts per gene and the stats file provides the user with information about Barcount performance. An overview of the analysis is provided in (**Figure 2.13**).

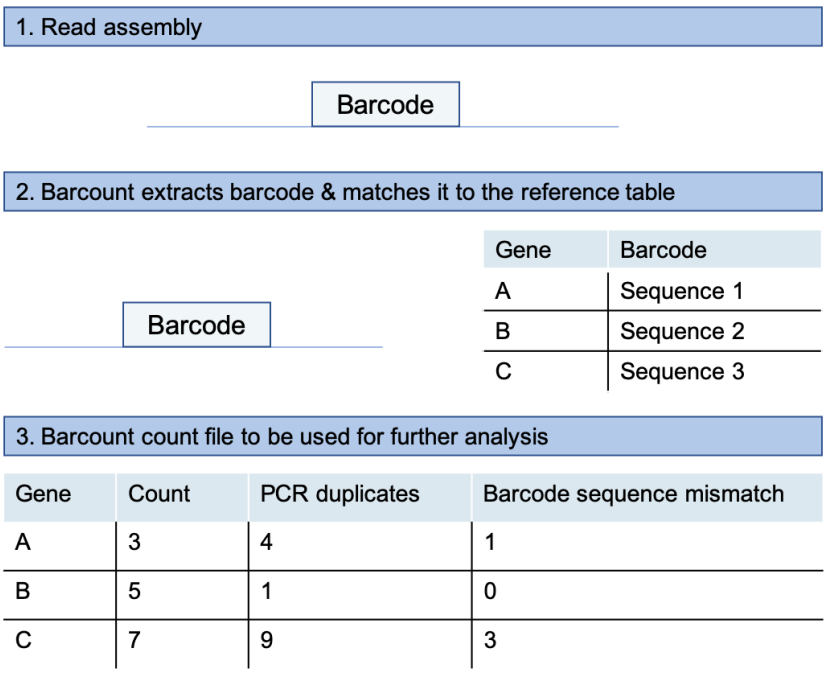

Figure 2.13: Bar-seq analysis.
After performing read assembly with PEAR ***(1)****, Barcount processes the assembled reads by comparing the barcode sequences to the reference database* ***(2)****. The reference database contains the barcode decoded genes segregated into separate uptag and dntag tables. The output obtained from Barcount includes details about which barcode from the sample matched to the reference along with the counts per gene for that barcode* ***(3)****. Note, the numbers in the table are arbitrary. The maximum number of barcode sequence mismatches allowed was three: one insertion, one deletion and one substitution.*

2.3.3 RNA-seq of cellular quiescent cells

RNA libraries were sequenced on the MiSeq Illumina instrument as paired-end reads of 75 bp with 165 cycles generating approximately 56 million reads. The reads were analysed using the public Galaxy server (`https://usegalaxy.org/`) following the steps depicted in (**Figure 2.14**).

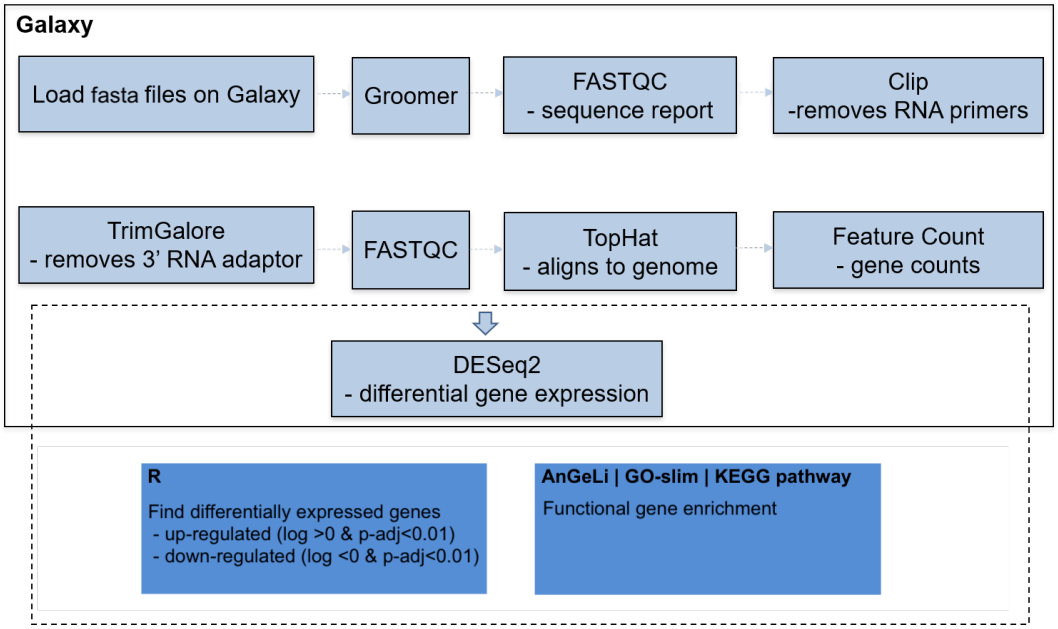

Figure 2.14: RNA-seq analysis.
Flowchart highlights the steps and the packages used for the RNA-seq data analysis.

2.3.4 Chronological ageing of competitively grown decoded strains

Following read assembly with PEAR, Barcount analysed the reads as outlined in the figure above. The stats and the count files were further processed with R (version 3.5.0). Count data were normalised for the sequencing depth and sample correlation plots were produced using the cor() function in R with the following parameters, method="pearson" and use="complete.obs". DESeq2 package (version 1.20.0) the recommended package for Bar-seq data analysis was used to perform differential gene expression (DGE) analysis (Love et al., 2014). DGE was performed on all the replicate time points at day 1 and day 6 for long-lived gene deletion mutants, and day 1 and day 3 for the short-lived gene deletion

mutants. Gene Ontology (GO) (http://geneontology.org/) using AnGeLi (www.bahlerlab.info/AnGeLi) was performed for functional using 'Biological Process' as the GO-term and the decoded genes as the background gene list. KEGG mapper (https://www.genome.jp/kegg/tool/map_pathway2.html) was further used on the differentially expressed gene list to identify the genes annotated to KEGG pathways.

2.3.5 Isogenic strain growth to validate CLS lifespan

In order to facilitate chronological lifespan studies at a larger scale, we have established a novel high-throughput colony forming unit assay which is much less labour- and resource-intensive than the traditional assay and can be largely automated using robotics. The method along with the analysis pipeline is due for publishing (Townsend et al., manuscript in prep.). Briefly, CLS of selected long-lived and short-lived Bar-seq gene deletion mutants were independently validated by growing the strains in isogenic cultures and comparing their lifespan to the lifespans obtained from the competitive pool growth. The validated mutants criteria used was based on the fold-change (FC) and p-adjusted value, plus the gene characterisation status i.e. unstudied genes. Thus, using eight mutants per group, the following gene groups were selected: Bar-seq top long-lived, Bar-seq unexplored long-lived, Bar-seq unexplored short-lived, day 10 and day 12 long-lived, and controls. Besides the wild-type (wt) control, three known long-lived mutant controls, *git3, pyp1 and tco89* and two known short-lived controls *sdh1 and coq5* were added for direct lifespan comparisons.

Mutant candidates were manually selected from fresh prototrophic cell colonies grown on YES plates, re-streaked onto new YES plates and grown at 32°C for 2 days. The grown colonies were then used to set individual pre-cultures grown in parallel in 20 ml YES (3% glucose) o/n at 32°C and 170 rpm. The following morning, individual cultures of 20 ml YES (3% glucose) at 0.002 OD_{600nm} were prepared from the corresponding pre-cultures and grown at 32°C and 170 rpm to saturation density. Once cultures reached saturation, the first time point (e.g. day 0)

was collected, where 100% cell viability is assumed. Thereafter, samples aliquots of 150 µl were collected every two days and transferred into 96-well plates to serially dilute the cells using the ASSIST PLUS liquid handling robot *(INTEGRA, UK)*. The dilutions were spotted as quadruple technical replicates on YES plates in a 384 format using the RoToR HAD robot *(Singer Instruments, UK)*. The plates were incubated at 32°C. After sufficient time to allow for growth, patterns of colonies appear on the plate, and the pattern is indicative of the number of viable cells in the ageing culture, akin to the traditional spot assay. These plates were then stored at 4°C to obtain colonies from all the time points before analysing the data.

To provide quantitative estimates of culture viability, images of plates are acquired and analysed using a custom analysis pipeline. For the general script used for the CLS mutant validation analysis (see **Appendix C**). Briefly, we have developed a statistical model which, given the number of viable cells in a culture, will be able predict probable patterns of colonies: given that a three-fold serial dilution has been performed, the number of viable cells will exponentially decline in each well, altering the probability that colonies will be present or absent at each dilution factor. Based on the concentration of cells per droplet (the volume of liquid transferred by the pinning robot) in each dilution factor, the number of viable cells pinned will follow a Poisson distribution. This distribution can be thresholded (0 or >1), giving the probability that at least one viable cell was pinned i.e. does a colony grow or not. Given that each dilution factor was pinned multiple instances, the number of times a colony grows for each dilution factor will hence follow a binomial distribution. This statistical model can then be used to perform a maximum likelihood estimation of the number of viable cells in the ageing culture based on the pattern of colonies generated by that sample and thus provide quantitative information on the lifespan of ageing cultures. An example of the statistical analysis used to estimate the number of viable cells for one gene deletion mutant (i.e. *vms1*) is shown in (**Figure 2.15**). However, the same principle applies to all of the analysed mutants.

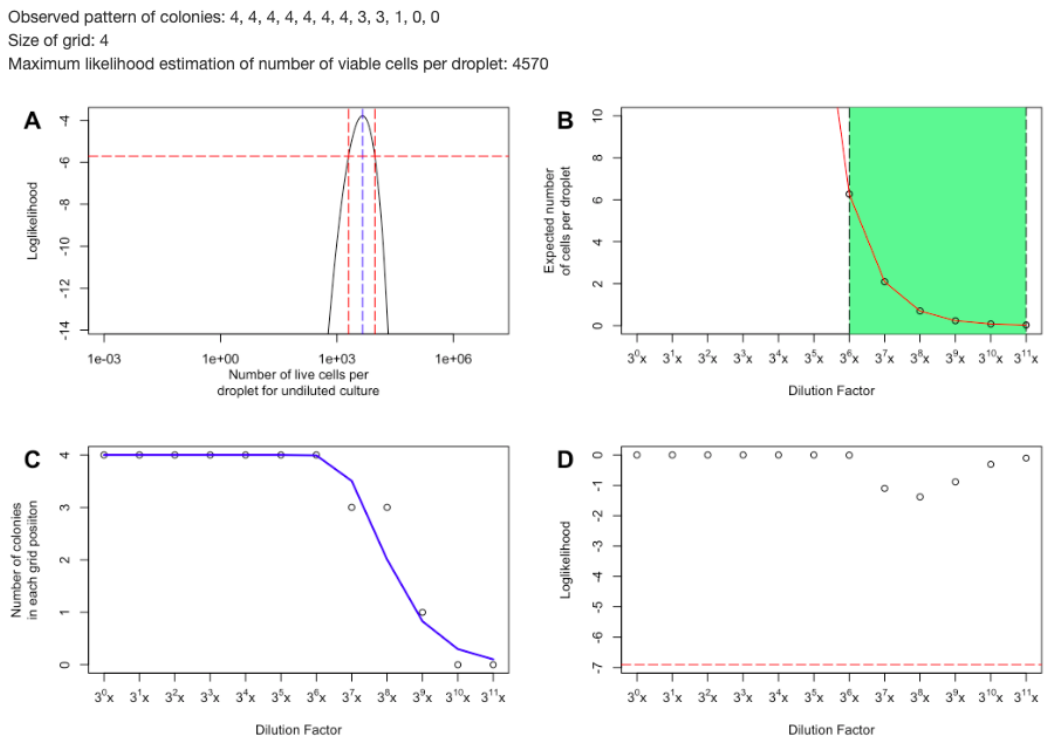

Figure 2.15: Example of the maximum likelihood estimation of the number of viable cells of a mutant

Maximum likelihood estimation plots. ***Figure A*** *shows the likelihood function. Blue dashed lines indicate the maximum likelihood, and red dashed lines show the bounds of the confidence interval for the desired probability.* ***Figure B*** *shows the expected number of cells per droplet at each dilution based on the maximum likelihood estimate. The shaded green area shows the expected informative region - i.e. the region in which there is reasonably probability that some positions will contain a colony and some will not.* ***Figure C*** *shows the pattern of observed colonies. The blue line indicates the expected distribution based on the maximum likelihood estimate.* ***Figure D*** *shows the likelihoods of observing each particular data point (i.e. the number of colonies observed at a particular dilution) for the maximum likelihood estimate. The red dashed line show the tolerance. If there any data points for which the probability of observing them is less than the tolerance, then it is assumed that something has gone wrong. In this case, the most troublesome data point will be excluded, and the maximum likelihood estimation will be performed again.*

2.4 Data Repository

Raw fastq files for each of the experiments described in this thesis, the decoding of the deletion strains, Bar-seq, RNA-seq and chronological lifespan experiments were deposited on the Lab's sequencing repository. For more information or to gain access to these files please contact the lab (`www.bahlerlab.info/`).

Raw count data for each of the experiment presented in this thesis can be found on GitHub (`https://github.com/Catalina37/Experiments_data`). A master file compiling all the CLS screen results, including the outcome of the validated mutants was compiled and can also be found by accessing the link above.

Our developed Barcount and BarSeqTools packages have also been made available on GitHub (`https://github.com/Catalina37/Barcount_BarSeqTools_Pipelines`).

Chapter 3

Barcode decoding of the deletion library strains

3.0.1 Background

Gene deletions are powerful genetic constructs that enable the study of gene function. Although systematic non-essential gene deletions first emerged in *S. cerevisiae* (Giaever and Nislow, 2014; Giaever et al., 2002), the fission yeast systematic non-essential gene deletion collection followed shortly thereafter (Kim et al., 2010). Since first developed, several versions of the fission yeast deletion library have been constructed in both haploid and diploid backgrounds, where with each version iteration more gene deletions have been added to the collection. These different versions, however, are commercially available and can be purchased from Bioneer (`http://www.bioneer.com/`).

For our work, we used the latest library known as version v5.0 deletion collection comprising of 3,420 systematic non-essential gene deletions. This mutant strain collection represents two-thirds of the fission yeast protein-coding genes and approximately 47% of these genes have human orthologs (Lock et al., 2019). The way this collection was constructed is analogous to the methods used to generate the *S. cerevisiae* deletion library collection (Kim et al., 2010; Giaever et al., 2002). In addition to the auxotrophies for the following marker genes; *ade6-M210* or *ade6-M216*, *ura4-D18* and *leu1-32*, the methods included targeted mutagenesis and ho-

mologous recombination. The two techniques were used to integrate the cassette construct containing a selection marker gene, two unique molecular barcodes, and two homology arms into the open reading frame (ORFs) of the gene of interest as shown in (**Figure 3.1**).

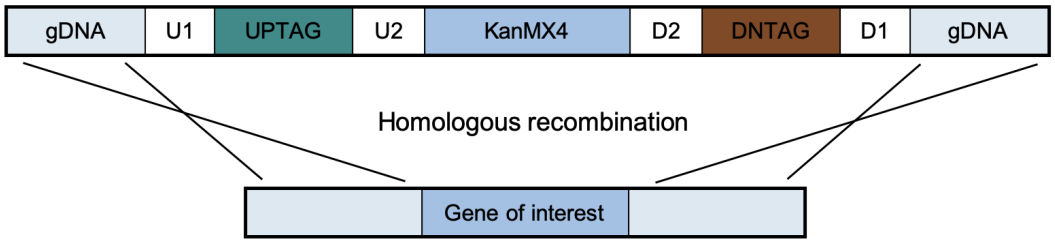

Figure 3.1: Cassette construct used to delete the non-essential genes.
Non-essential genes were deleted by replacing the ORFs with a cassette composed of an antibiotic marker gene, kanamycin (KanMX4), constant U1/U2 and D2/D1 sequences that surround the uptag and the dntag barcodes, and two homology arms containing sequences specific for each gene. The construct was approximately 1 kb in length, where the KanMX4 was roughly 780 bp, the U1/U2 sequences 26 bp and 24 bp, the D2/D1 sequences 24 bp and 26 bp, and the uptag and the dntag barcodes approximately 20 bp each. The figure was adapted from (Kim et al., 2010).

However, the majority of the genome-wide deletion library studies are limited to solid media screens (Giaever and Nislow, 2014; Wagih et al., 2013; Deutschbauer et al., 2002) where the identification of the strain uptag and dntag barcodes corresponding to the deleted gene can only be determined from the strains' position on the plate. This therefore limits the use of the deletion library collection to growing the mutant strains in 96-well plates, or on solid media, as liquid cultures by pooling the mutants together is currently not possible unless one decides to decode the mutant strains. Recent strain barcode decoding on an early version of the budding yeast deletion collection revealed that up to 30% of the mutants digressed from the original design (Smith et al., 2009; Eason et al., 2004), thus raising possible contentions about solid media screens. To date, fission yeast barcode decoding was performed on two earlier versions of the deletion collection (Sideri et al., 2015; Han et al., 2010). Thus to our knowledge the latest deletion library strains have not yet been decoded, or at least not made publicly available despite a recent publication reporting error rates within the barcode sequences (Lee et al., 2018). Here we set out to decode the latest fission yeast deletion library mutant strains collection using our in-house developed tools (see **Subsection 2.2.3.1**).

3.0.2 Experimental design

Analysis of our earlier attempts to decode the deletion collection strains (data not shown) led to the generation of independent pool of strains from each plate to allow for better strain representation during PCR amplification. For details about pool generation (see **Subsection 2.2.1**). Gene barcode decoding was initiated with auxotrophic strain pool aliquots representative of each plate being cultured competitively prior to preparing the collected samples into DNA libraries as depicted in (**Figure 3.2**).

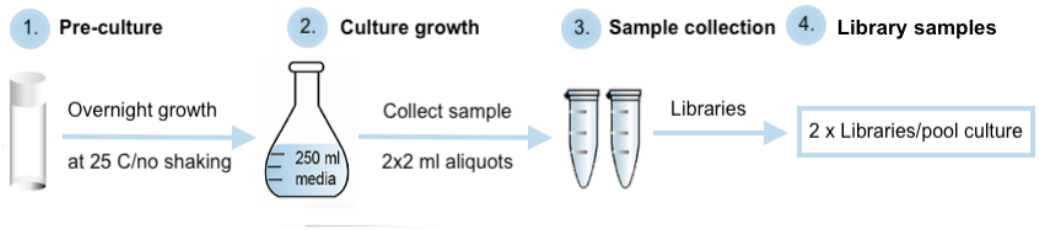

Figure 3.2: Barcode decoding experimental set-up.
Mutant strain pool aliquots from each plate were thawed on ice and cultured in parallel in 250 ml (YES 3% glucose) o/n at 25°C with no shaking ***(1 & 2)****. The next morning,* 2×2 *ml sample aliquots per culture were collected* ***(3)*** *which were then prepared into libraries* ***(4)****. Note, the schematic diagram represents only the process for one pool culture but all nine culture pools representative of the total plates containing the deletion library strains were processed in parallel and treated in the same manner.*

In total, 36 libraries ($9plates \times 2repeats \times 2barcodes$) were prepared using the method previously described (see **Subsection 2.2.3.1**). Libraries were sequenced on a MiSeq Illumina instrument as paired-end reads of 75 bp with 165 cycles generating approximately 30 million reads. Reads analysis was performed using our custom-developed analysis pipeline previously described (see **Appendix A.1**). Sequencing read depth per sample and barcode is provided in (**Table 3.1**).

Table 3.1: Barcode decoding sample names and sequencing depth per barcode.

Sample name	Uptag sequencing depth	Dntag sequencing depth
Plate 1.rep1	6.14	2.42
Plate 1.rep2	0.05	1.96
Plate 2.rep1	0.63	3.24
Plate 2.rep2	2.87	2.45
Plate 3.rep1	2.68	3.54
Plate 3.rep2	1.70	2.44
Plate 4.rep1	2.26	4.78
Plate 4.rep2	0.01	1.66
Plate 5.rep1	1.73	3.35
Plate 5.rep2	1.31	2.19
Plate 6.rep1	3.50	1.64
Plate 6.rep2	2.73	2.34
Plate 7.rep1	2.31	3.39
Plate 7.rep2	3.66	2.58
Plate 8.rep1	2.36	3.27
Plate 8.rep2	2.39	2.41
Plate 9.rep1	2.51	5.30
Plate 9.rep2	3.39	2.35

Sample name denotes the plate number (e.g. plate 1-9) and the two repeats (e.g. rep1 and rep2) per plate. Sequencing depth per sample and barcode (e.g. uptag and dntag) is presented for both reads as a percentage of the total reads (30,706,556) that have passed the quality filter.

3.0.3 Decoded barcodes show a strong correlation

In total, approximately 22% more dntag than uptag reads were found, suggesting that the dntag barcodes were more efficiently amplified than the uptag barcodes (see **Table 3.1**). However, the disparity in the total number of reads did not influence the analysis, and in fact, the uptag reads seem to have out-performed the dntag reads in all the subsequent steps of the analysis as shown in (**Figure 3.3**).

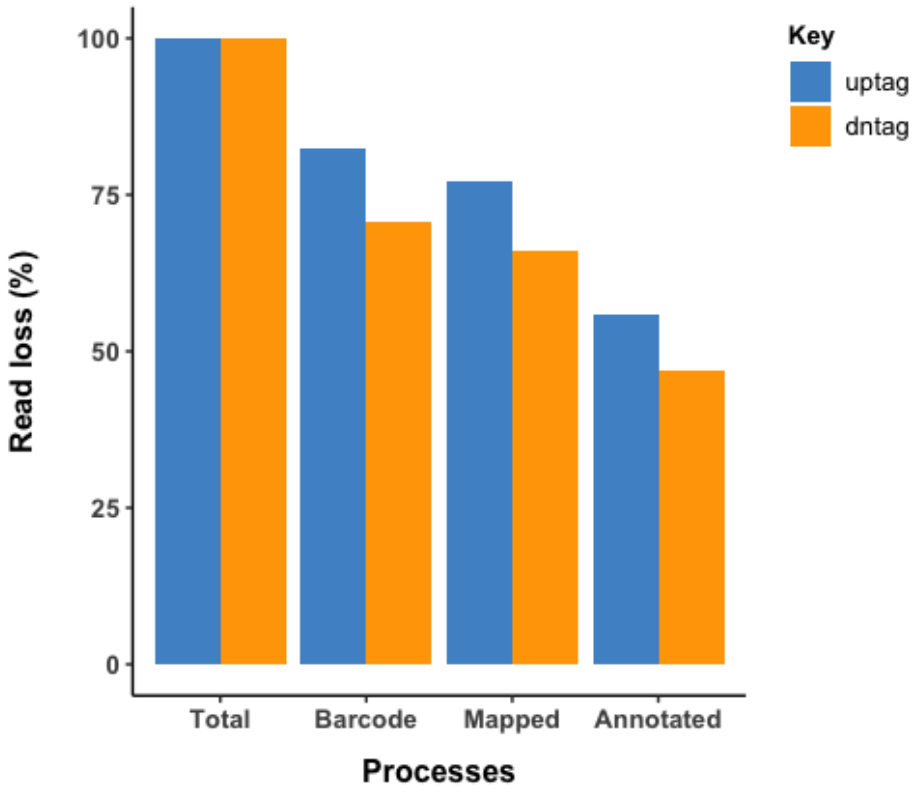

Figure 3.3: Dntag samples show higher read loss than uptag samples.
*Percentage read loss for the uptag and the dntag samples at each analysis stage, including the total number of reads (**Total**), the percentage of reads from which a barcode was extracted (**Barcode**), the percentage of reads which mapped uniquely to the reference genome (**Mapped**), and the percentage of uniquely mapped reads which mapped to the S. pombe protein-coding genes (**Annotated**).*

We found that a large proportion of the successfully mapped reads to the library deletion genes were sub-optimal frequency annotations, described as unique gene-barcode pairs with occurrences lower than ten as shown in (**Figure 3.4**). Furthermore, the exclusion of these sub-optimal gene-barcode pair frequencies from the analysis resulted in unique gene-barcode pairs which were more likely to be genuine gene-barcode mappings than annotations that occurred by chance alone.

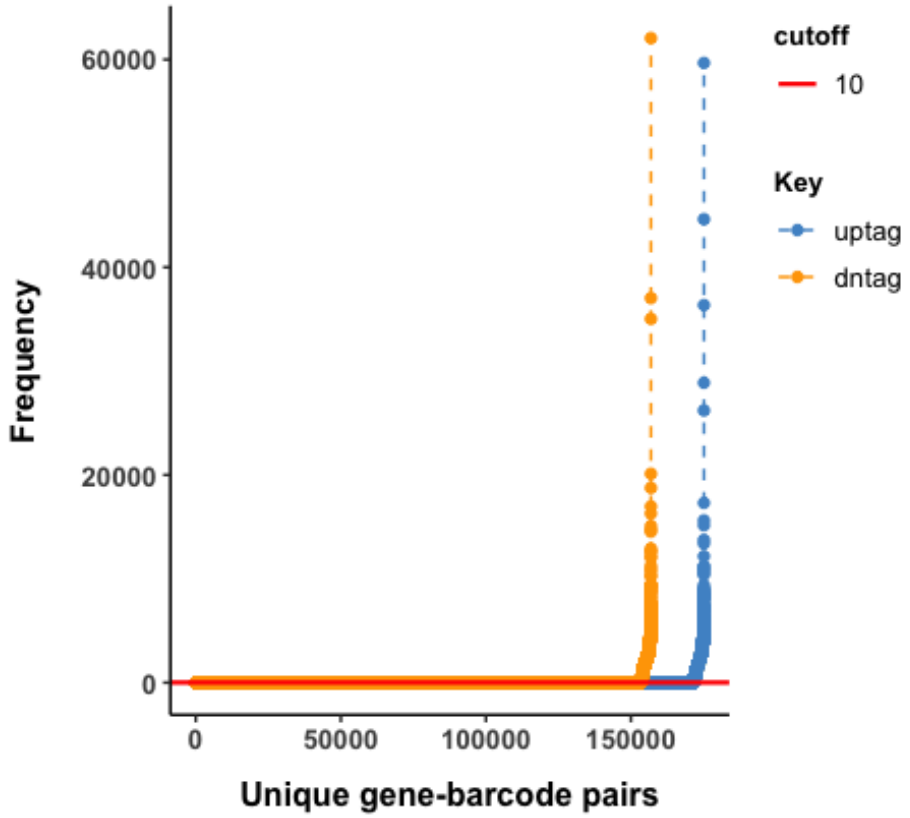

Figure 3.4: The exclusion of sub-optimal gene-barcode pairs frequency.
Unique gene-barcode pairs vs occurrence frequency revealed several gene-barcode pairs that were too low to be considered as true mappings. These pairs were observed for both uptag and dntag barcodes and were removed by setting the minimum frequency at ten.

To delineate which of these gene-barcode pairs were genuine mappings, and to examine the strength between these unique interactions, we decided to calculate the gene and barcode proportions between the pairs. For more details (see **Appendix A.1**). Distribution of these gene and barcode proportions for both uptag and dntag showed a distinct cluster of gene-barcode pairs. This allowed us to set a threshold of 80% for each of these proportions, and by combining this with the minimum gene-barcode pair frequency of ten, we were able to select high-confidence unique gene-barcode pairs as shown in (**Figure 3.5**).

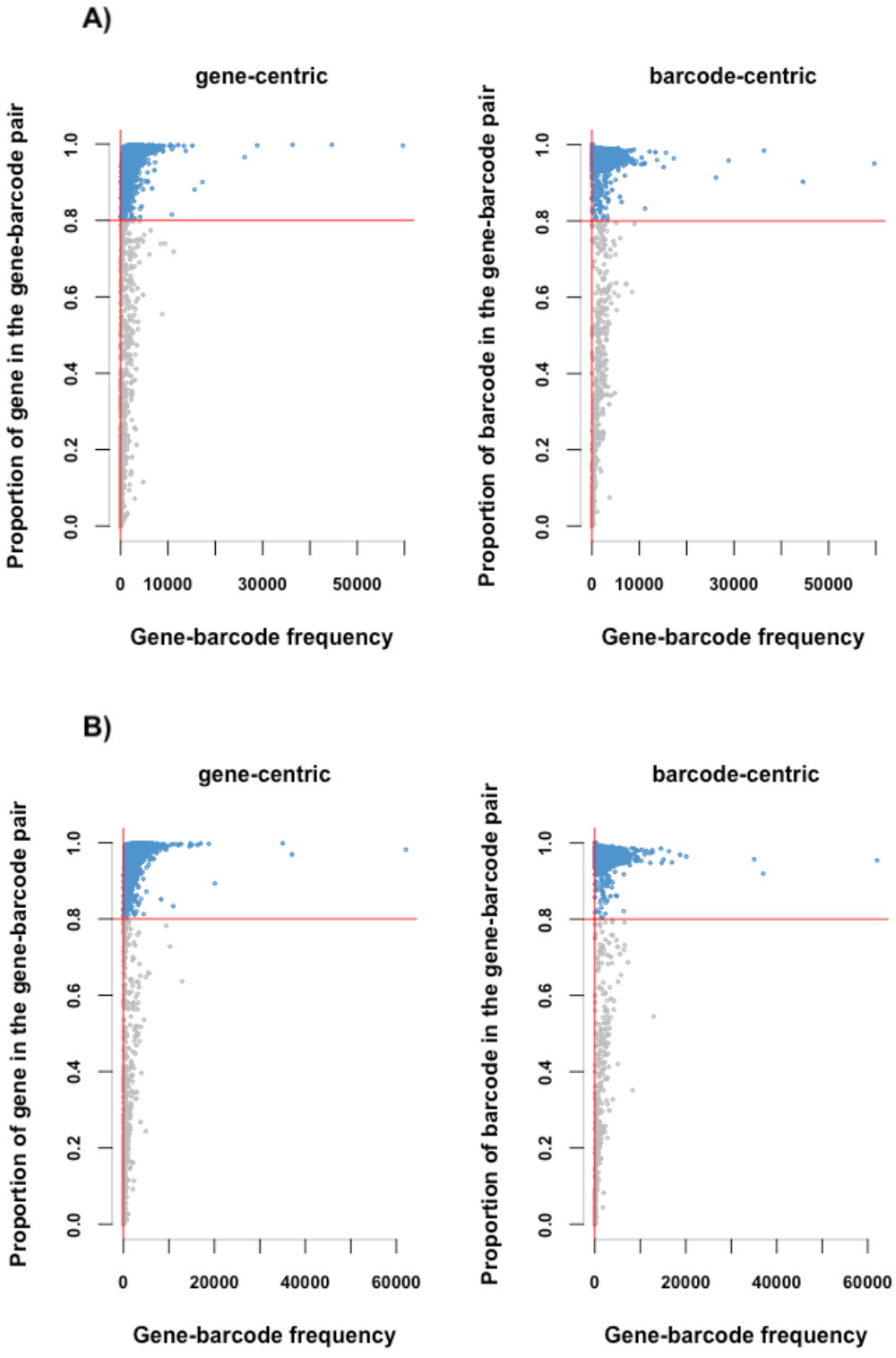

Figure 3.5: Automatic high-confidence gene-barcode pairs selection.
Unique gene-barcode pairs were selected for both uptag ***(A)*** *and dntag* ***(B)*** *by combining the gene and the barcode proportions at 80% with the minimum gene-barcode pair frequency at ten. These thresholds were selected to obtained strong interactions reflective of true mappings between the gene-barcode pairs.*

The chances that the selected high-confidence gene-barcode pairs were erroneous was highly unlikely, as the dual gene and barcode proportion of 80% each could not be serendipitous. Therefore, we were confident that the selected pairs were genuine interactions. We found that of these high-confidence gene-barcode pairs 2,791 genes and 2,890 genes were decoded with an uptag and a dntag respectively, the counts of which were also found to correlate (R^2=0.57, p-val= 2.2E-16) as shown in (**Figure 3.6**).

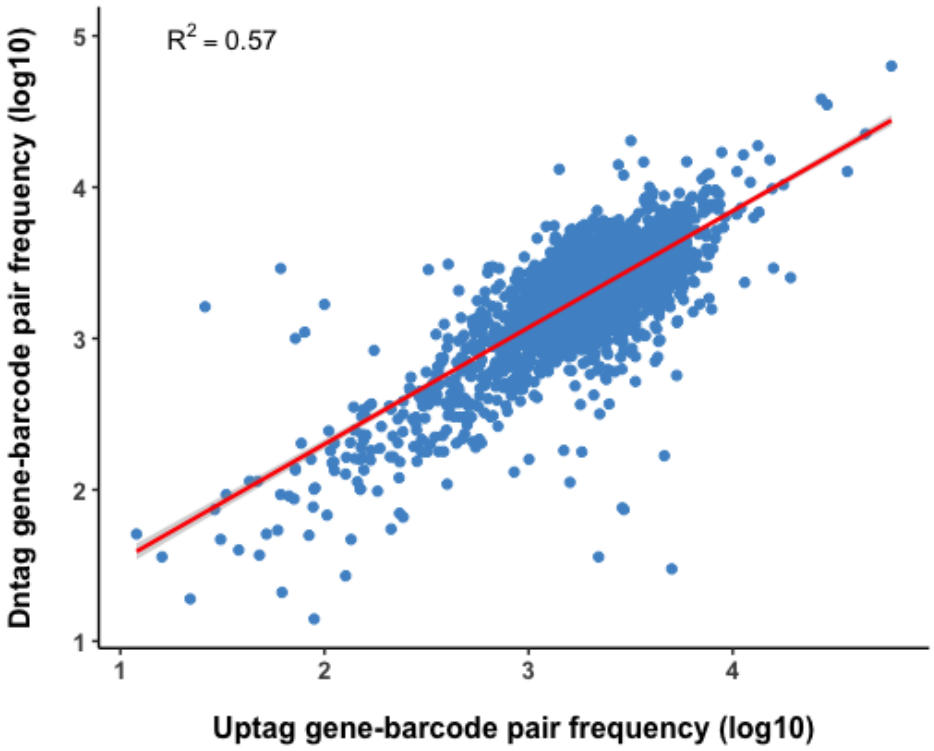

Figure 3.6: Uptag and dntag barcode decoded genes show a strong correlation.
Uptag and dntag gene counts correlation of the high-confidence gene-barcode pairs, with R^2=0.57 and p-value of 2.2E-16.

Furthermore, the genuine interaction of these high-confidence gene-barcode pairs was confirmed by visualising the decoded genes using our custom-developed gene browser. For more details (see **Appendix A.1**). In addition, the gene browser was also used to visualise the remaining genes and where possible these were manually decoded.

3.0.4 We successfully decoded 94% of the library gene deletions

From combining the manually and the automatically characterised genes, we were able to decode 3,206 genes or 94% of the library strains in total. Of these, 3,011 genes were decoded with both barcodes, 96 were decoded with only the uptag barcode, and 99 were decoded with only the dntag barcode. The remaining 214 genes were undecoded and comprised of 36 duplicate genes (e.g. genes that shared the same barcode) and 178 no barcode genes (e.g. genes with no identifiable barcodes). The characterisation of the deletion library is shown in (**Figure 3.7**).

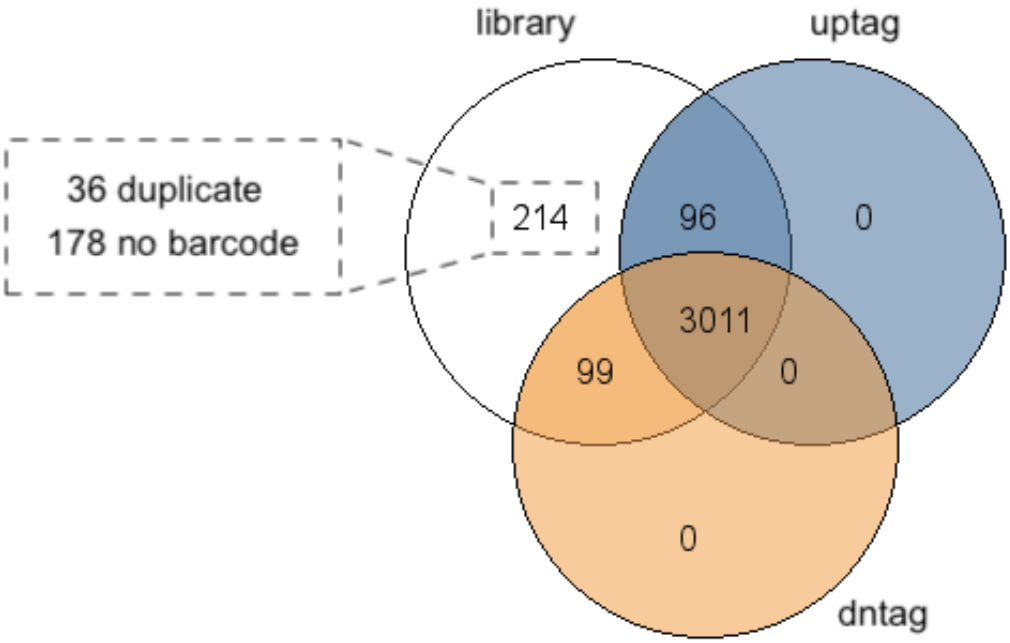

Figure 3.7: Barcode decoding of 94% of the library gene deletions.
We successfully decoded 94% of library collection gene deletions. Of these, 88% were genes decoded with both barcodes, 3% were genes decoded with either uptag or dntag and 6% of genes remained undecoded. Library refers to the version 5.0 deletion genes.

We found that among the 36 duplicate genes eight were indeterminable as these genes shared the same barcodes in equal proportions (see **Figure A.3**), thus we were unable to determine which barcodes corresponded to which genes. A closer investigation revealed that while four of these were telomeric genes (*SPBPB10D8.05c, SPBPB10D8.04c, SPBPB10D8.06c and SPBPB10D8.07c*) residing on chromosome II, another set of four genes were identified as long terminal repeat (LTR) genes (*SPBC1348.01, SPAC750.06c, SPAC212.04c and SPBPB2B2.19c*) with the

former two residing on chromosome I and the latter two genes residing on chromosome II. Subsequently, these genes were found to be annotated as either 'telomeric' or 'tandem duplication' and as a result were not investigated further but led us to treat these deletions with caution.

3.0.5 Not all verified mutants showed successful gene deletions

Intrigued by the fact that 178 genes representing 5% of the library gene deletions remained undecoded, we decided to investigate further. Initially, we speculated that perhaps these were low fitness genes which might have been under-represented in the pool or perhaps the genes were telomeric or LTR genes. To our surprise, all of these speculations were incorrect as we found no fitness correlation, neither for the decoded nor for the undecoded genes as shown in (**Figure 3.8**).

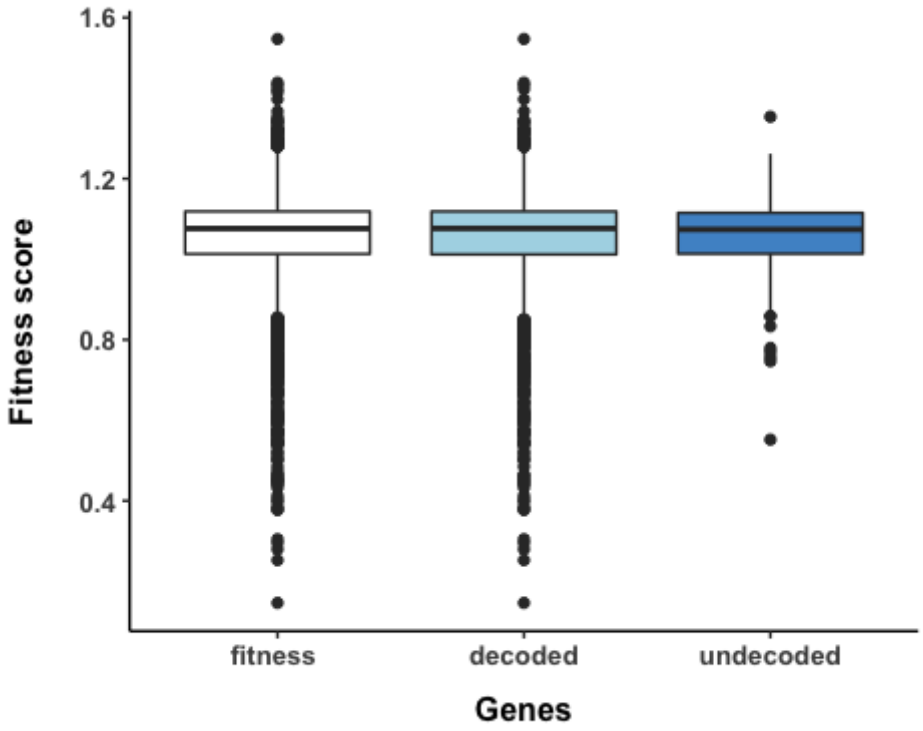

Figure 3.8: Decoded and undecoded genes show no fitness correlation.
Fitness correlation for the decoded and the undecoded genes. Only genes for which fitness data was available were plotted against all the fitness data genes (fitness). Gene fitness information was obtained from the data previously described (Malecki et al., 2016).

Similarly, no bias in the position of the undecoded genes along the chromosomes was found as shown in (**Figure 3.9**).

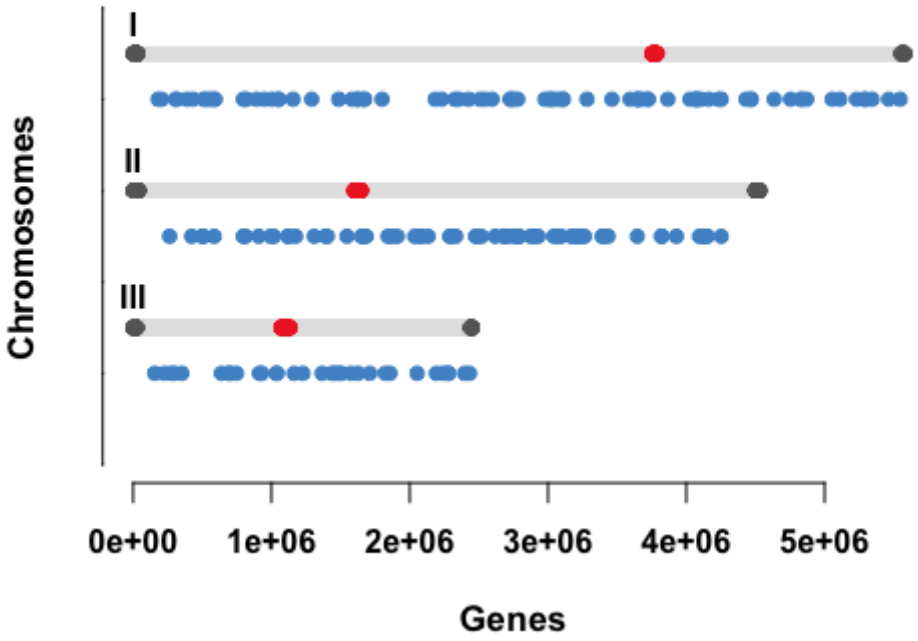

Figure 3.9: Undecoded genes are neither telomeric nor LTR genes.
The chromosomal position of the undecoded genes shows no region bias around neither the telomeres (dark grey) nor the centromeres (red).

Unsatisfied by these findings, we decided to examine whether there was a possible bias between the decoded genes and the plate from which they originated. Since this validation was easy to perform, we also checked the duplicate genes for the same bias. To our surprise, we found that while up to 75% of the undecoded genes originated from plate nine, there was no obvious bias observed for the duplicate genes as shown in (**Figure 3.10**).

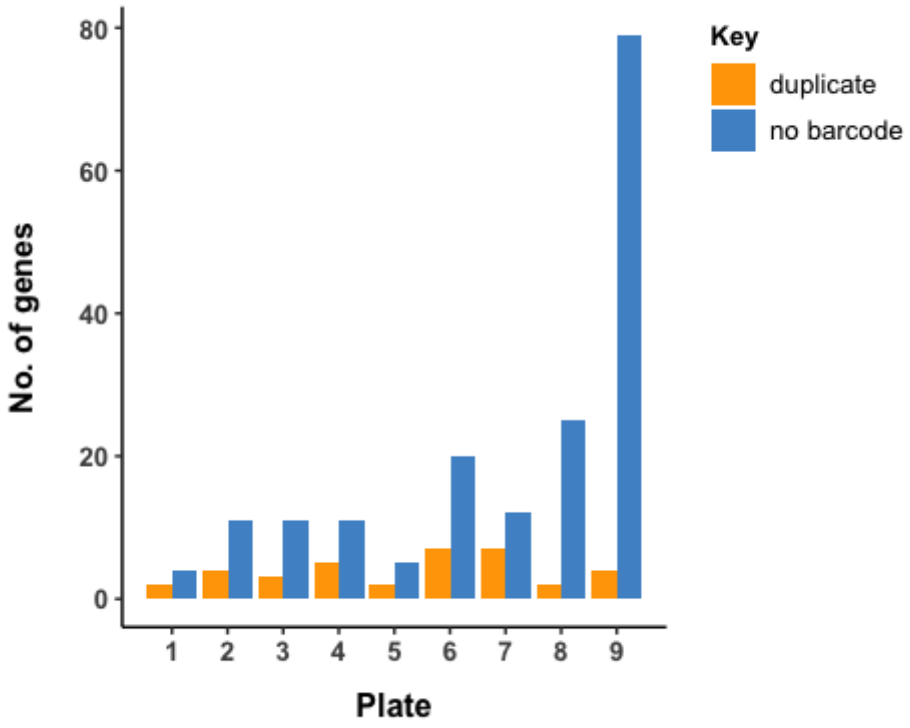

Figure 3.10: Possible mutant position bias for the undecoded but not for the duplicate genes.

The number of genes per plate for both duplicate and undecoded genes.

To further confirm whether the source of the mutants from different plates, especially plate nine influenced our ability to decode the strains, we constructed a new pool containing all the undecoded genes, all the duplicate genes and the genes decoded with only one barcode. For details about pool generation (see **Subsection 2.2.1**). We noticed that during the manual selection of the mutants, the strains from plate nine showed a modest biomass, suggesting that these were not low fitness gene deletions and that our analysis was not biased against the characterisation of the strains with origin from plate nine.

This was because despite modest sequencing depth (2.1 and 1.7 million for uptag and dntag reads respectively), we found that most of the reads mapped to genes previously decoded with one barcode. The analysis revealed only 20 additional decoded genes and confirmed our previous findings of duplicate and unde-

coded genes. Rather than questioning our analysis which was validated through our second gene decoding attempt, we decided to focus on the deletion of the undecoded genes. Therefore, we verified the deletion of ten random strains representing genes from each plate. Because plate nine contained the most undecoded genes, two genes instead of one were randomly selected. The other genes were selected as one gene at random from each plate. Out of ten validated strains, we confirmed the successful deletion of the strains with origin from plate 5, 6 and 7. The strains with origin from plate 1, 8, 9.1 and 9.2 did not produce any primer specific products suggesting unsuccessful deletions. Other than the plate 3 strain which arguably produced a product using the gene specific primers and the cassette primers, the strains with origin from plate 2 and 4 did not produce products across the three amplifications described (**Figure 3.11**), hence rendering the strains as inconclusive. These results were confirmed by Sanger sequencing using the CPN1 and CPC3 primers, and the raw data with supplementary information can be accessed here (`https://github.com/Catalina37/Experiments_data/tree/master/Bioneer_DeletionMutants_SangerSeq`). The extent of this finding was not investigated further but raised concerns about the nature of these gene deletions.

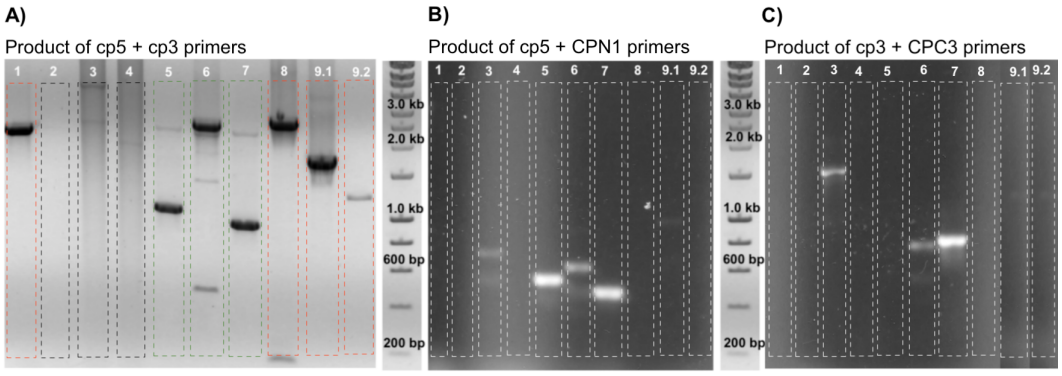

Figure 3.11: PCR analysis of the validated deletion strains hints at tenuous deletions.
PCR products of the ten target genes checked for their deletion. The target genes were independently amplified using the gene-specific UTR primers ***(A)****, 5' UTR gene primer (cp5) combined with the 3' cassette primer (CPN1)* ***(B)****, and the 3' UTR gene primer (cp3) combined with the 5' cassette primer (CPC3)* ***(C)****, where Bioline Hyperladder I was the marker used to determined fragment size. Note that the numbers refer to the plate origin of the target genes. The target genes and the primers used are listed in (****Table 2.2****).*

3.0.6 Our barcodes matched to previously decoded barcodes

Both the uptag and the dntag barcodes of the decoded genes were validated by comparing their sequences to the barcode sequences decoded from the previous deletion library versions; version 1.0 (Han et al., 2010) and version 2.0 (Sideri et al., 2015). Among the common decoded genes, over 87% of the uptag and the dntag barcode sequences were perfect matches to the barcode sequences decoded in the previous versions. However, most of the remaining genes with barcode sequences that did not match either version 1.0 or version 2.0 barcodes was due to different barcode sequence lengths. The proportion of the respective barcode sequence matches to the aforementioned versions was 12.2% and 12.5% for uptag and dntag, respectively, and 10.6% and 10.7% for uptag and dntag barcodes, respectively. Unlike the proposed barcode length of 20 bp (Kim et al., 2010), we found that the uptag and the dntag barcode sequences varied in length between 13-22 bp as shown in (**Figure 3.12**).

Our data was consistent with the findings obtained independently by Lee *et al.,* 2018, a report which found barcode sequence mutations introduced during Sanger sequencing (Lee et al., 2018). This base mutation accumulation within the barcode sequences may have therefore contributed to the difference in barcode lengths, thus explaining the mismatches between our decoded barcodes and the barcodes decoded in the previous library versions. Even so, there were also entire barcode mismatches between the versions, likely as a result of the barcode sequence changes that occurred upon upgrading to new deletion library versions.

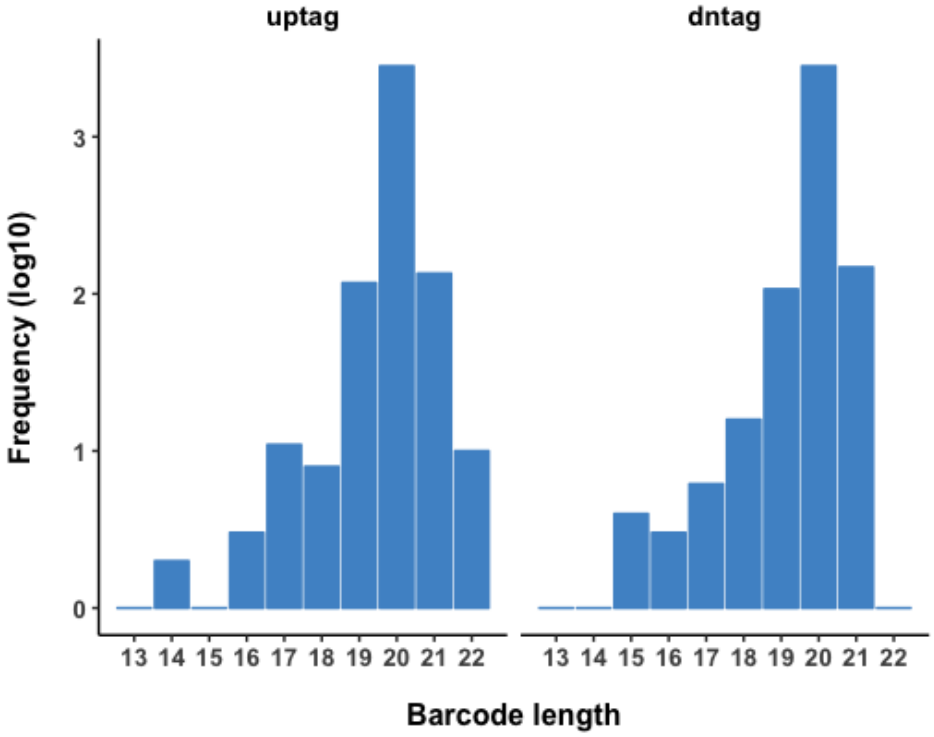

Figure 3.12: The barcode decoded sequences vary in length.
Barcode equence lengths for the decoded genes were found to vary from 13-22 bp for uptag and 14-23 bp for dntag. However, most of the barcodes were 20 bp in length.

Chapter 4

Bar-seq strain fitness comparison to colony screen

4.0.1 Background

First proposed over two decades ago, Barcode sequencing (Bar-seq) is a molecular cell tagging technique developed to facilitate the tracking of cells across time and space (Walsh and Cepko, 1988). Its ability to study thousands of cells simultaneously while collecting information at the individual cell level led to the application of Bar-seq to several areas of expertise with developmental biology (Kebschull and Zador, 2018) and gene function (Smith et al., 2009) being two such examples. The first time gene function analysis was carried out in yeast using Bar-seq, dates back to the generation of the budding yeast deletion library strain collection (Han et al., 2010; Kim et al., 2010; Smith et al., 2009). Because of its high-throughput power, Bar-seq has successfully been applied to several model organisms including yeast (Smith et al., 2009; Giaever et al., 2002), wasps (Cruaud et al., 2017) and plants (Wilkinson et al., 2017).

In fission yeast alone, Bar-seq was used to perform several genome-wide studies (Kim et al., 2016; Robinson et al., 2014; Ucisik-Akkaya et al., 2014; Delneri, 2010). This was possible because of the constant U1/U2 and D2/D1 sequences surrounding the uptag and the dntag barcodes respectively. These sequences were used to amplify the two barcodes separately, thus allowing the study of the deletion strains

in parallel. Without Bar-seq, this study would be carried out using the standard CFUs method (see **Subsection 2.1.2**), an otherwise laborious and impractical task. The application of Bar-seq to profile the deletion strains was previously described in both fission (Robinson et al., 2014) and budding yeast (Smith et al., 2016) using the method depicted in (**Figure 4.1**). Though one major drawback of these Bar-seq studies was data irreproducibility. Although there are several data analysis tools available (Simpkins et al., 2019; Zhao et al., 2018; Lee et al., 2017; Mun et al., 2016) their use typically require a specific design and author correspondence can often be challenging. Unsurprisingly, the data reproducibility bottleneck is mainly caused by the lack of a standard Bar-seq data analysis tool and the experimental differences that exist across laboratories.

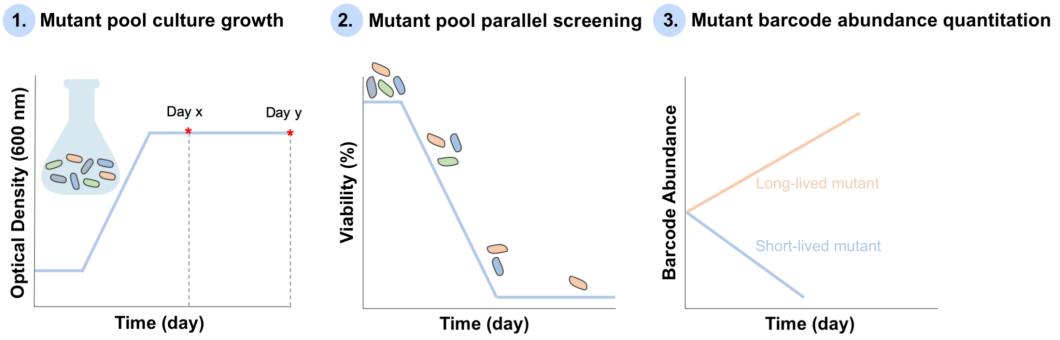

Figure 4.1: Bar-seq profiling of the decoded strains grown competitively in a pool. *Strains are cultured together and grown competitively in a pool to saturation density* ***(1)****. Samples are collected while simultaneously plating CFUs as a function of time throughout the course of the experiment* ***(2)****. Collected samples are then processed to obtain the barcode abundance for each strain which acts as a proxy for fitness* ***(3)****. This enables high fitness strains (light orange) to easily be distinguished from low fitness strains (light blue).*

In this work, we set to improve Bar-seq and custom-develop a robust data analysis pipeline. As proof-of-principle we apply Bar-seq to identify mitochondrial-protein encoding gene deletion mutants. Solid screening on the deletion collection has previously been performed using the same deletion collection which we characterised (Malecki and Bähler, 2016), thus direct comparison of our Bar-seq results to the solid screen results would allow the simultaneous validation of both our developed Bar-seq analysis method and our decoded deletion library strains. Bar-seq screen was performed by culturing the decoded strains on both glucose and glycerol media. When cultured in glucose the mutants grow by fermentation, but when cultured

in glycerol only mutants that can undergo respiration would be able to grow. Mutant growth on these two media types therefore allows for the identification of the mitochondrial-protein encoding gene deletion mutants as these mutants would be able to grow on glucose but not on glycerol.

4.0.2 Experimental design

Three independent pools containing the auxotrophic deletion library strains prepared using the method previously described (see **Subsection 2.2.1**) were grown competitively in parallel, and in either glucose or glycerol media. These cultures were grown in a BioLector® plate *(m2plabs, GmbH, Baesweiler, Germany)* to saturation as shown in (**Figure 4.2**).

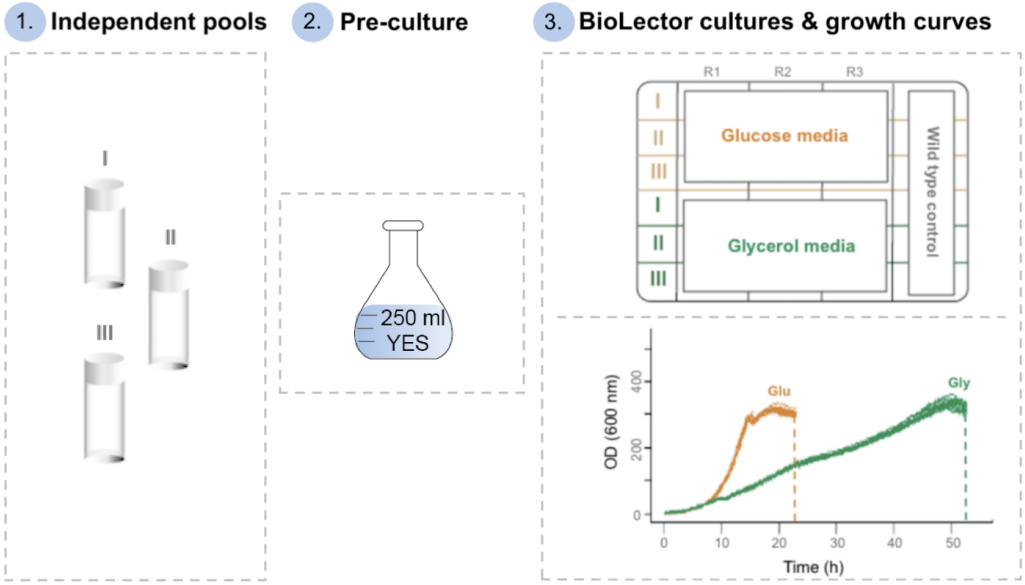

Figure 4.2: Bar-seq proof-of-principle experimental set-up.
Aliquots of 500 µl representing each auxotroph pool (e.g pool I, II and III) were thawed on ice ***(1)*** *and the cells re-suspended in 250 ml YES medium (3% glucose) as independent cultures grown o/n at 25°C and no shaking* ***(2)****. The following morning, the first time point (e.g. T0) culture aliquots of 1 ml were collected from each pre-culture before inoculating the cultures in a 48-well microtiter plate at 0.20* OD_{600nm} *in a total volume of 1.4 ml in either glucose (Glu) or glycerol (Gly) media. Cultures were grown in a BioLector® plate at 32°C to saturation density where* 2×1 *ml samples aliquots were collected (indicated by dotted lines)* ***(3)****. Each pool was cultured in triplicate (e.g. R1, R2 and R3 repeats) with wild-type added as a control. Note, only two repeats (e.g. R1 and R2) were used for further processing.*

In total, 36 DNA libraries ($3 CollectionTimepoints \times 3 pools \times 2 repeats \times 2 barcodes$) were prepared using the method previously described (see **Section 2.2.3.3**). Libraries were sequenced on a MiSeq Illumina instrument as paired-end

reads of 75 bp with 165 cycles generating approximately 20 million reads. Reads were analysed using our custom-developed Barcount package described previously (see **Subsection 2.3.2**). Sequencing read depth per sample and barcode is provided in (**Table 4.1**).

Table 4.1: Bar-seq strain fitness comparison sample names and sequencing depth per barcode.

Sample name	Uptag sequencing depth	Dntag sequencing depth
T0.I.rep1	2.03	2.19
T0.I.rep2	1.78	2.07
T0.II.rep1	1.88	1.69
T0.II.rep2	2.40	1.66
T0.III.rep1	3.60	1.73
T0.III.rep2	1.84	2.37
Glu.I.rep1	2.96	2.00
Glu.I.rep2	2.00	2.16
Glu.II.rep1	3.10	1.54
Glu.II.rep2	2.14	1.85
Glu.III.rep1	1.54	1.67
Glu.III.rep2	2.10	1.95
Gly.I.rep1	2.50	2.22
Gly.I.rep2	2.07	2.16
Gly.II.rep1	2.36	1.91
Gly.II.rep2	1.80	1.93
Gly.III.rep1	1.86	1.33
Gly.III.rep2	2.75	2.01

Sample names denote the media type 'Glu' and 'Gly' for glucose and glycerol, the three independent pools(e.g. pool I, II and III), and the two independent experimental repeats (e.g. rep1 and rep2). Sequencing depth per sample and barcode (e.g. uptag and dntag) is presented for both reads as a percentage of the total reads (20,317,552) that have passed the quality filter.

4.0.3 Bar-seq detected 94% of the decoded genes

We found that the average read assembly for both uptag and dntag samples was 81% as shown in (**Figure 4.3 A**). However, the percentage of the assembled reads with barcodes that matched to the reference barcode database was higher for the uptag than for the dntag samples as shown in (**Figure 4.3 B**). This difference probably arose as a result of lower mutation rates within the uptag than the dntag barcode sequences. Although the lowest barcode matching efficiency was for a dntag sample (e.g. Gly.III.rep2), over 85% of the assembled reads contained barcodes which matched successfully to the reference barcode database, thus providing adequate sequencing depth.

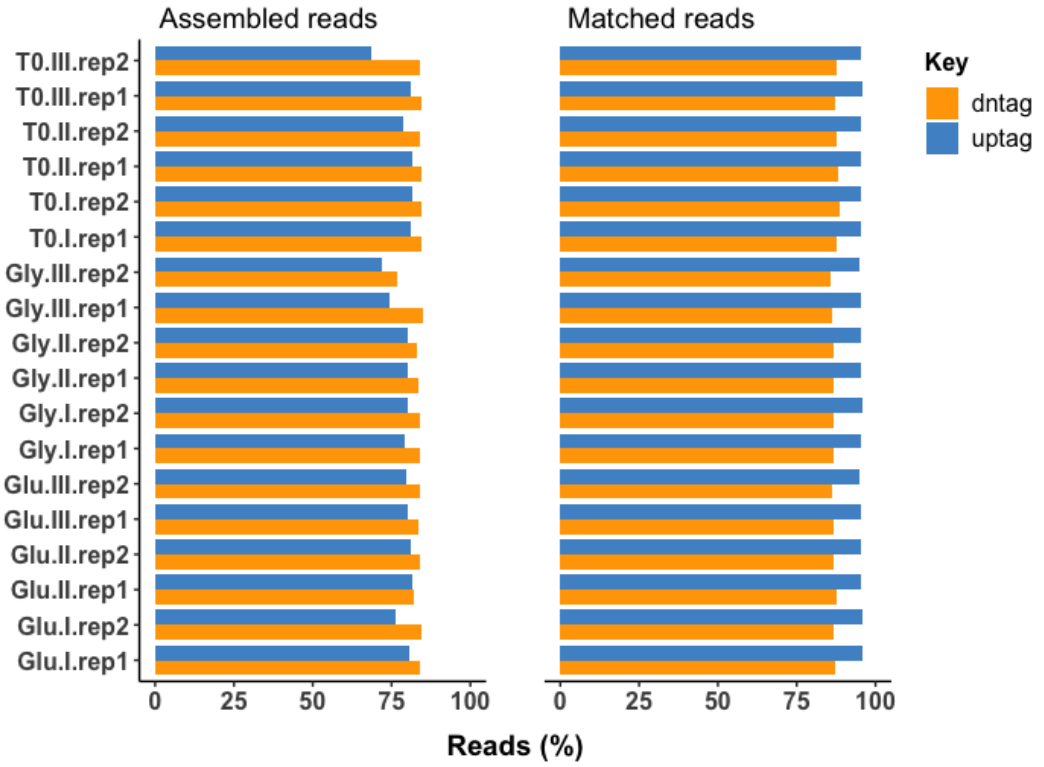

Figure 4.3: Uptag barcodes match to the reference barcodes with higher efficiency than the dntag barcodes.

The percentage of reads post-read assembly for each barcode per sample ***(A)*** *with the proportion of assembled-reads that match to the reference barcode database* ***(B).***

Further examination of the sample gene counts showed the expected uptag and dntag counts correlation as depicted in (**Figure 4.4**).

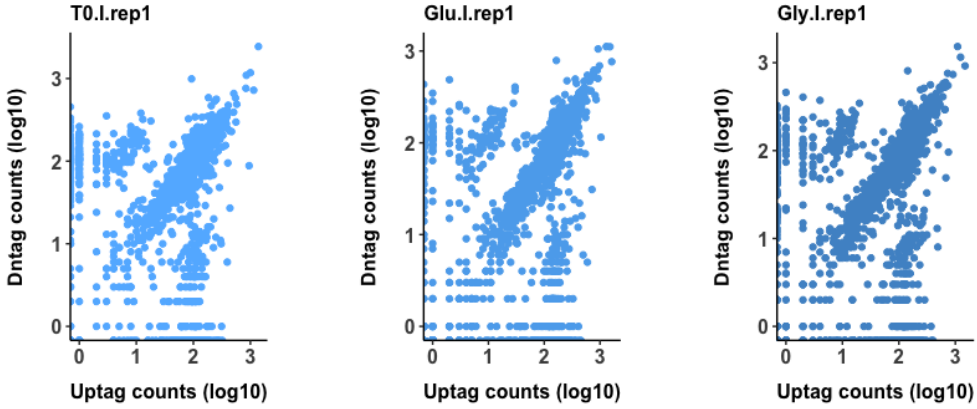

Figure 4.4: Uptag and dntag gene counts show a good correlation.
Correlations between uptag and dntag gene count for the following samples: T0, Glu and Gly, only for one repeat and one pool. Note, similarly strong correlations were also observed for the rest of the samples though these were not shown.

4.0.4 Bar-seq identified 112 mitochondrial protein-coding genes

Because the uptag and the dntag gene counts showed strong correlations for all the samples, we were able to streamline the analysis by combining the uptag and the dntag counts per gene for each sample. Following the combination of the uptag and the dntag gene counts, we identified 3,008 genes, or 94% of the decoded strains. However, sample count correlation analysis revealed that the T0 samples were introducing bias across the repeats (**Figure 4.5 A**) and that while removing them eliminated this bias, the batch effect between the pools persisted (**Figure 4.5 B**). However, the observed pool batch effect might have been caused by preparing pool II and pool III from the same deletion library working stock, while pool I was constructed from a previous library working-stock. Additionally, the difference might have also been caused by the difference in mutants represented within the pools which would have arisen during pool generation (see **Subsection 2.2.1**).

Analysis of the glucose/glycerol gene count ratio on the 3,008 decoded genes using the recommended Bar-seq data analysis package, DESeq2 which accounts for batch effects (Robinson et al., 2014) identified 112 high glucose/glycerol gene count ratio

and 36 low glucose/glycerol gene count ratio shown in (**Figure 4.6**).

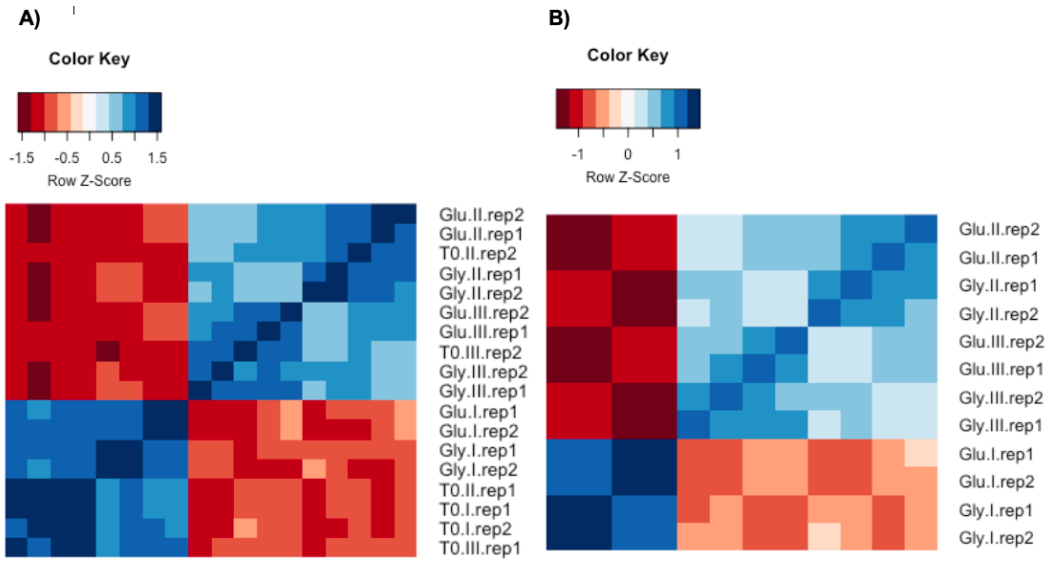

Figure 4.5: Sample barcodes and gene counts show good correlations.
Sample gene count correlations show strong T0 bias ***(A)****. Removing T0 samples revealed the expected correlation between the glucose and glycerol samples, plus the remaining batch effect between the pools* ***(B)****.*

4.0.5 Bar-seq out-performed the colony screen

Despite the difference in the Bar-seq and the colony screen methods, the two screens correlated (R^2= 0.33, p<2.2E-16) as shown in (**Figure 4.7 A**). By comparing the 112 mitochondrial protein-encoding genes identified with Bar-seq and the 204 mitochondrial protein-encoding genes obtained with the colony screen, we discovered 57 genes (p<5.47E-51) in common (**Figure 4.7 B**).

We next compared the number of Bar-seq and colony screen genes to determine which method was more efficient at identifying mutant strains defective in mitochondria genes. The analysis was carried out with AnGeLi, a public database (`http://bahlerweb.cs.ucl.ac.uk/cgi-bin/GLA/GLA_input`) containing information about the fission yeast gene function and molecular processes. Gene ontology (GO) analysis was performed using the 'Cellular Component' as our GO-term and the list of our decoded genes as the background gene list. We found that compared to the colony screen genes (p<4.7E-16), Bar-seq detected approximately two-fold more mitochondrial genes (p<1.1E-30), suggesting that Bar-seq was more sensitive at identifying mitochondrial protein-encoding

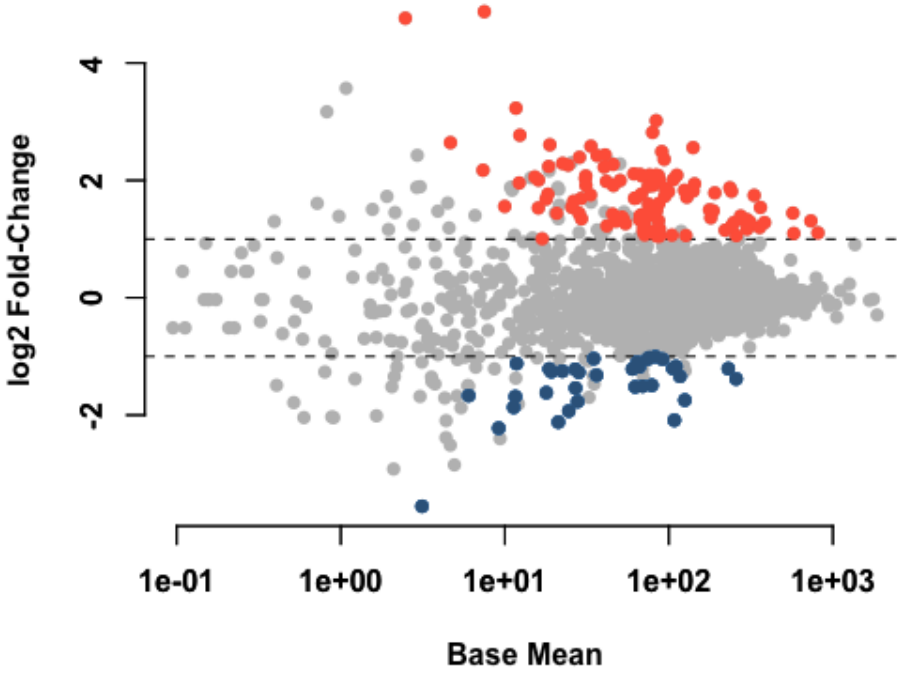

Figure 4.6: Bar-seq identified 112 mitochondrial protein-encoding genes.
Glucose/glycerol gene count analysis identified 112 high glucose/glycerol gene count ratio gene deletions (orange), or mitochondrial protein-encoding genes, and 36 low glucose/glycerol gene count ratio (dark blue). The genes were selected based on log2FC= ±1 and p-adj<0.05.

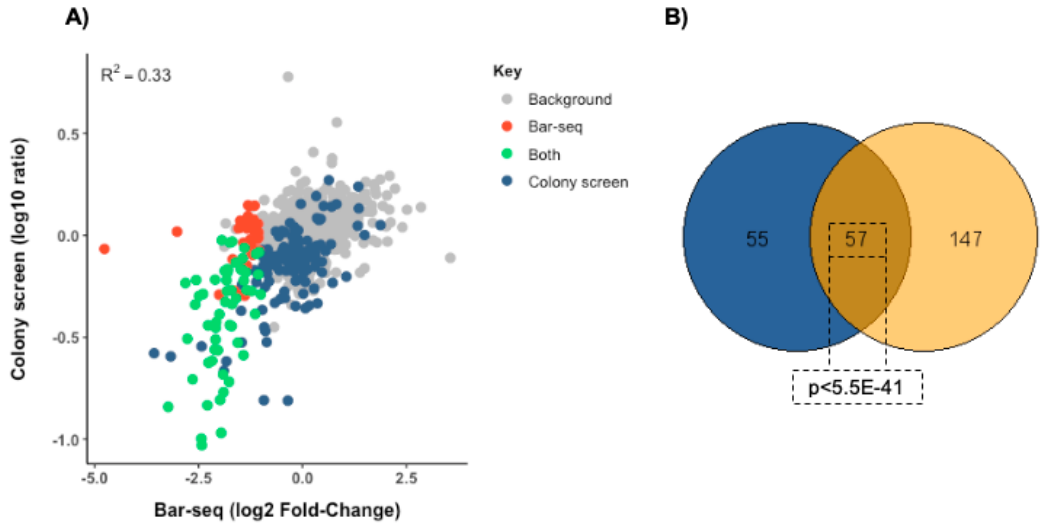

Figure 4.7: Bar-seq and colony screens show good correlation.
Correlation between Bar-seq expressed as log2 fold-changes and colony screen expressed as log10 glucose/glycerol ratio ***(A)****. Gene comparison between the screens revealed 57 common genes* ***(B)****.*

defective mutants than the colony screen (**Figure 4.8**).

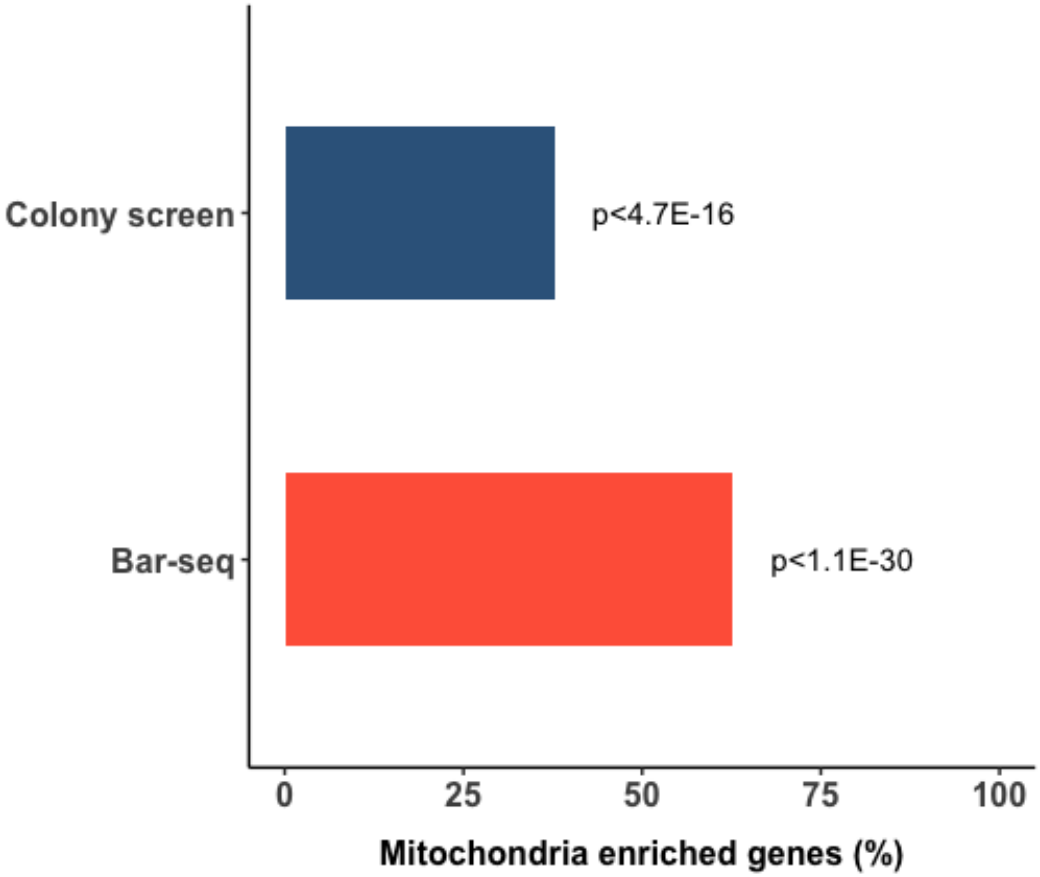

Figure 4.8: Bar-seq vs colony screen of the mitochondrial protein-encoding gene enrichment.
Bar-seq identified more and higher statistically significant mitochondrial protein-coding genes than the colony screen.

Further functional analysis was performed using the Gene Set Enrichment Analysis (GSEA) package previously described (Chen et al., 2018). The results were similarly encouraging and revealed that Bar-seq not only identified genes involved in several molecular processes but identified genes involved in processes not previously detected with the colony screen as shown in (**Figure 4.9**).

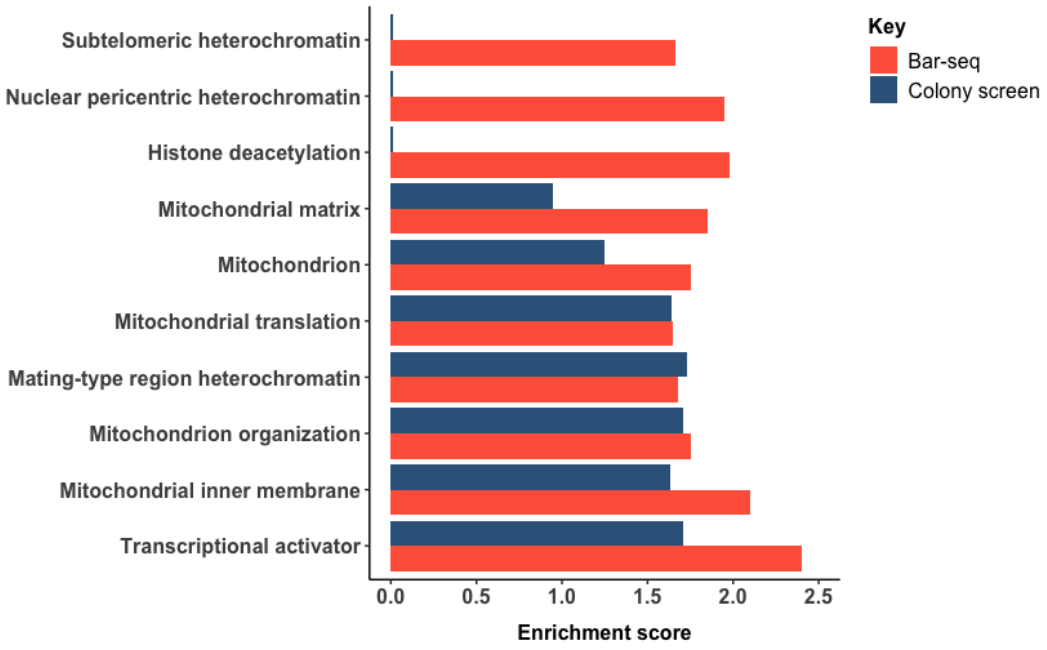

Figure 4.9: Bar-seq was more efficient at identifying respiratory defective mutants than the colony screen.

Bar-seq identified new genes with molecular processes not previously found with the colony screen.

Chapter 5

RNA-seq of cellular quiescent cells

5.0.1 Background

Improvements in sequencing technologies allow for genome-wide sequencing at a relatively low cost. Combining this sequencing advantage with the fission yeast genetic tractability, we sought to optimise an RNA-seq method to investigate the quiescence of old yeast cells. RNA-seq of long-term quiescent cells would provide insights into the transcriptional changes that occur during cellular quiescence. These would complement the genomic analyses obtained with Bar-seq, thus presenting an opportunity with the potential to unravel new links between the molecular factors underlying ageing. As such, we planned for the future use of Bar-seq on quiescent cells though our efforts to date focused on establishing an RNA-seq method for old yeast cells. This was because while in yeast quiescence can easily be induced by growing and ageing cell populations in EMM-N, the isolation of high-quality RNA from aged cells up until now proved challenging due to the instability and degradation of RNA. Here, we establish a new RNA protocol that allows for high-quality RNA isolation from long-term quiescent cells suitable for RNA-seq.

5.0.2 Experimental design

To confirm the extraction of high-quality RNA from old quiescent cells, we prepared independent wild-type cell cultures as per the method described in (**Figure 5.1**). Two technical repeats per time points; exponential, day 0, 2, 4 and 21 days old quiescence samples were processed for RNA isolation using the method previously

described (see **Subsection 2.2.2**). qPCR was prepared for all the samples, though only 2 days and and 21 days old quiescent samples were prepared into RNA libraries using the method previously described (see **Subsection 2.2.3.4**). Libraries were sequenced on a MiSeq Illumina instrument as paired-end reads of 75 bp with 165 cycles generating 28 million reads. The reads were analysed using the standard DESeq2 package as previously described (see **Subsection 2.3.3**). Sequencing depth per sample and barcode is provided in (**Table 5.1**).

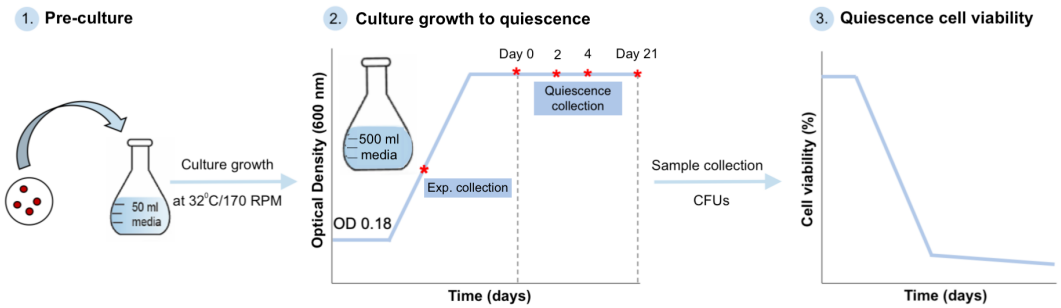

Figure 5.1: Quiescence cells culture set-up.
Fresh wild-type cell colonies were used to set-up three independent pre-cultures in 50 ml EMM medium with o/n growth at 32°C and 170 RPM ***(1)****. The following morning, cell pellets from each pre-culture were re-suspended in 500 ml EMM-N medium at 0.18* OD_{600nm} *and grown at 32°C and 170 RPM to saturation density, the point at which cells enter quiescence and where the first collection time point (e.g. day 0) was collected* ***(2)****. Samples of* 2×50 *ml aliquots were collected (red asterisks) at the following time points: exponential, day 0, 2, 4 and 21 days with subsequent CFUs which were used to obtain population cell viability* ***(3)****. Note, quiescent cultures were maintained for 21 days by replacing the EMM-N medium once/week for the first week and twice/week after that to replenish the media glucose levels.*

Table 5.1: RNA-seq quiescent sample names and sequencing depth per barcode.

Sample name	Sequencing depth
Day2.rep1	14.2
Day2.rep2	14.4
Day21.rep1	15.8
Day21.rep2	12.4

The sample names denote the number of days spent in quiescence (e.g. day 2 and day 21 refers to 2 days and and 21 days old quiescence samples, respectively) and the two repeats (e.g. rep1 and rep2) per sample. Sample sequencing depth is expressed in millions.

5.0.3 The new RNA isolation protocol preserves RNA

The adaptation of an RNA extraction protocol previously described in plants (Chan et al., 2007) enabled us to circumvent the challenge of high-quality RNA extraction from long-term quiescent cells. We next tested the protocol by comparing RNA integrity from 4 days old quiescence samples treated with both the new and the standard protocol, and found that the quiescent samples treated with the new protocol showed a higher RNA integrity than the sample treated with the standard protocol as shown in (**Figure 5.2**).

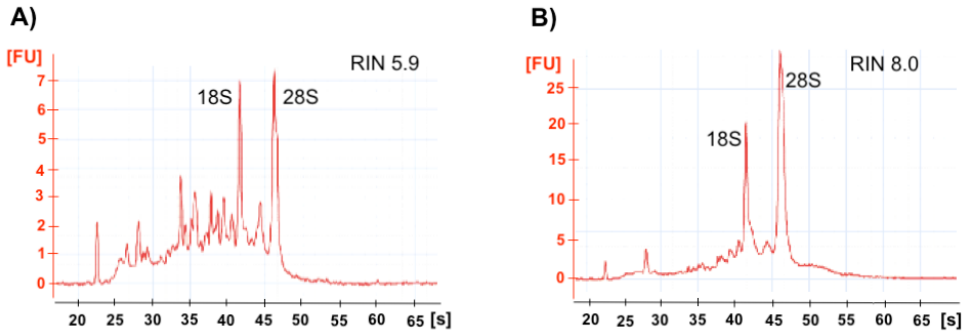

Figure 5.2: The new RNA protocol preserves RNA integrity.
Samples treated with the standard RNA isolation protocol showed a lower RNA integrity number (RIN) ***(A)*** *than the samples isolated with the new protocol* ***(B)*** *as determined by the two ribosomal subunits, 18S and 28S respectively. Note that the RIN values can be between 0-10, where 0 represents the lowest and 10 the highest integrity, thus defining RNA stability. Y-axis defined as [FU] denotes the peak fluorescence measured as 'Height threshold'. X-axis defined as [s] denotes the size of the rRNA peaks and corresponds to the time of their detection.*

Unlike the old protocol, the new protocol contained polyvinylpyrrolidone (PVP), a chemical thought to bind to phenolic compounds which are then eliminated during EtOH precipitation. In addition to one cold 70 % (v/v) EtOH precipitation, 10 mM LiCl precipitation is also required. LiCl precipitation is thought to remove cDNA synthesis inhibitors and increase RNA stability (Chan et al., 2007). Therefore, these differences contributed to the RNA integrity maintenance with the knock-on effect of improved cDNA synthesis. Although cDNA synthesis efficiency was higher for the quiescent samples treated with the new protocol, the underlying issue of gDNA contamination was also present in these samples, albeit much less than for the samples treated with the old protocol, and in both cases higher for actin than the cdc2

primer as illustrated in (**Figure 5.3**).

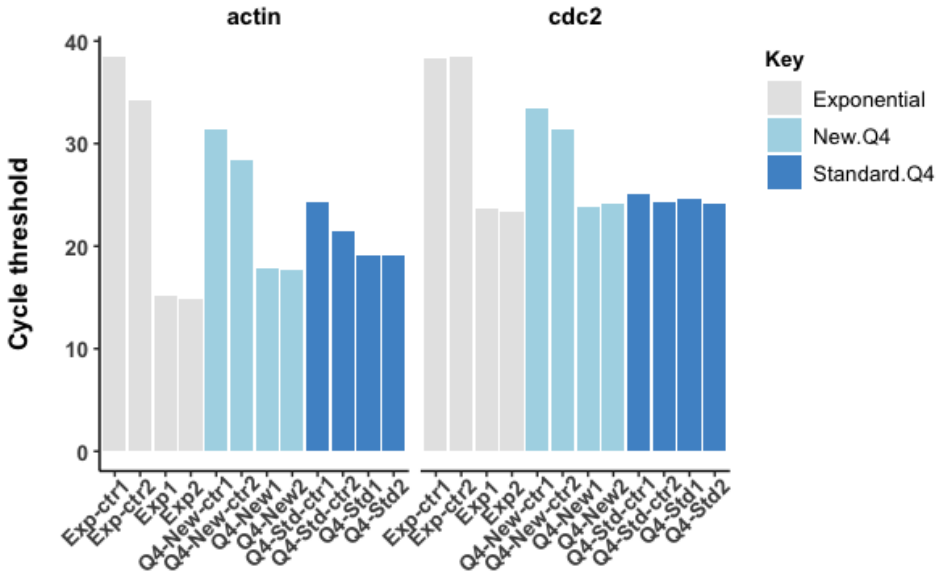

Figure 5.3: Samples treated with the new protocol show higher cDNA synthesis efficiency.

cDNA synthesis comparison between wild-type quiescent cells treated with both RNA isolation methods showed higher cDNA synthesis and less gDNA contamination for samples treated with the new protocol (New.Q4) than the samples treated with the standard protocol (Standard.Q4), where 'Q4' refers to the number of days in quiescence. qPCR was performed using actin and cdc2 primers with exponentially growing cells used as a control. Internal controls lacking the reverse transcriptase enzyme were also prepared.

5.0.4 Quiescent cells gene enrichment shows ncRNA regulation

Given that the new protocol proved effective at extracting high-quality RNA from quiescent cells, we tested the method by extracting RNA from 2 days and 21 days old quiescent cells. RIN and cDNA synthesis were both efficient, thus we processed the samples into RNA libraries using the method previously described (see **Subsection 2.2.3.4**). As over 85% of the reads mapped successfully to the genome, we performed differential gene expression (DGE) analysis using the recommended RNA-seq data analysis package, DESeq2. DGE analysis between 2 days and 21 days old quiescent samples revealed 292 up-regulated and 272 down-regulated genes as shown in (**Figure 5.4**). Of these, 43.8% and 49.0% were genes with predicted and

unknown functions, while the remaining genes were genes with known functions. We found that within the down-regulated genes with predicted function 22 were intergenic ribosomal RNAs (rRNAs), 9 were noncoding RNAs (ncRNAs) and 2 were anti-sense RNAs. Also, within the up-regulated genes with predicted function, we found only 1 intergenic RNA and 2 ncRNAs.

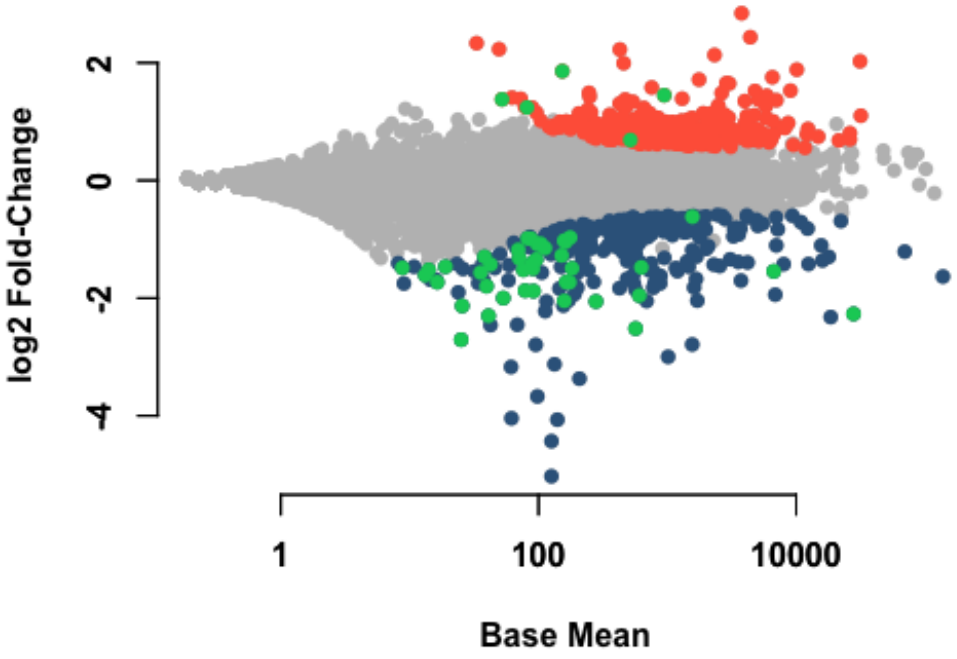

Figure 5.4: Differential gene expression of long-term quiescent cells.
Differential gene expression of 2 days and 21 days old quiescent cells revealed 292 up-regulated (orange) and 272 down-regulated (dark blue) genes. Within the up-regulated and the down-regulated genes, we highlighted (green) the 5 and the 41 ncRNAs respectively. Genes were selected based on log2FC=±1 and p-adj<0.05.

We next examined the up-regulated and the down-regulated genes for GO enrichment, a standard procedure used to obtain the functional and molecular processes of the genes provided in the query list (Bitton et al., 2015). We performed this using AnGeLi with the 'Biological Process' as our GO-term and all the 'protein-coding genes' as the background gene list. We found that while the up-regulated genes were mainly enriched for "rRNA processing" (p<5.29E-18), "ribosome biogene-

sis" (p<1.57E-16) and "ncRNA processing" (p<4.17E-13), the only enrichment obtained for the down-regulated genes was "import across the plasma membrane" (p<7.10E-4) with significant under-enrichment for "cellular component organization or biogenesis" (p<3.85E-05, data not shown). Surprised by this, we investigated further and found that 20.6% of the down-regulated genes were omitted from the analysis as these represented different RNAs not yet characterised. These RNAs were: 41 ncRNAs, 4 small nuclear RNAs (snRNAs), 4 small nucleolar RNAs (snoRNAs), 2 tRNAs, 4 pseudogenes and 1 rRNA. In contrast, the up-regulated genes contained only 5 ncRNAs which were also omitted from the analysis. These ncRNAs were also depicted in (see **Figure 5.4**).

To gain additional information about these genes, we continued our functional analysis using GO-slim, a comprehensive database available on PomBase that summarises the biological function of yeast genes (Wood et al., 2019; Lock et al., 2019). We found that while 15.3% and 8.90% of the down-regulated and the up-regulated genes, respectively, did not have any slim annotations, several statistical significant categories (p<0.01) were found as shown in (**Figure 5.5**).

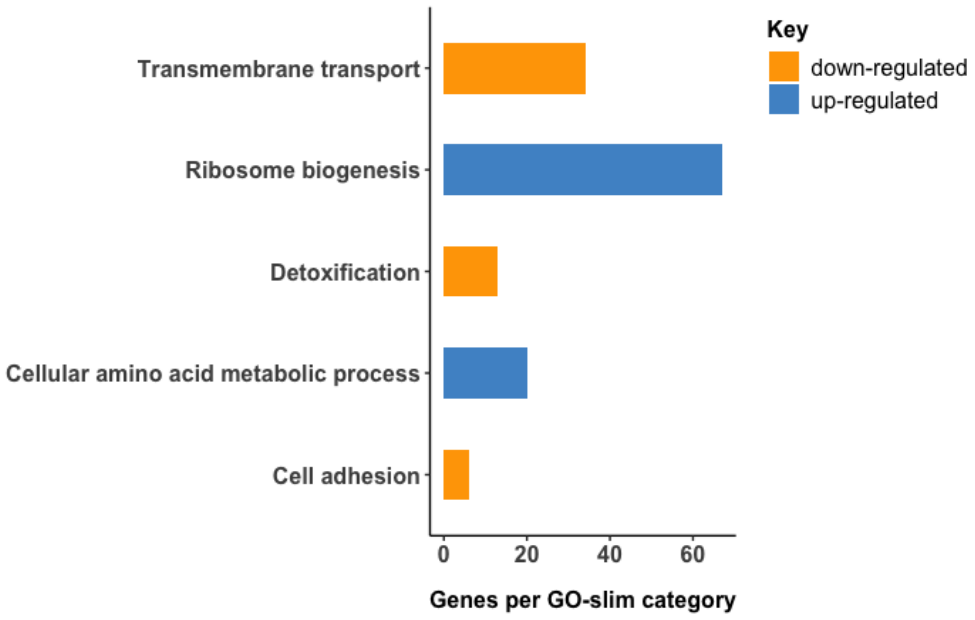

Figure 5.5: Ribosome biogenesis and trans-membrane transport are the main RNA-seq GO-slim signatures.

Up-regulated and down-regulated GO-slim categories defined at p<0.01.

Advancing with our analysis, we also used KEGG (`https://www.genome.jp/kegg/tool/map_pathway2.html`), a publicly available database that contains the molecular pathway genes of different organisms, including the fission yeast genes. Although several pathways were found, we arbitrarily present only the top 5 enriched pathways for each gene category as depicted in (**Figure 5.6**).

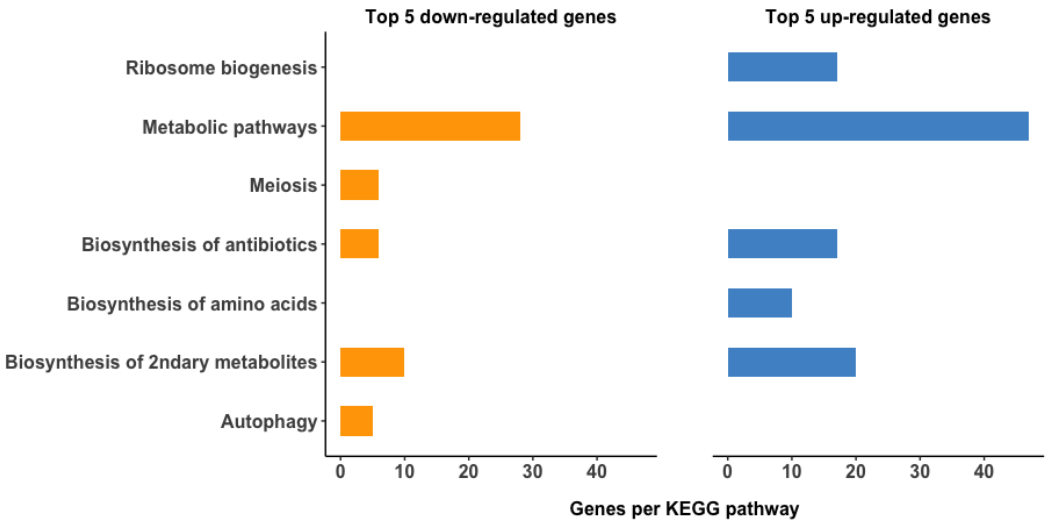

Figure 5.6: Metabolic pathway genes seem to play a role in long-term cellular quiescence.

Top 5 up-regulated and down-regulated KEGG gene enrichment pathways.

As expected, we found consistency between the enrichment analyses performed with the GO-term, GO-slim and KEGG. This comprehensive enrichment analysis enabled the identification of distinct functional pathway signatures for each gene category. These included ribosome biogenesis and trans-membrane transport as central transcriptional up-regulation and down-regulation, respectively, required for the maintenance of long-term cellular quiescence in yeast.

Chapter 6

Chronological ageing of competitively grown decoded strains

6.0.1 Background

Studying ageing in a high-throughput fashion to determine genome-lifespan relationships, mainly due to the lack of tools remained a subject not very well explored. This scarcity of high-throughput studies to identify genes important for ageing, motivated us to apply Bar-seq to profile the chronological lifespan of our decoded strains. We anticipated that the results obtained from this high-throughput study would improve our understanding of which gene deletions are important for cellular ageing, especially since 47% of these systematic gene deletions have human orthologs (Lock et al., 2019). Because the methods to perform such genome-wide studies are not very well defined, screening for gene deletions relevant for lifespan regulation are almost non-existent, even in budding yeast (Fabrizio et al., 2010) with Bar-seq studies in fission yeast being just as elusive.

To our knowledge, however, only one fission yeast Bar-seq study was previously performed. Unlike our screen, this study used version 3.0 of the deletion library to screen for gene deletion mutants with pro-ageing effects following long-term quiescence survival (Sideri et al., 2015). Since during quiescence cells arrest at a G0-like cell-cycle phase and during the glucose-starvation model cells arrest at the G2-phase of the cell-cycle, the mechanisms responsible for life extension, thus are different

and so the gene candidates obtained from these two models are likely very different also. Therefore, our glucose-starvation screen will not only be complementary to the quiescence screen but will provide new insights about gene deletions important for cellular ageing. Furthermore, the difference in methods, where we use our improved Bar-seq and our custom-developed data analysis pipeline, in addition to the most up-to-date fission yeast deletion library comprising approximately 400 more strains than the previous library, should further substantiate the significance of our chronological lifespan Bar-seq screen.

6.0.2 Experimental design

Three independent prototroph deletion library pools made using the method previously described (see **Subsection 2.2.1**) were grown in parallel and cultured as described in (**Figure 6.1**).

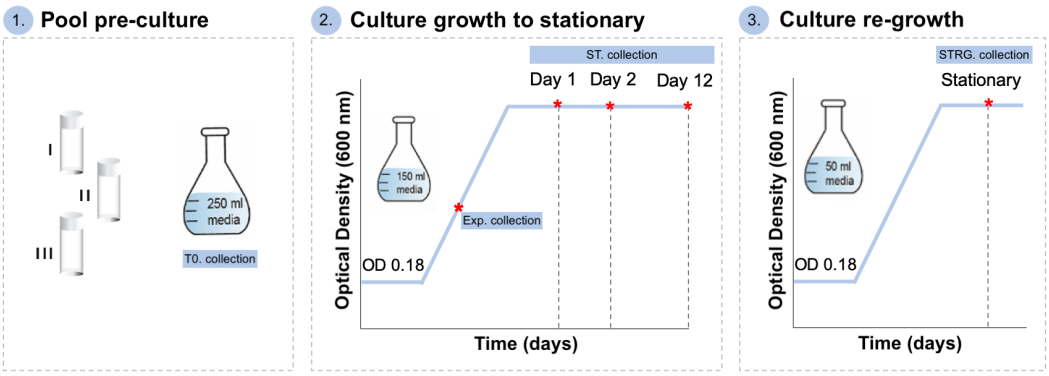

Figure 6.1: Chronological ageing experimental set-up.
Prototroph deletion library pool aliquots of 500 µl were used to prepare individual pre-cultures in 250 ml YES (3% glucose) grown in parallel o/n at 25°C with no shaking ***(1)****. The following morning, T0 samples were collected from each culture. The pre-cultures were then used to inoculate cultures at 0.18 OD_{600nm} in 150 ml YES (3% glucose) and grown at 32°C and 170 RPM to saturation density, during which exponential samples were also collected. Following saturation density, the first time point (e.g. day 1) where 100% viability is assumed was collected. Subsequent sample collections, CFUs and culture re-growths were prepared daily and where possible at the same time* ***(2)****. Culture re-growths were prepared at 0.18 OD_{600nm} in 50 ml YES (3% glucose) and grown under the same conditions as the main cultures to saturation density where samples were collected* ***(3)****. Note, sample collection was made as* 2×20 *aliquots.*

In total, 96 libraries ($7 collectionTimePoints \times 3 pools \times 2 cultures \times 2 barcodes$) plus T0 and exponential samples per pool and barcode were prepared using the method previously described (see **Subsection 2.2.3.3**). Libraries were sequenced first, on a NextSeq, and then on a MiSeq Illumina instrument as paired-end reads

of 75 bp using 165 cycles. NextSeq was unsuccessful due to low library complexity, hence MiSeq was run on selected samples to obtain enough reads for all the samples. The reads were analysed with Barcount, our custom-developed Bar-seq analysis pipeline previously described (see **Subsection 2.3.2**). Reads per sample and barcode from both runs were combined and are shown in (**Table 6.1**).

Table 6.1: Chronological lifespan sample names and sequencing depth per barcode.

Sample name	Uptag St	Dntag St	Uptag RG.St	Dntag RG.St
T0.I	0.27	0.72	-	-
T0.II	0.52	0.38	-	-
T0.III	0.39	0.44	-	-
Exp.I	0.55	0.33	-	-
Exp.II	0.42	0.36	-	-
Exp.III	0.26	0.58	-	-
Day1.I	0.28	0.63	0.33	0.61
Day1.II	0.27	0.71	0.42	0.61
Day1.III	0.24	0.47	0.29	0.63
Day3.I	0.22	0.76	0.55	0.39
Day3.II	0.51	0.49	0.41	0.44
Day3.III	0.22	0.42	0.44	0.26
Day4.I	0.20	0.39	0.49	0.68
Day4.II	0.46	0.35	0.65	0.35
Day4.III	0.28	0.31	0.48	0.33
Day6.I	0.30	0.77	0.22	0.27
Day6.II	0.14	0.67	0.64	0.28
Day6.III	0.27	0.39	0.24	0.30
Day8.I	0.32	0.49	0.27	0.72
Day8.II	0.43	0.50	0.55	0.43
Day8.III	0.24	0.75	0.13	0.05
Day10.I	0.23	0.26	0.04	0.10
Day10.II	0.31	0.45	0.12	0.18
Day10.III	0.29	0.54	0.11	0.07
Day12.I	0.27	0.45	0.29	0.30
Day12.II	0.35	0.33	0.13	0.21
Day12.III	0.21	0.39	0.12	0.13

Sample names denote the collection time points where 'T0' stands for pre-culture cells, 'Exp.' for exponential cells, days 1-12 stand for the days spent in the stationary phase, and I, II and III are the three pool replicates. 'St' and 'RG.St' stand for samples collected at the stationary and the re-growth stationary phase. Reads per million were provided for each sample time point at the stationary and the re-growth stationary for each barcode. Note, there were only one T0 and exponential sample collection per pool. These reads were added under the barcodes of the stationary culture.

6.0.3 Chronological ageing with Bar-seq requires culture re-growth

In our pool experiments, we found consistency in the population viability with our previous observations, where the viability of the pool declined at a faster rate compared to the rate of decline observed for the wild-type cell cultures (**Figure 6.2**).

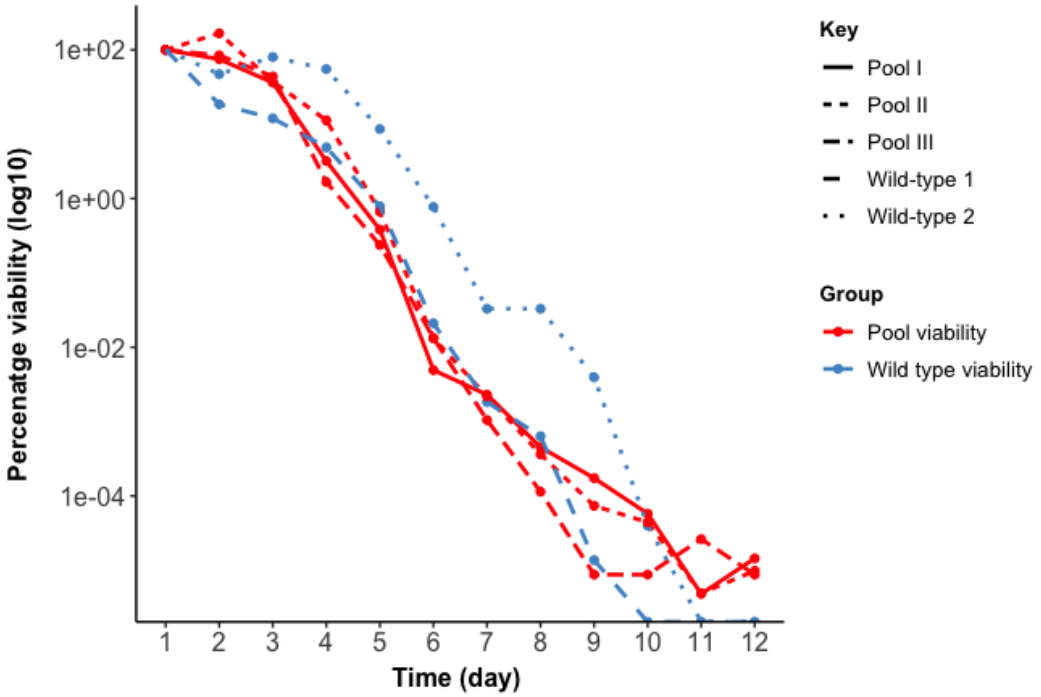

Figure 6.2: Competitive strain pool growth shows reduced cell viability.
Population viability curves for the wild-type and the decoded strains cultures. Percentage viability calculated for each of the three biological pools and the two wild-type repeats is shown on a log scale.

Chronological ageing studies from budding yeast (Fabrizio et al., 2010; Matecic et al., 2010) included cellular re-growth before DNA extraction to avoid bias from the dead cells. Our preliminary chronological ageing data on wild-type cells also showed that the DNA of the dead cells is maintained for several days following cell necrosis. For an explanation of this finding (see **Figure 6.4**). However, to avoid this DNA bias, we also performed cellular culture re-growths, and found that during the culture re-growth the time taken by the cells to saturate the cultures positively correlated with the time the cells spent in the stationary phase of the aged culture (data not shown). This observation seemed sensible as the number of viable cells

over time is expected to decrease linearly with the population viability, and thus the remaining cells would take longer to saturate the same volume of culture not only because there are fewer cells, but due to delayed signalling as a result of cells being more dispersed.

We next examined the reads and found that both the read assembly and their uptag and dntag barcode matching to the reference barcode database was more efficient for the stationary growth samples than the stationary re-growth samples as shown in (**Figure 6.3**).

Although among the assembled reads for the re-growth samples, one sample (e.g. Day8.III) in particular stood out as having the lowest read assembly efficiency, 80% of these reads contained barcodes that matched successfully to the reference barcode database. The opposite relationship of high read assembly and low dntag barcode matching to the reference barcode database was observed for yet another re-growth sample (e.g. Day8.II). However, we found that the number of gene counts for both of these samples was high enough to sequence each gene over 200 times, and thus twice the recommended sequencing depth (Han et al., 2010).

Since we found that the DNA of the dead cells is maintained for several days following cell necrosis even when 99% of the cells lost viability as shown in (**Figure 6.4 A**), we wondered whether this phenomenon featured in our data. Indeed, we found that while the total number of genes for the stationary growth samples remained constant throughout the time points, the re-growth samples showed the expected decrease in the total number of genes; a phenomenon consistent across the time points, but only up to day 6 as shown in (**Figure 6.4 B**).

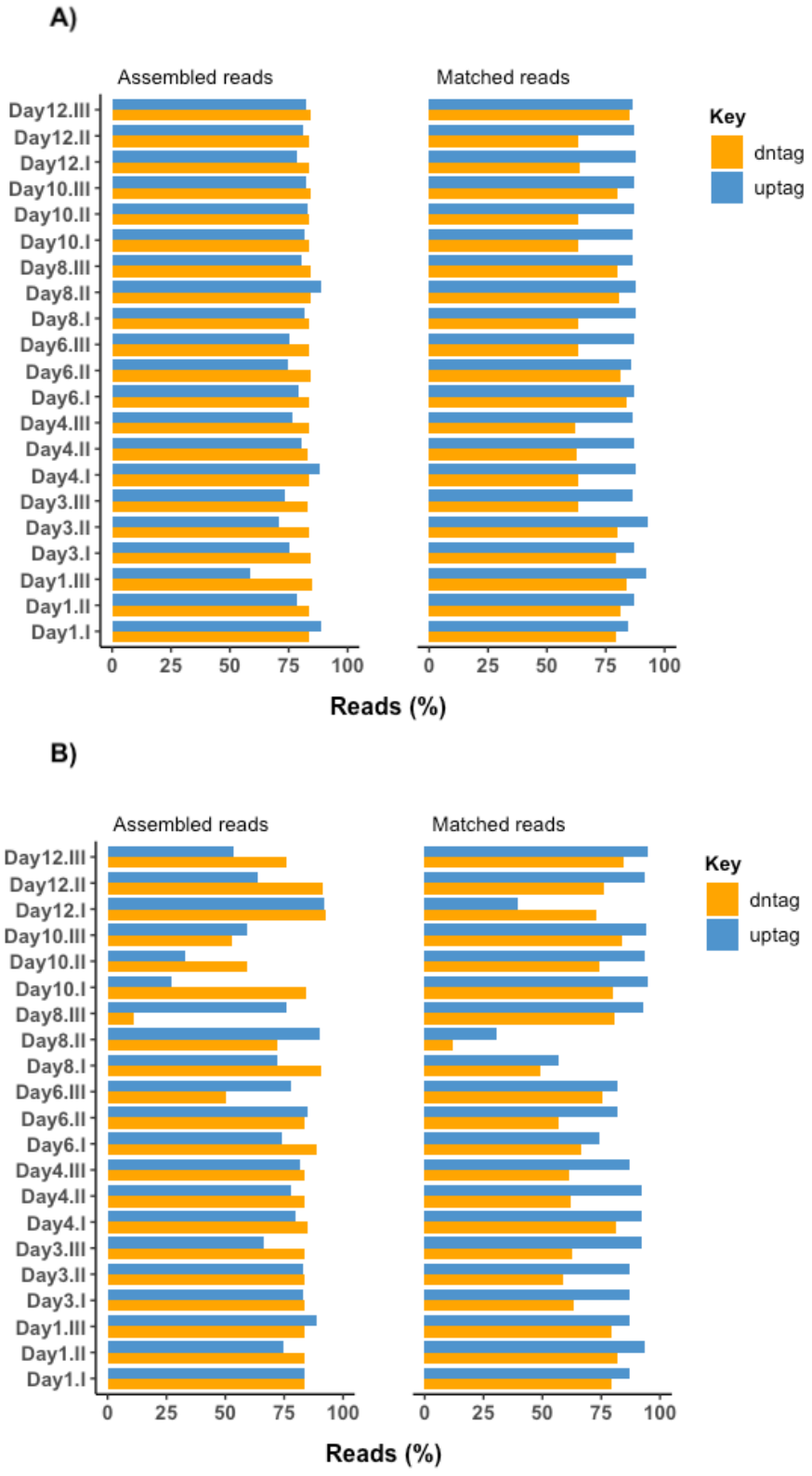

Figure 6.3: Uptag reads matched more efficiently to the reference barcode database than the dntag reads.

Percentage of assembled reads and the proportion of the reads which matched to the reference barcodes for the stationary growth ***(A)*** *and the stationary re-growth samples* ***(B)****.*

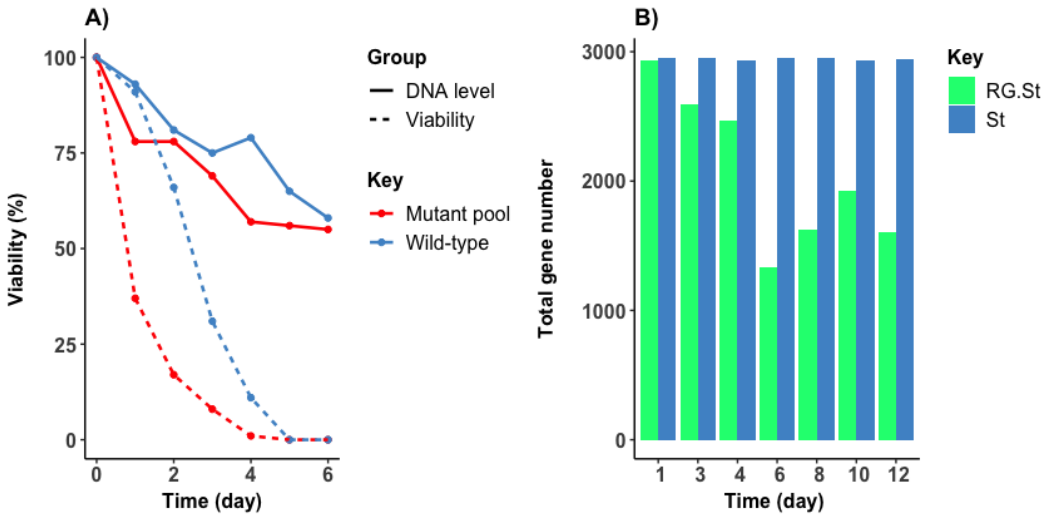

Figure 6.4: DNA persistence following cell death impacts the sequencing depth.
DNA maintenance following cell death ***(A)*** *and the total number of post-filtered genes with at least 5 counts for the stationary growth (St) and the stationary re-growth samples (RG.St),* ***(B)****. DNA level and cell viability were based on three biological repeats. DNA level and cell viability were measured using a Qubit instrument and the CFUs method, respectively.*

Contrary to our expectation, we found that relative to the day 6 samples, day 8, 10 and 12 re-growth samples showed a slightly higher and roughly constant number of genes. We discovered that this was caused by an increase in the DNA carryover where more volume of mainly dead cells was inoculated to compensate for the few surviving cells at the later time points. Consequently, the DNA of these cells was amplified and sequenced, hence contributing to the increase in the total number of genes observed for samples a day 8, 10 and 12.

Consistent with this finding, the raw gene counts across the time points for both the stationary and the re-growth stationary samples recapitulated the phenomenon (**Figure 6.5** and **Figure 6.6**). Considering that the number of viable cells is expected to decrease linearly with the decrease in population viability and that only a few surviving cells that display a fitness advantage will dominate the culture, a decrease in the total gene number as a function of time is expected, thus supporting our observation. Analysis of the stationary samples whereby the total number of genes across the collected time points remains unchanged, or in other words, analysis of these saturated samples would lead to very subtle differences which in reality may actually be large changes. Therefore, Bar-seq barcode abundance obtained from the

stationary growth samples would be misleading, hence the re-growth data is more suitable for the analysis over the stationary growth data.

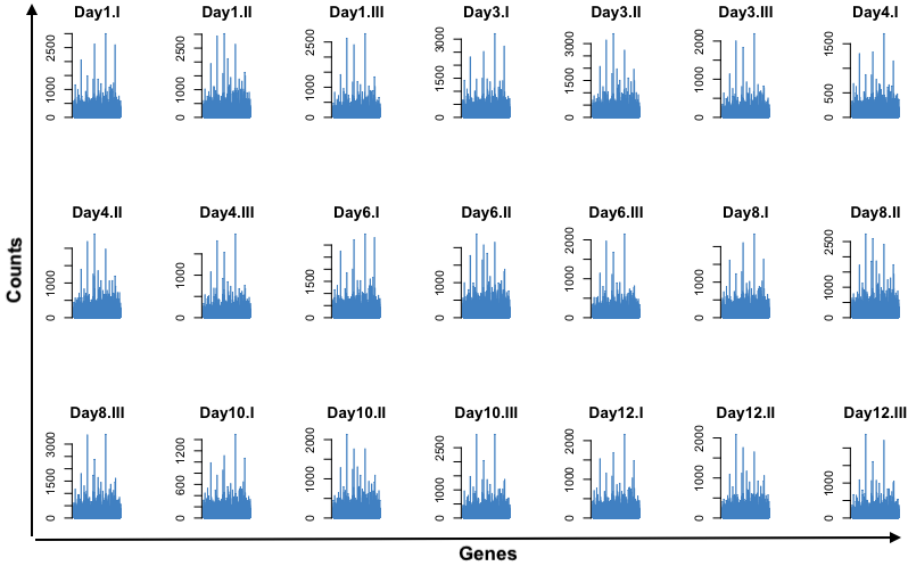

Figure 6.5: Over time the growth samples show no change in gene counts.
Sample gene counts for the stationary growth samples with no overall change in the gene number. Day 1-12 represent the days spent in the stationary phase while I,II and III refers to the pool replicates.

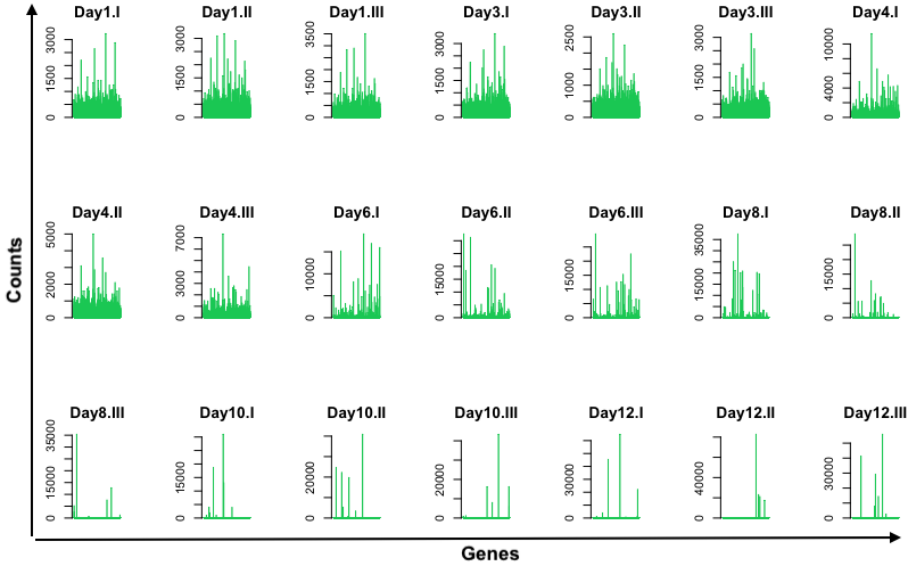

Figure 6.6: Over time the re-growth samples show a decrease in gene counts.
Sample gene counts for the stationary re-growth samples with a downward trend in the gene number throughout the time points. Day 1-12 represent the days spent in the stationary phase while I,II and III refers to the pool replicates.

6.0.4 Re-growth data analysis identified more short-lived than long-lived mutants

Similar to our previous finding, gene count sample correlations showed sample similarities across time points, but only up to day 6 as the later time points (e.g. day 8, 10 and 12) were too different from the early time points (e.g. day 1, 3, 4 and 6) as shown in (**Figure 6.7**).

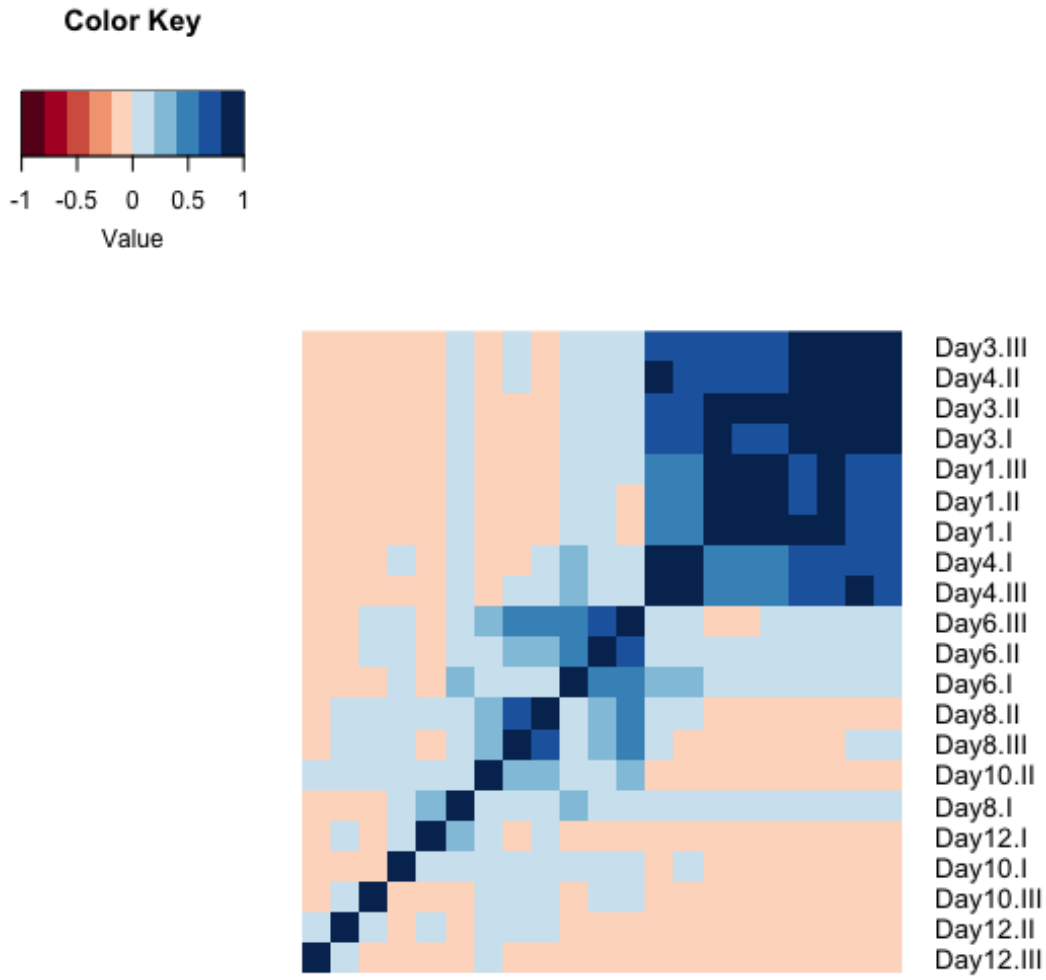

Figure 6.7: Chronologically aged re-growth samples show strong correlations but only up to day 6.
Gene count correlation between the re-growth samples time points was calculated using the cor() function and plotted with the heatmap.2 package using R (version 3.5.0).

Given that the later time points (e.g. day 8, 10 and 12 samples) were too different from the rest of the samples, we decided to investigate further. We found that at these time points the number of reads was lower than the number of inoculated

cells as shown in (**Table 6.2**). Therefore, inclusion of these samples in the analysis due to the large discrepancy in the number of genes between the early time points where there are thousands of genes, and the later time points where there are only tens of genes, would result in genes with over-inflated p-values, hence no true hits would be produced. Also, no available statistical packages can model this type of data, thus day 8, 10 and 12 samples were excluded from further analysis.

Table 6.2: The number of cells used to inoculate the re-growth cultures.

Pool	Time point	CFUs/ml	Inoculate/ml	Live cells
I	1	5.30E+07	0.65	34,700,000
II	1	5.70E+07	0.71	40,400,000
III	1	6.00E+06	0.71	4,250,000
I	3	9.00E+07	0.86	77,700,000
II	3	9.20E+07	0.82	75,300,000
III	3	4.00E+07	0.82	32,700,000
I	4	5.20E+04	0.86	4,490,000
II	4	1.78E+05	0.79	14,100,000
III	4	1.80E+04	0.79	1,430,000
I	6	1.66E+02	0.83	13,800
II	6	2.06E+02	0.91	18,700
III	6	7.60E+01	1.00	7,600
I	8	1,280	1.05	1,340
II	8	1,080	0.83	900
III	8	260	0.95	247
I	10	120	0.95	114
II	10	80	1.05	84
III	10	20	1.05	21
I	12	60	0.90	54
II	12	20	0.95	19
III	12	20	0.95	19

Since day 1 re-growth samples were most similar to the rest of the re-growth samples, day 1 samples were the reference time point used for the Bar-seq barcode abundance analysis. The analysis was performed using the recommended Bar-seq data analysis package, DESeq2 (version 1.20.0), where we identified 3,110 genes or 97% of the decoded strains in total. The long-lived mutants were identified using the relative difference between the day 1 and day 6 samples as well as using the time course approach for day 1, 3, 4 and 6 samples. Since day 3 was the latest

time point relevant for the identification of short-lived mutants, these were identified using only the relative difference between the day 1 and the day 3 samples. The Bar-seq barcode abundance analysis results are shown in (**Figure 6.8**).

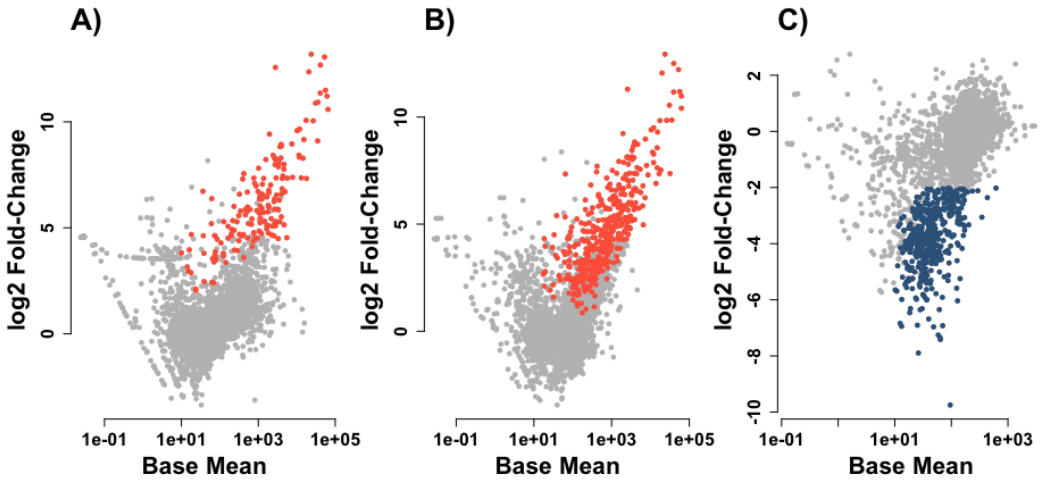

Figure 6.8: The relative difference method was more suitable over the time course analysis.

The long-lived mutants were obtained using the relative difference at day 1 and day 6 samples ***(A)*** *and the time course approach between day 1, 3, 4 and day 6 samples* ***(B)****. The short-lived mutants were identified only by using the relative difference between day 1 and day 3 samples* ***(C)****. DESeq2 analysis was performed on raw gene counts with genes selected at log2FC=±2 and p-adj <0.01.*

In comparison to the relative difference, the number of long-lived mutants obtained from the time course analysis was two-fold higher. However, the relative difference analysis was more suitable for the identification of the long-lived mutants as all of these mutants were also included in the list of the long-lived genes obtained with the time course approach, albeit with more stringent statistics and overall the same biology. Therefore, the relative difference method was used to identify the 168 long-lived and the 441 short-lived mutants detected from our chronological lifespan Bar-seq screen. The difference in the gene number between the short-lived and the long-lived mutants reflects the expected higher number of deleterious over advantageous gene deletions, thus supporting our Bar-seq barcode abundance analysis and the relative difference method used to identify the short-lived and the long-lived mutants.

6.0.5 Enrichment of differentially expressed genes suggests the identification of age-relevant genes

To identify the molecular processes the differentially expressed genes were enriched for, we performed GO analysis using the 'Biological Process' as the GO-term and our decoded genes as the background gene list. The processes were selected based on the statistical significance of $p<0.001$ and $p<0.005$ for the short-lived and the long-lived mutants, respectively, as shown in (**Figure 6.10**). This statistical significance was selected to ensure the inclusion of as many processes as possible while retaining a high level of statistical power. However, because this selection resulted in a few molecularly enriched processes, we further queried our genes for molecular pathway enrichment using KEGG, the results of which are shown in (**Figure 6.11**).

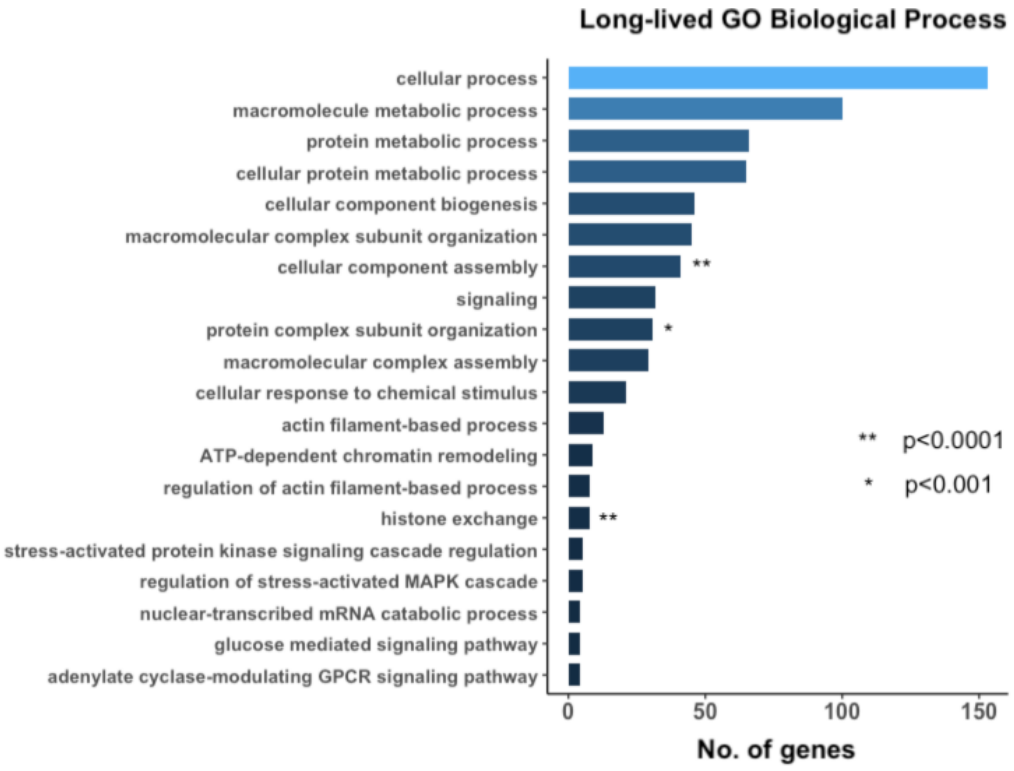

Figure 6.9: Long-lived GO-term enrichment.
The biological processes of the long-lived genes obtained as a result of GO enrichment analysis performed with 3,824 two-sided Fishers exact tests and corrected according to FDR with an alpha of 0.01.

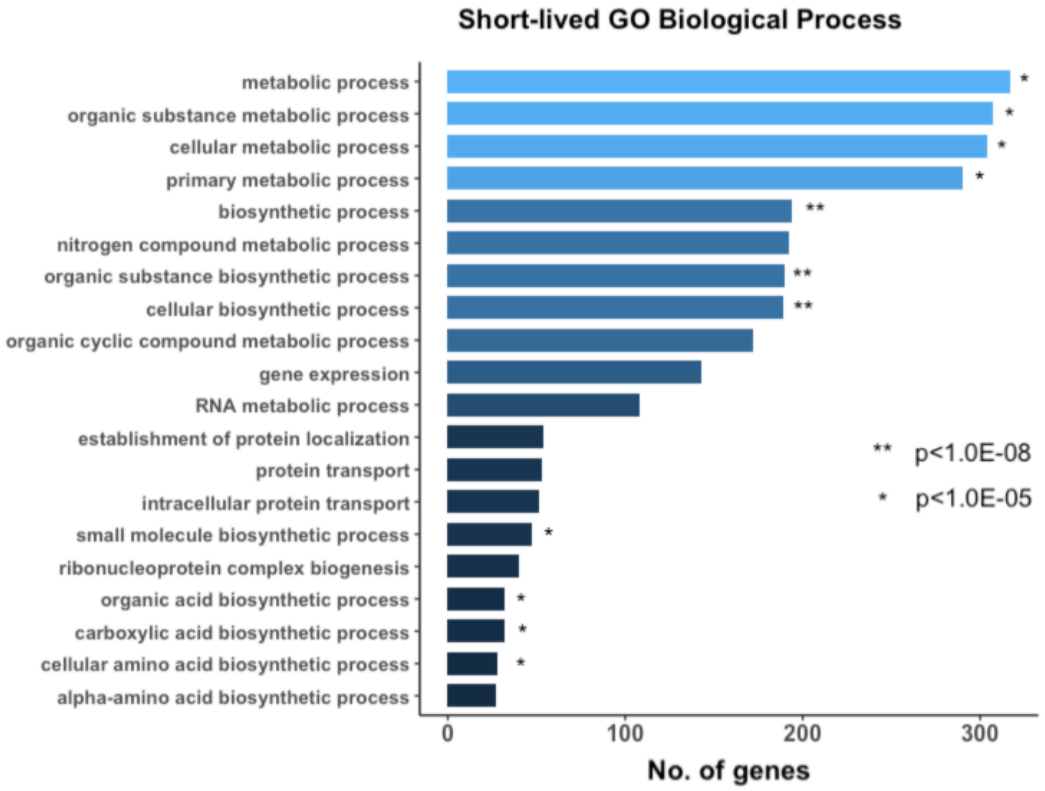

Figure 6.10: Short-lived genes GO-term enrichment.
The biological processes of the short-lived genes obtained as a result of GO enrichment analysis performed with 3,824 two-sided Fishers exact tests and corrected according to FDR with an alpha of 0.01.

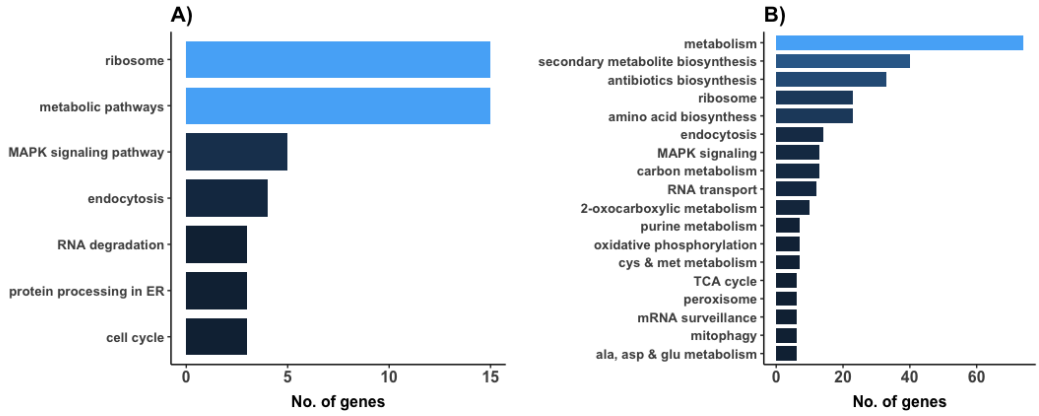

Figure 6.11: Long-lived and short-lived genes KEGG enrichment.
Metabolic pathway enrichment of the long-lived ***(A)*** *and the short-lived* ***(B)*** *gene deletion strains. The top association pathways were defined by at least 3 and 6 genes for the long-lived and the short-lived genes, respectively.*

The GO-term and the KEGG enrichment were both complimentary and in agreement. For example, while the main enriched biological processes for the long-lived mutants were catabolic metabolism, signalling response and genome re-organisation, KEGG showed metabolic enrichment for pathways such as, RNA degradation and protein processing, MAPK signalling and cell cycle; all of which are associated with processes known to promote lifespan (Smith et al., 2018; Slack, 2017; Borbolis and Syntichaki, 2015; Okuyama et al., 2010). A similarly strong resonance between GO-term and KEGG enrichment was also observed for the short-lived mutants. For instance, while the main biological processes represented the biosynthesis of molecular components, KEGG showed enrichment for amino acid metabolism and secondary metabolites. These processes seem to reflect the mutants desire for anabolism similar to the 'live fast die young' concept representative of short-lived cells (Jimenez et al., 2015). This finding thus suggests identification of relevant gene candidates.

Furthermore, we found that the relative difference between the short-lived annotated and unannotated KEGG genes was two-fold higher than the relative difference between the long-lived annotated and unannotated KEGG genes as shown in (**Figure 6.12**). Therefore, more long-lived gene deletions lack KEGG annotations than short-lived genes, implying that our long-lived gene deletions corresponded to genes with unknown function, not previously explored or studied. Indeed, 41.5% and 42.3% of the short-lived and long-lived gene deletions respectively, were genes with predicted and unknown functions, hence supporting the idea that our Bar-seq analysis identified relevant age-associated genes, and possibly novel gene candidates as well.

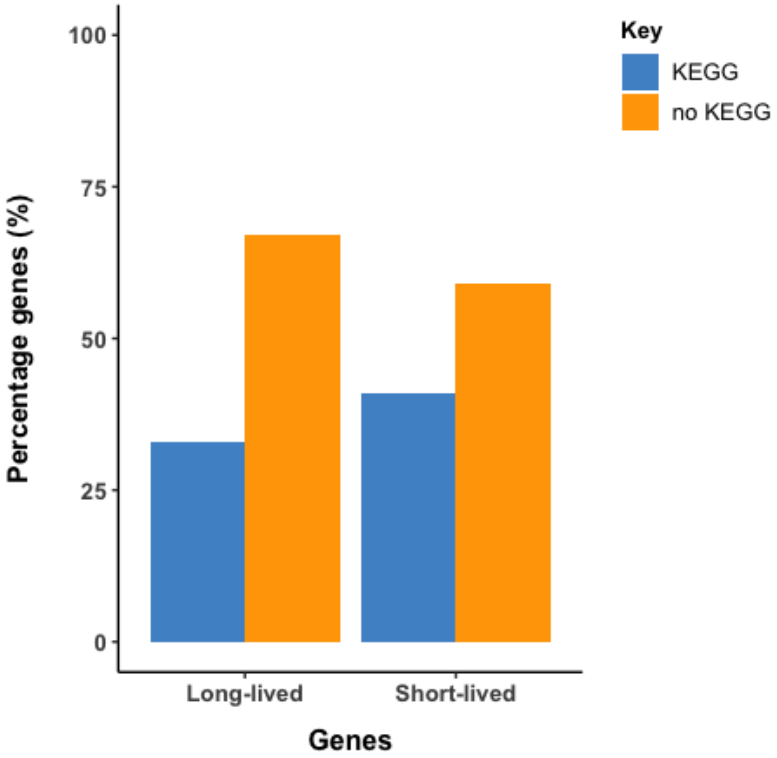

Figure 6.12: The long-lived deleted genes have fewer KEGG annotations than the short-lived deleted genes.

The number of long-lived and short-lived genes with and without KEGG annotations.

Chapter 7

Isogenic strain growth to validate CLS lifespan

7.0.1 Background

The CLS of the top short-lived and the long-lived gene deletion mutants obtained from our Bar-seq analysis were validated by growing the mutants in isogenic cultures. Since our focus was on the identification of gene deletions with pro-ageing effects most of the validated mutant strains were long-lived. The gene candidates were selected based on statistical significance described by fold-change (FC) and p-adjusted values, plus novelty criteria defined by unexplored, or unstudied gene characterisations obtained from PomBase (`https://www.pombase.org/`). Applying the above criteria, two groups of eight long-lived gene deletion mutants were made: ”Bar-seq top long-lived” and ”Bar-seq unexplored long-lived”, as mutants defined by top FC and p-adjusted values, and a modest FC and p-values and unexplored genes, respectively. As for the short-lived candidates, we decided to select eight mutants with the top FC and unexplored gene status. This group was referred to as ”Bar-seq unexplored short-lived”. Because fewer genes than the number of reads prevented the use of the day 10 and day 12 samples in the Bar-seq barcode abundance analysis, we also selected the top two/three mutants from each pool to include in our validation assay. These two groups of eight mutants each were referred to as day 10 and day 12 long-lived mutants.

In addition to these groups, we also included another group of six genes which apart from the wild-type, included three long-lived controls and two short-lived controls. *Git3* (G-protein coupled receptor), *pyp1* (MAPK tyrosine phosphatase) and *tco89* (TORC1 subunit) were the long-lived controls and *sdh1* (TCA, succinate dehydrogenase gene) and *coq5* (mitochondrial, c-methyltransferase gene) were the short-lived controls.

7.0.2 Experimental design

In total, we assayed the CLS of 48 mutants in parallel. These were divided into six plates of eight mutants each, including, Bar-seq top long-lived, Bar-seq unexplored long-lived, Bar-seq unexplored short-lived, day 10 long-lived, day 12 long-lived and controls. The lifespan of each mutant was determined using the method described in (**Figure 7.1**). For details about the method and the analysis (see **Subsection 2.3.5**).

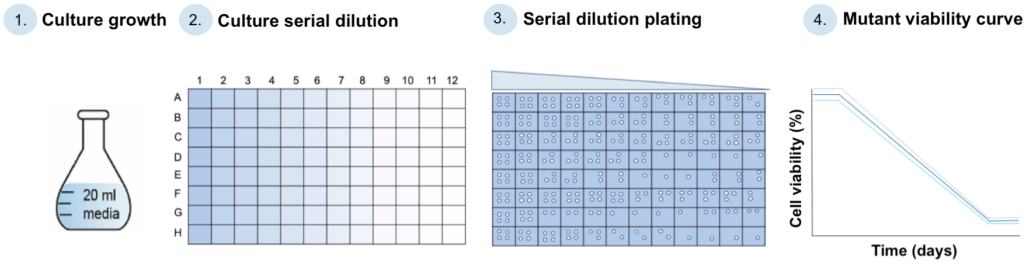

Figure 7.1: Mutant CLS validation experimental set-up.
A prototroph deletion library was thawed and arrayed onto 9×384 *PlusPlates (Singer Instruments, Watchet, UK) containing YES medium. The plates were incubated for 2 days at 32°C prior to manually selecting the mutants of interest (except for the wild-type which was obtained from an independent glycerol stock) to set-up individual pre-cultures in 50 ml flasks containing 20 ml YES (3% glucose) grown o/n at 32°C and 170 RPM. These were then used to inoculate cells at 0.002* OD_{600nm} *in 50 ml flasks containing 20 ml YES media (3% glucose) and grown at 32°C and 170 RPM to saturation density, where the first time point (e.g. day 0) was collected* ***(1)****. Sample collection consisted of 150 µl aliquots from each culture with cells transferred into the first well of a fresh 96-well Corning® plate (Sigma-Aldrich, Dorset, UK). The plates were then passed onto the ASSIST PLUS liquid handling robot (INTEGRA, Berkshire, UK) to serially dilute the cells as 50 µl in a total volume of 150 µl fresh YES medium* ***(2)****. The dilutions were plated using the RoToR HAD robot (Singer Instruments, Watchet, UK) as four independent technical replicate spots onto 384-PlusPlates containing solid YES* ***(3)****. The plates were then incubated upside-down at 32°C until colonies appeared. These were then stored at 4°C until the colonies from the plates across the time points were ready to be scanned using the Epson Perfection V800 Photo scanner (ice technologies, London, UK). Colonies were analysed using a custom-script (Townsend et al., manuscript in prep.) generating viability curves as colony forming units per droplet for each strain at each time point* ***(4)****. The steps were repeated for each mutant culture every two days and up to day 12 (e.g. day 0, 2, 4, 6, 8, 10 and 12).*

7.0.3 Of the Bar-seq validated mutants, 90% showed the expected lifespan observed from the competitive pool growth

Out of the 48 validated mutants, only 30 were Bar-seq derived comprising both the short-lived and the long-lived mutants. The remaining 18 validated mutants were not found in our Bar-seq analysis as these included the 16 strains from day 10 and day 12 samples, the wild-type which does not have any molecular barcodes and the *tco89* long-lived control mutant. Despite being one of the decoded genes identified in the Bar-seq screen, the *tco89* mutant was abundant enough at day 3 not to be included as a short-lived mutant, while its minimal abundance at day 6 prevented it from being selected as a long-lived candidate. Therefore, even though *tco89* is a known long-lived mutant, hence its use as a long-lived control in our validation assay, it was not a long-lived mutant identified by our Bar-seq analysis.

To compare our lifespans, other than the wild-type control, we also included control sets of known short and long survival fission yeast cells. These allowed us to compare the short-lived and long-lived sets of genes directly and independently to the wild-type. We found that while the wild-type lifespan in our assay was consistent with previous results obtained independently and using the same method, both control types displayed the expected lifespans (**Figure 7.2 A**). For instance, in comparison to the wild-type, the *coq5* and the *sdh1* short-lived control mutants showed lower lifespans, while relative to the wild-type, the *pyp1*, the *tco89* and the *git3* long-lived controls showed higher lifespans. This finding and the fact that both control types exhibited the expected lifespan phenotype relative to the wild-type, suggests that the wild-type relative to the mutants in the pool shows an average lifespan, thus enabling direct mutant to wild-type comparisons.

Furthermore, all of the Bar-seq short-lived validated strains displayed the expected short lifespan phenotype when compared to the wild-type, thus providing 100% lifespan consistency with the lifespan obtained from growing the mutants competitively in a pool as shown in (**Figure 7.2 B**).

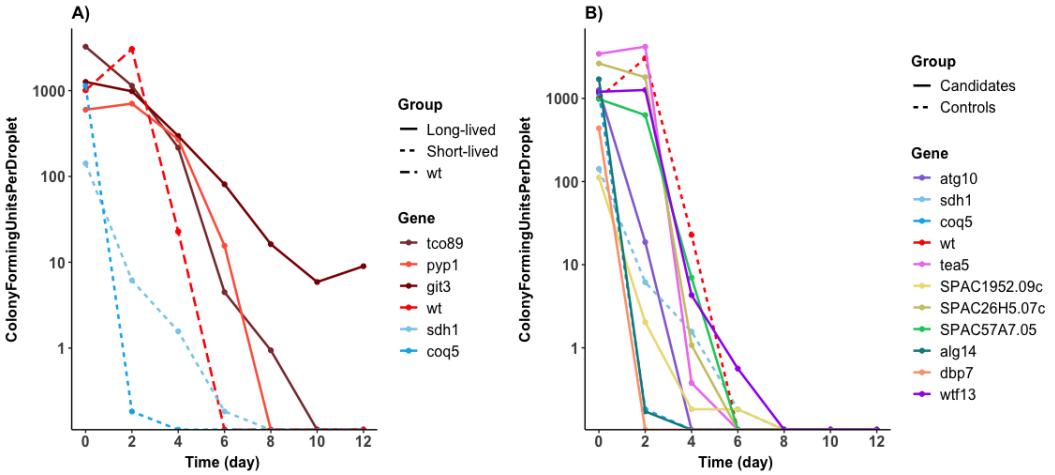

Figure 7.2: All validated control and Bar-seq short-lived mutants display the expected lifespan phenotype.

*CLS of the validated controls **(A)** and Bar-seq short-lived mutants **(B)**. Mutants lifespan curves were defined as the number of colony forming units per droplet at each time point.*

Of the validated long-lived mutants, we found that while 100% of the Bar-seq top long-lived mutants recapitulated the lifespan observed from growing the mutants in competition, only 75% of the Bar-seq unexplored long-lived validated mutants showed the expected long life phenotype as depicted in (**Figure 7.3**). For the long-lived Bar-seq mutant list, including the CLS validated mutants (see **Table D.1**).

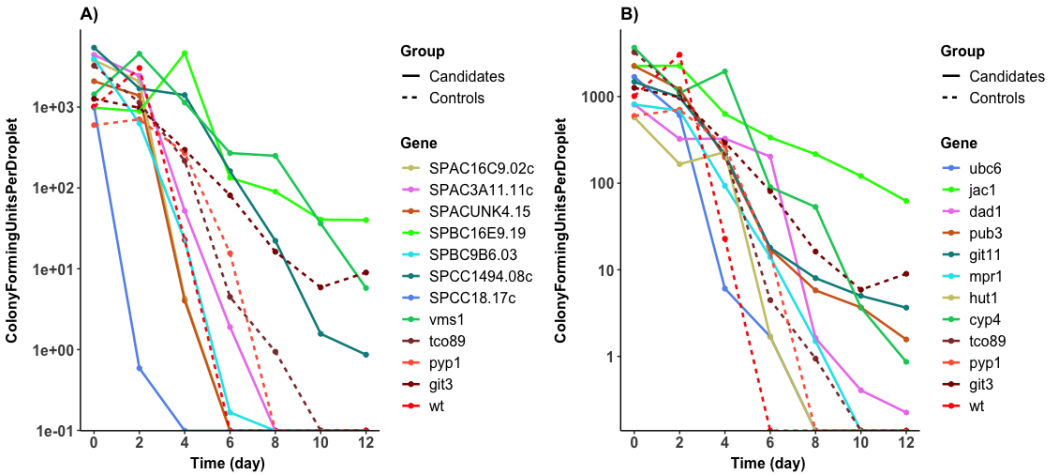

Figure 7.3: Validated Bar-seq top and unexplored long-lived mutants show 100% and 75% lifespan recapitulation to the competitively grown mutants.

*Individually validated unexplored **(A)** and top **(B)** long-lived Bar-seq mutants with CLS lifespan curves defined as the number of colony forming units per droplet at each time point.*

The results, therefore, show that 90% of the Bar-seq validated mutants displayed the expected lifespans, thus confirming our Bar-seq analysis.

7.0.4 Only 25% of day 10 and day 12 long-lived mutants showed the expected lifespan observed from the competitive pool growth

Of the 16 validated strains representing the top mutants that dominated the growth cultures at day 10 and day 12, in comparison to the wild-type, only two mutants from each time point showed the expected long lifespan phenotype as observed from the competitive pool growth. Therefore, 75% of these mutants showed the opposite lifespan to the one expected (**Figure 7.4**).

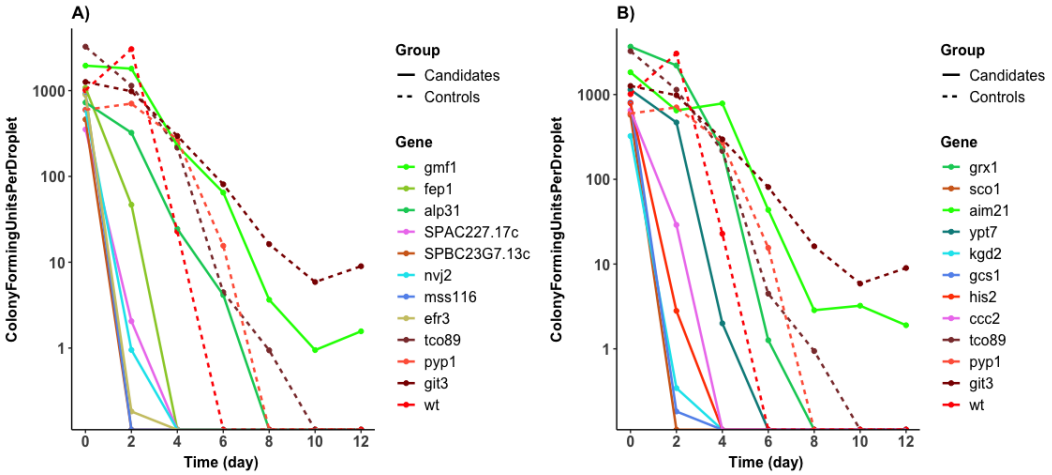

Figure 7.4: Most of the validated day 10 and day 12 long-lived mutants show lifespans opposite to what was observed from the competitive pool growth.
*Individually validated day 10 **(A)** and day 12 **(B)** long-lived mutants with CLS lifespan curves defined as the number of colony forming units per droplet at each time point.*

7.0.5 Validation of the top 10 long-lived mutants distinguishes several novel longevity genes

The top long-lived mutants were pooled together to compare their lifespans relative to the wild-type and each other. The selection was based on the mutant lifespans greater than the wild-type, the *pyp1* and the *tco89* long-lived controls. These top 10 mutants included 5 genes (*jac1, dad1, pub3, git11 and cyp4*) from the Bar-seq top long-lived list, 3 genes (*vms1, SPBC16E9.19 and SPCC1494.08c*) from the Bar-seq unexplored long-lived mutant list, with *gmf1* and *aim21* gene from day 10 and day 12 samples, respectively.

Analysis of these 10 long-lived candidates revealed 7 genes (*vms1, jac1, SPBC16E9.19, SPCC1494.08c, gmf1, cyp4, and pub3*) with identifiable human orthologs (Lock et al., 2019). We also found that the functional impairment of either *cyp4*, a peptidyl-prolyl cis-trans isomerase or *pub3*, a predicted HECT-type ubiquitin-protein ligase E3 has reportedly been linked to the development of osteogenesis imperfecta type 2, 3, 4 and 9 and kidney disease in humans (Rappaport et al., 2017). Additionally, CLS comparison between the top 10 mutants revealed three genes (*jac1, vms1 and SPBC16E9.19*) with absolute maximum lifespans as shown in (**Figure 7.5**).

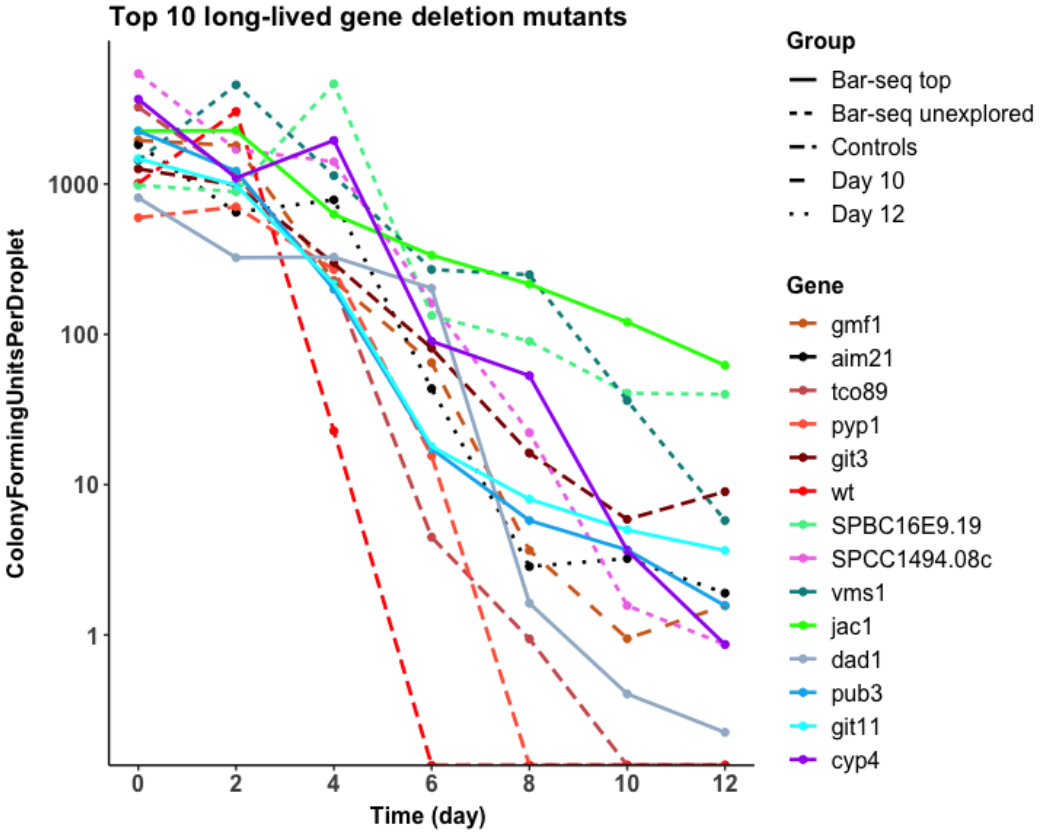

Figure 7.5: Comparison of long-lived mutants with top lifespans revealed novel longevity genes.

Lifespan curves of the top validated long-lived mutants with CLS greater than the wild-type, pyp1 and tco89 long-lived controls.

The predicted function of these genes included, a mitochondrial Fe-S cluster assembly co-chaperone (*jac1*), a predicted ER-associated ubiquitin-dependent protein with function in protein catabolism (*vms1*) and a predicted proteasome assembly

chaperone (*SPBC16E9.19*), (Lock et al., 2019). Since the function of neither of these genes, nor the remaining top long-lived validated genes has not been previously annotated as ageing relevant genes, and given that at least some genes (e.g. *jac1* and *vms1*) were previously shown to be implicated in processes associated with ageing in *S. cerevisiae* (Matecic et al., 2010; Nielson et al., 2017; Voisine et al., 2001) demonstrates that our Bar-seq screen included novel pro-ageing genes conserved across the yeast species. This conservation, at least for the (*jac1*) gene, extends to humans also (Dutkiewicz and Nowak, 2018), a conservation which may also be relevant for the remaining novel pro-ageing genes discovered in our CLS screen. Therefore, our Bar-seq screen provides a rich dataset which can be used to advance further CLS studies.

Chapter 8

Discussion

This chapter continues with a global discussion of the decoded fission yeast library strains, the importance of Bar-seq and RNA-seq, the chronological lifespan genes obtained from the re-growth data analysis of the competitively grown strains, and the validation of our top chronological lifespan hits. Future directions will also be described.

8.1 The fission yeast library decoded genes

8.1.1 Comparison of our decoded genes to previous versions shows that our characterisation was the most successful

The high-throughput sequence analysis of the deletion library strains with our custom-developed pipeline led to the characterisation of 3,206 strains, or 94% of the genes decoded with at least one barcode. Compared to the decoded genes of previously characterised deletion libraries with 2,560 strains, or 90% of gene deletions (Han et al., 2010) and 2,473 strains, or 82% of gene deletions (Sideri et al., 2015), our characterised deletion library, consisted of 1.3 times more decoded genes on average. Additionally, while we characterised 3,011 genes with both barcodes, the number of decoded genes characterised with both barcodes was 2,235 and 1,871 for version 1.0 (Han et al., 2010) and version 2.0 (Sideri et al., 2015), respectively. Therefore, unlike the previous versions of the deletion library characterisations we successfully decoded the highest number of genes and most of our strains have been decoded for both barcodes.

8.1.2 One-fifth of the undecoded genes represent genes with multiple copies and most are likely unsuccessful gene deletions

Despite the successful characterisation for most of the gene deletions in the library, 214 genes, or approximately 6% of the strains remained undecoded. In comparison to the number of undecoded genes from the previous library characterisations, to date these represent the lowest number of undecoded genes. It seems however, that the identification of genes which share the same barcodes and thus cannot be characterised is a common challenge also previously experienced during the characterisation of the earlier deletion library version (Han et al., 2010).

We found that within the 214 undecoded genes 36 were identified as duplicate genes and 178 as genes with no identifiable barcodes. Of the duplicate genes only half were found to have multiple copies. Thus, the other half likely represent cases of gene deletions with the same barcodes. Among the 178 genes with no identifiable barcodes only 35 were found to have more than one copy. The remainder were single-copy genes. The remaining genes with unidentifiable barcodes may include strains with low biomass, their absence from the pool, or loss of cell revival capacity. During the selection process for the strains used in our second re-characterisation attempt we found modest size colonies for most of these strains, thus the aforementioned explanations are unlikely the reason behind our identification of the undecoded genes. However, our ability to confirm three gene deletions out of the ten validated, together with a similarly low gene deletion validation rate previously described (Sideri et al., 2015) points at the 178 genes with no identifiable barcodes as likely examples of unsuccessful deletions.

8.2 Bar-seq & RNA-seq

8.2.1 Genome-wide analysis with Bar-seq distinguishes the method as a powerful genetic tool

Successful establishment of Bar-seq in budding yeast (Smith et al., 2009) led to the adaptation of the method to bacterial systems (Wetmore et al., 2015; Hobbs et al., 2010). Further improvements to the technique yielded the development of transposon pools available for a diverse range of organisms, including human cells (Mesarich et al., 2017; Brutinel and Gralnick, 2012; Carette et al., 2011; Gallagher et al., 2011). Because Bar-seq works by the deep sequencing of cellular molecular barcodes, fission yeast deletion library collection genome-wide analyses are limited to either study on solid media, or liquid growth in 96-well plates. Despite this limitation, numerous genome-wide screens were performed, including screens on spore formation (Ucisik-Akkaya et al., 2014), mating phenotype (Li et al., 2013), DNA damage response (Pan et al., 2012; Deshpande et al., 2009), catalase expression (García et al., 2016), sensitivity to ionising radiation (Li et al., 2014), drug sensitivity (Yang et al., 2018; Doi et al., 2015; Calvo et al., 2009; Kennedy et al., 2008), anti-fungal and anti-cancer drug target screens (Zhang et al., 2015; Fang et al., 2012; Takeda et al., 2011) and mutant fitness profiling (Lie et al., 2018; Malecki and Bähler, 2016; Sideri et al., 2015; Rallis et al., 2014; Roux et al., 2009).

Comparing our Bar-seq result with the solid screen result previously described where the aim was to screen for mitochondria defective mutants (Malecki and Bähler, 2016), we found that Bar-seq was more sensitive at identifying mitochondrial-protein encoding genes than the colony screen. Our identification of biologically relevant genes not previously identified with the colony screen supported our Bar-seq analysis approach, including its custom-developed pipeline. Therefore, this proof-of-principle experiment not only validated our upgraded version of Bar-seq and its custom-developed pipeline but also the fission yeast decoded library strains.

8.2.2 Transcriptomic analysis renders regulation of ncRNAs as important elements required for quiescence maintenance

The independence of the cellular barcoding from the experimental conditions enables Bar-seq application to genome-wide transcriptomic studies. The fundamentals of the method were also successfully adapted for RNA molecules (Kebschull and Zador, 2018). However, to combine the transcriptomics of long-term quiescent cells with the genomics of chronologically aged yeast cells, we focused on establishing an effective RNA-seq method to overcome the bottleneck of RNA instability to perform RNA-seq on long-term quiescent cells.

Our analysis on wild-type long-term quiescent cells distinguished several types of RNA molecules, and specifically, several ncRNAs with a putative role for long-term quiescence maintenance by fine-tuning the down-regulation of metabolic genes. This interesting finding highlights ncRNAs' importance in cellular metabolic control. In light of gaining further functional insight the ncRNAs were compared to the ncRNAs previously described (Atkinson et al., 2018). However, no further information was derived from this comparison.

8.2.3 Bar-seq whole-genome analyses & RNA-seq transcriptomic analyses are highly complementary

To gain a comprehensive understanding of the genetic elements underlying cellular ageing in yeast and the factors that drive their expression, genomics and transcriptomics integrative analyses is required. However, our work has helped bridge these two worlds together through our established and validated Bar-seq and RNA-seq methods, both of which can be carried out in a relatively fast and cost-effective manner. The two powerful genetic tools complement each other well and their application to future parallel functional profiling of non-essential gene deletions can provide new insights and exciting avenues to better investigate the molecular factors that contribute to the complexity of ageing.

8.3 Chronological lifespan of competitively grown decoded strains

8.3.1 Comparison of long-lived mutants to published data

Chronological lifespan (CLS) studies on the fission yeast gene deletion mutants are lacking, and to date only one cellular Bar-seq quiescence study (Sideri et al., 2015) and one study that uses the standard glucose-starvation model (Lie et al., 2018) exists. In addition to CLS studies, fission deletion collection was also used for chemical genomics screens (Doi et al., 2015; Rallis et al., 2014, 2013). A summary table describing the screens conducted with the Bioneer library and the number of the CLS genes overlapping with the genes identified from the previously described screens is presented in (**Table 8.1**).

Table 8.1: Long-lived CLS gene overlap with the previously described Bioneer library screens.

Deletion mutants tested for	**Gene no.**		**Ref.**
	Total	**Overlap**	
Torin1 sensitivity	139	8	Lie et al., 2018
Rapamycin sensitivity	59	2	Doi et al., 2015
Long-term quiescence fitness	103	9	Sideri et al., 2015
Rapamycin & caffeine resistance	726	21	Rallis et al., 2014,2013

*Note that the total gene number overlap comes from comparing our 168 long-lived CLS genes with the CLS genes identified from the referenced screens. For details about the overlapped genes (see **Table D.1**).*

While CLS quiescence studies using the budding yeast library are similarly scarce (Li et al., 2015), there are several studies performed using the CLS glucose-starvation model (Smith et al., 2016; Garay et al., 2014; Fabrizio et al., 2010; Matecic et al., 2010; Powers et al., 2006).

Comparison to budding yeast data

Despite differences in methods, comparison of our long-lived gene deletion mutants to the mutants obtained from the budding yeast studies showed some degree of similarity. The methods previously used to determine cell survival varied from plating the stationary-phase cells on solid media and using cell colony as a measure of fitness (Smith et al., 2016), pooling stationary-phase growths to determine fitness

using either DNA microarrays (Fabrizio et al., 2010; Matecic et al., 2010) or RNA microarrays (Kim et al., 2014) to individual growth on plates using either optical density (Powers et al., 2006) or CFUs (Garay et al., 2014). Although two studies showed no gene orthologs (Fabrizio et al., 2010; Powers et al., 2006), we identified one ortholog each (*sgp1, rps402 and swc3*) from the following studies (Smith et al., 2016; Garay et al., 2014; Matecic et al., 2010). We also found three ortholog gene deletions (*SPBC9B6.03, SPACUNK4.15 and ogm1*) common between our list of long-lived mutants and the long-lived quiescent mutants identified using the RNA microarrays method (Kim et al., 2014).

Comparison to fission yeast data

The comparison of our long-lived gene deletion mutants to the fission yeast mutants previously described (Lie et al., 2018) revealed eight common gene deletions (*vps71, sen1, pex1, mug161, pub3, gim3, pac10 and nfs1*). Notably, *pub3* was among the list of our CLS validated mutants. Our validation confirmed that the CLS of the *pub3* gene deletion mutant cultured individually recapitulated the lifespan observed from the competitive mutant pool growth, and that *pub3* CLS was greater than the wild-type and both the *pyp1* and the *tco89* long-lived controls.

We next compared our list of long-lived mutants to the mutants identified from the drug screens previously described (Doi et al., 2015; Rallis et al., 2014, 2013) and found one (*SPBP4H10.16c*), two (*ctu1 and ypa1*) and four common gene deletions (*SPCPB16A4.06c, mfs3, SPAC1002.17c and SPACUNK4.15*), respectively. The comparison of our long-lived mutants to the fission yeast mutants identified from the only high-throughput Bar-seq CLS screen revealed nine common genes (*pyp1, swc3, alg12, tea4, ppk33, ctu1, mug161, clg1 and SPBC1921.04c*). We noticed that the *pyp1* gene deletion mutant was a long-lived controls used in our validation assay and relative to the wild-type, the mutant displayed the expected long lifespan phenotype. As well as the two common genes (*swc3 and ctu1*) between the screens described above, we found that the genes were budding yeast orthologs with the latter gene also described as a rapamycin-sensitive gene deletion (Matecic et al., 2010; Doi et al., 2015). Interestingly, the deletion and mutation of the *clg1* gene

was also previously associated with mutant lifespan extension (Chen et al., 2013). While the *SPBC1921.04c* gene common between the quiescence-model and our glucose-starvation model evaded our validation assay, its lifespan was confirmed independently by the quiescence validation assay previously described (Sideri et al., 2015). The fact that the lifespan of the *pyp1* and the *SPBC1921.04c* genes was validated independently, and the genes were identified as long-lived in both our glucose-starvation model and the quiescence-model, suggests that the mutant lifespan is robust and independent of the ageing model used.

Our list of long-lived gene deletion mutants do not show the expected slow growth-lifespan relationship

We were curious about whether our long-lived mutants recapitulated the slow-growth phenotype ubiquitous for long-lived mutants as previously described (Rallis et al., 2014). Comparison of our long-lived gene deletions to the slow-growth long-lived annotated genes from PomBase (Lock et al., 2019) revealed only 25 common genes ($p<0.02$) of no statistical significance. However, this was not surprising as 42.3% of the long-lived gene deletions identified were genes with unknown functions. Therefore, these genes may still have the expected slow-growth lifespan relationship though not yet characterised as such.

8.3.2 Comparison of short-lived mutants to published data

The comparison of our short-lived mutants to the literature data distinguished several gene deletion mutants common between the studies. Reassuringly, more mutant gene deletions with similar lifespans were found for the comparison between our list of short-lived mutants and the fission yeast data than budding yeast data. For example, from the fission yeast data we found 38 common genes (Rallis et al., 2013) and only 10 genes from the budding yeast data; six genes (*sgf29, ada2, pex14, gcn5, did2 and gid8*) from (Smith et al., 2016), three genes (*SPAC227.17c, upf2 and kgd1*) from (Matecic et al., 2010) and only one gene (*ubp14*) from (Garay et al., 2014).

Some of our short-lived mutants were found to have the opposite lifespan when compared to published fission yeast data

To our surprise, several gene deletions previously described as long-lived corresponded to our list of 441 short-lived gene deletion mutants identified from the Bar-seq analysis of the re-growth data. These included 16 lifespan-opposing genes, half obtained by screening the deletion mutants for rapamycin sensitivity/resistance (Doi et al., 2015) and included the following genes (*pro1, csk1, elp1, php5, etr1, asc1, kgd1 and SPCC16C4.10*). Unlike the first four genes which had both a modest FC and p-values though not high enough to be included in our lifespan validation assay, the FC and p-values of the latter four genes (except for the *etr1* gene which had a high FC but low p-value) were among the lowest in the list of our short-lived gene deletion mutants, thus also evaded our validation assay.

The other half of the lifespan-opposing gene deletions emerged from the deletion library Bar-seq screen for long-lived mutants following revival from long-term quiescence (Sideri et al., 2015). From the re-growth screen, we found only six genes in common to our list of short-lived gene deletion mutants (*hrp3, rsv2, zhf1, rei1, SPBC56F2.05c and SPAC9.02c*) and two gene deletion mutants (*abp2 and SPAC8E11.05c*) from the standard screen. Because cell re-growth before Bar-seq analysis was an alternative CLS screen, Sideri *et al.,* 2015 did not include any of these genes in the validation screen. This was despite the *SPBC56F2.05c* gene, a predicted transcription factor involved in DNA transcription regulation showing the highest lifespan score among all the re-growth genes (Sideri et al., 2015). Interestingly, this gene in our screen had a considerably low FC and p-value, thus also evaded our validation assay. However, we found that only the *SPAC8E11.05c* gene that encodes a conserved fungal protein associated with clathrin-coated vesicles (predicted) has had its long lifespan validated and confirmed (Sideri et al., 2015). In our screen, however, neither *abp2* nor *SPAC8E11.05c* had high FC and p-values, hence these too were excluded from our validation assay.

Other than the experimental difference and the tools used to analyse the data, the CLS difference between the mutants identified in our Bar-seq screen and the colony drug screen resulted from the fact that in comparison to the colony screen, Bar-seq is a more robust and effective method. Because both studies applied Bar-seq to identify the competitively grown mutants, the same argument cannot be used to explain the opposite mutant lifespan phenotype observed between our screen and Sideri's screen. The CLS variation likely arose from the analysis method because while the previously published study used the time course approach to compute lifespan scores (Sideri et al., 2015), we instead used DESeq2 and rather than using the time course approach, due to the nature of our data, we used the relative difference analysis method. Additionally, our analysis was performed only on the re-growth data, which was not the case for the previously discussed study (Sideri et al., 2015).

8.4 Validation of top mutants

8.4.1 Validation of short-lived mutants uncovered eight novel genes important for the maintenance of early life

Apart from high FC and p-values, the short-lived gene deletion mutants selected for CLS validation were also gene deletions not previously studied, thus genes with unknown functions. We found that out of the eight validated mutants all had short lifespans relative to the wild-type, thus the validated mutant lifespan recapitulated the lifespan observed from the Bar-seq analysis. Although the *dbp7* and the *SPAC1952.09c* genes have previously been associated with a decreased growth phenotype when cultured on galactose and glycerol media (Malecki et al., 2016), our eight short-lived validated genes (*SPAC26H5.07c, tea5, oct1, dbp7, wtf13, alg14 and SPAC1952.09c*) are neither well-explored nor previously reported as short-lived mutants. Therefore, we propose these gene deletions which we identified from our Bar-seq screen as novel candidates with roles in the maintenance of early life.

8.4.2 Most of the validated long-lived mutants that saturated the re-growth cultures are likely cooperative cells

To identify as many novel longevity genes, and because we had to exclude the later time points (e.g. day 10 and day 12 samples) from the differential expression analysis, we also included the top mutants from each of the time points in our validation. The considerably lower number of mutants than reads observed for these samples led to their exclusion from the analysis. However, since the wild-type is not molecularly tagged, we do not know exactly where the wild-type lifespan is relative to the mutants in the pools, so theoretically any comparison to the wild-type alone would not be accurate. Therefore, in our validation, we included known short-lived and long-lived controls and because their CLS resided on either side of the wild-type confirmed that relative to the mutants in the pool, the wild-type CLS is representative of the average pool lifespan, thus allowing for a direct mutant to wild-type comparison.

Therefore, our finding that only four out of the 16 validated mutants obtained from day 10 and day 12 samples showed the expected long lifespan phenotype and that despite their culture dominance, the remaining 14 mutants are short-lived, suggests that their short-lifespans might be due to different reasons other than interfering with ageing *per say*. However, one plausible explanation of why these gene deletions did not show extended lifespans as expected from the competitive pool growth might be that these mutant cells are cooperative. This means that these mutants show a growth advantage in the stationary phase (GASP) only when cultured heterogeneously, as only in the presence of other cell types these mutants can act as 'scavengers' feeding on the nutrients released by the 'fast-dying' cells. Noteworthy, the GASP phenotype was also previously described in bacteria (Bacun-Druzina et al., 2007).

8.4.3 The validation of the long-lived mutants uncovered ten novel pro-ageing genes

The comparison between our top validated long-lived mutants with similar lifespans between the individually *vs* the competitively grown mutants, distinguished several mutants with lifespans considerably longer than both the wild-type and two of our long-lived controls. The genes identified were (*vms1, jac1, SPBC16E9.19, git11, aim21, pub3, cyp4, SPCC1494.08c, gmf1 and dad1*). Although all, but three genes (git11, aim21 and dad1) have identifiable human orthologs, none of the genes were previously associated with any lifespan phenotype, let alone annotated as gene deletions with pro-ageing effects. However, phenotypic data is available for two of these genes (*vms1 and cyp4*) which were previously phenotyped as slow growers when cultured in minimal media and glycerol media (Malecki et al., 2016; Sideri et al., 2015) but their phenotype in glucose media is currently unknown.

Out of the ten novel pro-ageing genes, three showed incredibly long lifespans with related function in protein catabolism, mitochondria degradation and biogenesis

Three out of these ten gene deletion mutants emerged as interesting candidates with CLS greater than the lifespan of any other validated mutant. Their function identifies them as a predicted proteasome assembly chaperone (*SPBC16E9.19*), a predicted ATPase complex subunit involved in ER-associated ubiquitin-dependent protein with a role in catabolism (*vms1*) and a mitochondrial (2Fe-2S) cluster assembly co-chaperone (*jac1*), (Lock et al., 2019). The dual role of the *vms1* gene in both mitochondria function and protein degradation (Heo et al., 2010; Wallace, 2005) links *jac1* and *SPBC16E9.19* gene function. Functional impairment of the *HSCB* gene, the human ortholog of *jac1* leads to fatal diseases in humans (Dutkiewicz and Nowak, 2018; Voisine et al., 2001). Unlike in the budding yeast, the deletion of the fission yeast *jac1* gene, similar to that of humans was previously reported to be lethal (Delewski et al., 2016). However, identification of the *jac1* deletion in our screen shows otherwise. Contrary to the report where the *jac1* gene deletion mutant grown on EMM media was found to be lethal (Delewski et al., 2016), the *jac1* gene

deletion in our screen was in a haploid background with the mutant cultured in rich YES media. Therefore, this finding suggests that the *jac1* gene deletion lethality is dependent on the environmental conditions, and this phenomenon might be true for other mutants also.

However, the fact that the *jac1* and the *vms1* genes are involved in maintaining the critical roles of mitochondria biogenesis, function and degradation, the dysfunction of which is associated with many age-related diseases in humans, such as cancer and neurodegenerative diseases mark these gene candidates highly intriguing (Wallace, 2005). The predicted function with no previous link to longevity and preserved functional conservation led us to propose these 10 gene deletions as novel longevity candidates discovered from our Bar-seq screen. Providing their relevance to humans, studying them may unlock some of the conundrums underpinning human cellular ageing.

8.5 Future Directions

A future study could include Sanger sequencing of the 178 undecoded and the 36 duplicate gene deletion mutants. This study will confirm both the deletion of these mutants and their barcode sequences.

The established and validated Bar-seq method, including its robust analysis pipeline allows for future high-throughput genome-wide studies in a fast and affordable manner. Future Bar-seq studies include calorific restriction, genetic drug targeting and global genetic-phenotypic relationships such as synthetic-genetic-arrays (SGAs) as well as studies requiring the testing of several experimental conditions simultaneously, a task that would otherwise be laborious and impractical.

One study which would have helped our re-growth analysis would have been the pooling of cell colonies formed at each time point before Bar-seq analysis and comparing the data with the data obtained from the liquid cultures. Additionally, despite the validation of our 48 mutants revealing new gene candidates with roles in ageing, a future study would include the validation of more genes. Also, the mutants which upon validation were found to show the opposite lifespan, and thus were suggested to be cooperative could prompt a future study validating their cooperativity.

Finally, an interesting future direction would be the follow-up of our proposed novel longevity gene candidates. One such experiment could test for the effect of certain lifespan extension drugs, such as rapamycin on the lifespan of these mutants. These attractive experiments could help delineate the mechanistic details of these genes/-drugs. Since these genes have identifiable human orthologs and their gene function impairment leads to human diseases, a similarly daring, yet attractive experiment could be the follow-up of these genes in mammalian systems. This study would be instrumental not only for providing insights relevant for mechanistic conservation but may also reveal new drug-targets important to human ageing.

Chapter 9

Concluding Remarks

Up to 67% of the fission yeast genes have identifiable human orthologs. This remarkable conservation is also shown across the fission yeast deletion library, where 47% of the 3,420 gene deletions have human orthologs. Since gene deletions are an example of genetic interventions shown to improve the lifespan of several organisms, the deletion library is a valuable resource for ageing studies, thus allowing for translation potential from yeast to humans.

Our work established and validated two complementary methods, Bar-seq and RNA-seq. To the best of our knowledge, the integration of both of these powerful genetic tools to gain genomic and transcriptomic data from aged cells has not been attempted before, though we now provide the necessary resources to enable this study. Therefore, the coupling of both of these methods can help characterise in more depth the molecular factors underlying yeast cellular ageing.

Bar-seq analysis of the data obtained from screening the fission yeast decoded genes grown competitively in a pool generated a rich data-set worthy of further exploration. Validation of top hits uncovered novel genes with roles in the maintenance of early life and genes with pro-ageing effects. The fact that these genes have human orthologs and their impairment influence human diseases marks them as perfect candidates for further study.

Appendix A

Deletion library barcode strain decoding analysis pipeline

A.1 Analysis

A custom pipeline to analyse the characterisation reads was custom developed as a result of a successful collaboration with two talented PhD students in the lab. BarSeqTools, an R package used to characterise the deletion library has been made available and can be accessed from this public Git repository (`https://github.com/Catalina37/Barcount_BarSeqTools_Pipelines`). Step by step details of the analysis are provided below.

Step 1: Reads filtering

R1 and R2 reads were separately combined into one large R1 and one large R2 fastq file for both uptag and dntag samples, enabling the streamlining of the uptag and the dntag files which were analysed separately. The barcode sequences were extracted using the pre-defined U1/U2 and D2/D1 uptag and dntag flank sequences defined as 5'-CAAGCTAAGATATC-3'(U1), 5'-TTTAAATGCGAAGTAA-3'(U2), 5'-TTTAAAATCCCCCCTA-3'(D2) and 5'-AGTGTCGAAAAGTATC-3'(D1). Three base pair mismatches were allowed within each flank sequence, and these were one deletion, one substitution and one insertion, thus enabling for barcode sequence extraction flexibility, irrespective of the barcode sequence length, which as a result of sequencing error rates within the barcode sequences were found to

vary between 13-23 bp in length, a phenomenon also previously described (Lee et al., 2018; Smith et al., 2009). R2 reads were then filtered based on the previously filtered R1 reads to match the read header information, thus selecting only for the R2 reads that correspond to the R1 barcodes.

Step 2: R2 customisation

The prerequisite to decoding the version 5.0 genes is the association between each barcode and a small part of the genomic region which can be used to map to the reference genome, thus linking the genomic coordinates to the corresponding barcode sequences. We meet this prerequisite by removing the index sequence from the R2 header and replaceing it with the barcode sequence extracted in step 1. Therefore, we retained the R2 header information required for alignment to the genome while tethering the gDNA to its corresponding barcode sequence.

Step 3: gDNA extraction & processing

The gDNA sequences of 43 bp in length were trimmed from the R2 reads using the fastx_trimmer (`http://hannonlab.cshl.edu/fastx_toolkit/`), and then filtered against the U2/D2 primer sequences, as the reads were found to contain part of the primer sequence. Given that the U2 and D2 sequences were directly adjacent to the barcodes and immediately downstream of the gDNA region, these were likely added during DNA shearing.

Step 4: Genome mapping

The gDNA sequences obtained from step 3 were mapped to the *S. pombe* genome using Bowtie2 (`http://bowtie-bio.sourceforge.net/bowtie2/index.shtml`). Only the uniquely mapped reads were considered for further analysis.

Step 5: Gene annotation

The uniquely aligned reads from step 4 were annotated to the *S. pombe* genes using Bedtools (`https://bedtools.readthedocs.io/en/latest/`). The protein-coding genes genomic coordinates were matched to the version 5.0 genes, thus extracting the genes coordinates required for further analysis.

Step 6: Barcode consensus generation

BarSeqTools was used to extract the barcode sequences previously added into the R2 header. These were then quantified each time they were annotated to a gene, thus forming unique gene-barcode pairs. Pairs which occurred ≤ 2 were omitted from further analysis to reduce the data processing time. The remaining barcodes were then used to generate barcode consensuses, by forming different variants of a barcode defined by the number of base-pair mismatches allowed from the original barcode sequence, thus grouping barcodes with similar sequences. Therefore, the base-pair mismatches within the barcode consensuses can be used to select for a barcode consensus sequence with a specific number of mutations from the original sequence.

Step 7: Barcode consensus matching

The barcode consensuses defined by three allowed base-pair mismatches were matched to the sequences of the extracted barcodes. Therefore, only the barcode sequences with three or fewer mismatches to the initially extracted barcode sequences were processed further.

Step 8: Gene-barcode pair frequency

BarSeqTools was used to calculate the frequency or the occurrence for each unique gene-barcode pair. This measure quantified the number of times each gene was associated with a unique barcode sequence, thus a unique gene-barcode pair.

Step 9: Gene-barcode pair proportion

BarSeqTools was also used to calculate the proportion, defined by a percentage for each gene and barcode separately within a unique gene-barcode pair. This was calculated as the relative gene/barcode proportions within a unique gene-barcode pair. These gene-centric and barcode-centric proportions can be used to define the stringency of the gene-barcode pair strength, thus selecting for unique-barcode pairs.

Step 10: High-confident gene-barcode pairs selection
High-confidence unique gene-barcode pairs were selected by combining the gene and barcode proportions as 80% each with the minimum unique gene-barcode pair frequency of 10, generated from step 9 and 8, respectively.

Step 11: Gene browser
As part of our analysis, we developed a gene browser, also incorporated in the BarSeqTools package. The browser comprises of several tracks: a chromosomal track, two gene tracks, one for the protein-coding genes and one for the deletion library genes, and a barcode track where the uptag and the dntag associated reads, defined by genomic coordinates, are on the upper and the lower track, respectively. The barcode read coordinates aligning to the version 5.0 gene genomic coordinates and the uptag and dntag barcodes residing downstream and upstream of the deleted gene, enable easy gene characterisation with the visual gene browser. Besides, the different colours for each barcode sequence, the gene browser highlights the gene of interest in black while the neighbouring genes are coloured in grey. An interaction network is also used to display the proportion between the gene and its associated barcode/s. Furthermore, the gene browser is interactive with adjustable visualisation parameters for each track and the query gene. Several pre-requisite files are required to use the gene browser, including the gene annotation files, the calculated aggregation files, the automatically decoded genes and the version 5.0 reference genes. However, these pre-requisites have been made available and can be accessed from this public Git repository (`https://github.com/Catalina37/Barcount_BarSeqTools_Pipelines/tree/master/BioneerV5.0_Characterisation/Browser`).

The gene browser was extremely useful not only for visualising and cross-validating the automatically selected gene-barcode pairs but also for the manual gene characterisation. The subsequent figures represent examples of the different gene-barcode pairs observed in the gene browser, including unique gene-barcode pairs (**Figure A.1**), duplicate genes (**Figure A.2**), dubious genes (**Figure A.3**) and genes with no identifiable barcodes (**Figure A.4**).

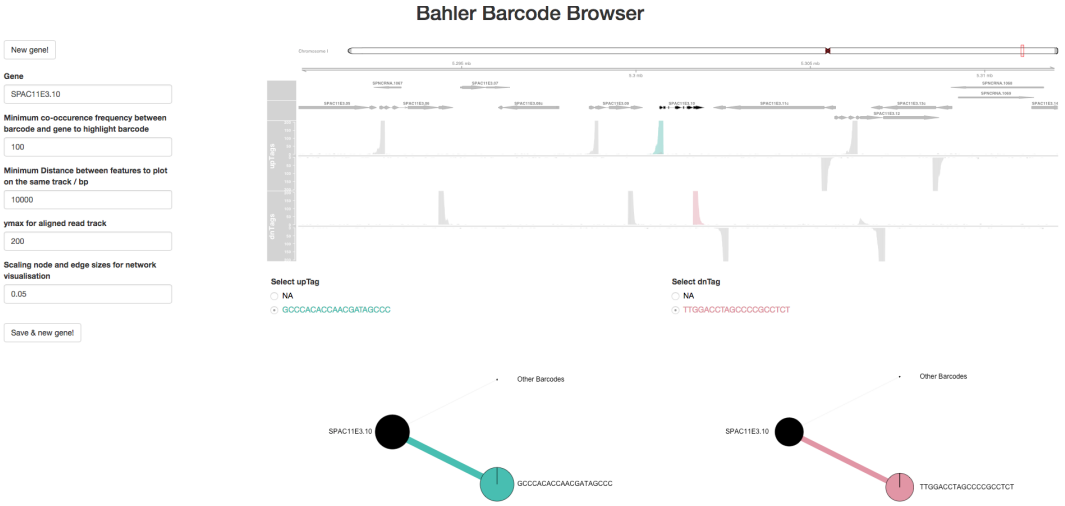

Figure A.1: Gene-browser example of a fully decoded gene.
Each barcode tag maps uniquely to each gene.

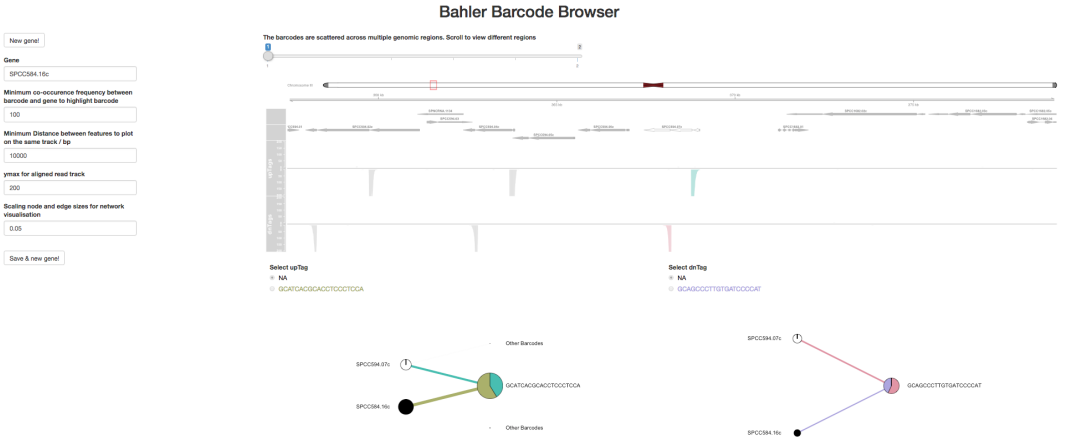

Figure A.2: Gene browser example of duplicate genes.
The barcode tag shared between two genes. The network indicates the proportion with which the barcode tag maps to each unique gene.

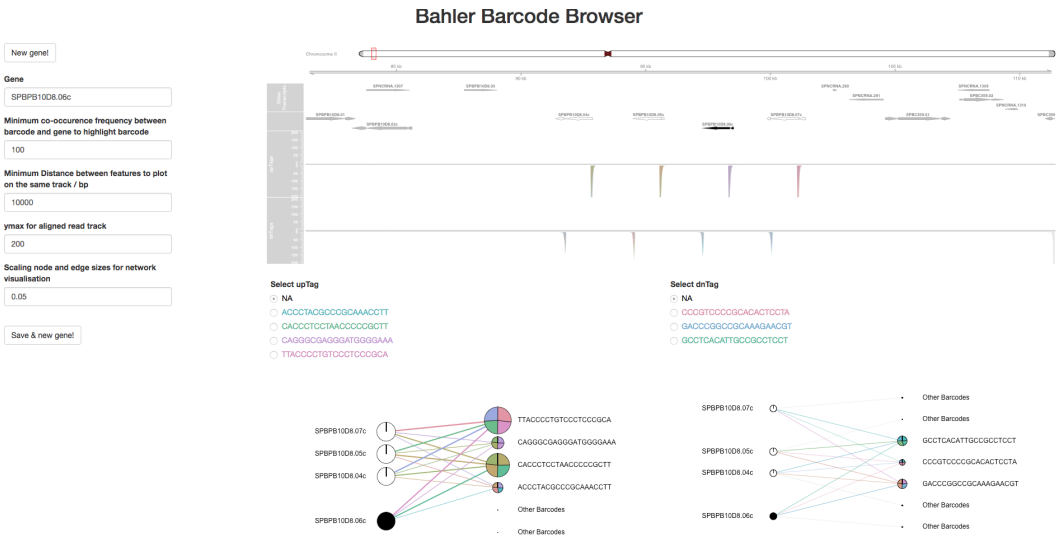

Figure A.3: Gene browser example of dubious genes.
Each barcode tag is mapping in nearly equal proportions over the four different genes rendering them as dubious, thus indeterminate genes.

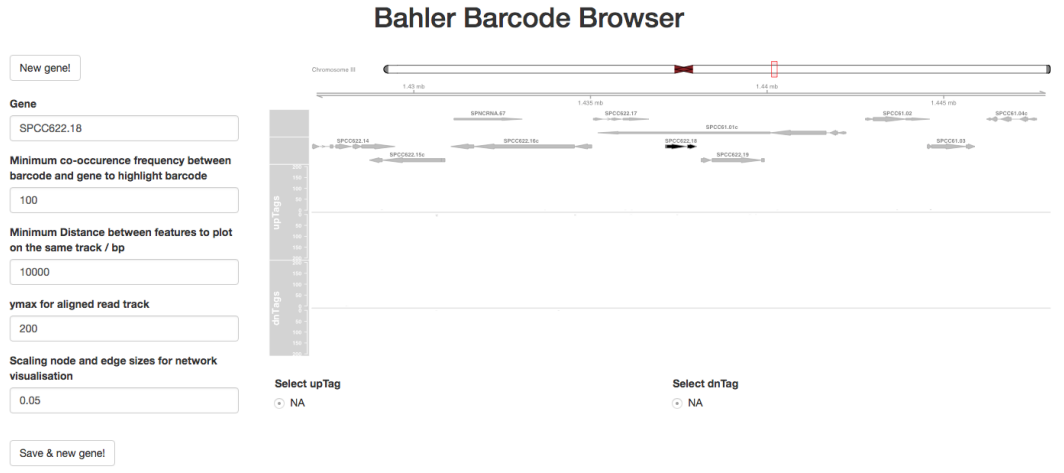

Figure A.4: Gene browser example of genes with no identifiable barcodes.
The absence of barcode tags around the query gene renders it as undecoded.

A.2 Script

The script used for the characterisation sequence data analysis, developed in collaboration with StJohn Townsend and Stephan Kamrad, two talented PhD students in the lab is provided below.

```
---
Title: "Deletion Library Characterization: Script "
---
# Load the custom built R package called BarSeqTools.
# All of the custom-built scripts are embeded within BarSeqTools.
library(BarSeqTools)

####### Sample Preparation #######
#1. Set working directory, one for uptag and one for dntag files
setwd("~/Desktop/Cat_BionnerV5_Recharc_May18-77598553/FASTQ_
    Generation_2018-05-22_07_28_46Z-96698246/uptag/")
setwd("~/Desktop/Cat_BionnerV5_Recharc_May18-77598553/FASTQ_
    Generation_2018-05-22_07_28_46Z-96698246/dntag/")

#2. Combine all uptag and dntag R1 and R2 fastq files
# uptag
system("cat */*R1* > upTagsR1.fastq")
system("cat */*R2* > upTagsR2.fastq")
# dntag
system("cat */*R1* > dnTagsR1.fastq")
system("cat */*R2* > dnTagsR2.fastq")

#3. Load MAGIC: cutsom-built script linking barcodes & genomic
    sequences together
MAGIC("upTagsR1.fastq", "upTagsR2.fastq", "up")
MAGIC("dnTagsR1.fastq", "dnTagsR2.fastq", "dn")

####### Sample analysis is performed with Barcount #######
#1. Set working directory for Barcount analysis
setwd("~/Desktop/Cat_BionnerV5_Recharc_May18-77598553/FASTQ_
    Generation_2018-05-22_07_28_46Z-96698246/up_dn_UniqueMatches/")
```

```
28
29#2. Extract barcodes from bed files & save as csv
30ExtractBarcodefromBEDTOOL("upBedtools.Unique.bed", "upTags.Unq.csv
    ")
31ExtractBarcodefromBEDTOOL("dnBedtools.Unique.bed", "dnTags.Unq.csv
    ")
32
33#3. Name & load the cvs files
34upTags <- read.csv("upTags.Unq.csv")
35dnTags <- read.csv("dnTags.Unq.csv")
36
37#4. Load package & count barcode presence and sequence length
38library(plyr)
39
40upTagsCount <- ddply(upTags, .(tag), summarise, n=length(tag))
41dnTagsCount <- ddply(dnTags, .(tag), summarise, n=length(tag))
42# Files are too large for R to process, discard reads.
43
44#5. Discard barcodes appearing <=2 to a unique gene
45upTagsUncommon <- upTagsCount$tag[upTagsCount$n<=2]
46dnTagsUncommon <- dnTagsCount$tag[dnTagsCount$n<=2]
47
48# 6. Filter for the barcode reads mapped to a unique gene >2
49upTags <- upTags[-which(upTags$tag %in% upTagsUncommon),]
50dnTags <- dnTags[-which(dnTags$tag %in% dnTagsUncommon),]
51
52#7. Create barcode consensus by allowing 2 bp mismatches
53upConsensus <- createConsensus(upTags$tag, 2, "upTags.pdf")
54dnConsensus <- createConsensus(dnTags$tag, 2, "dnTags.pdf")
55
56#8. Plot the distribution of the created barcode consensuses
57hist(upConsensus$n, breaks=1500, xlim=c(0,10000), ylim=c(0, 100),
    main="Uptags", xlab="Frequency of Consensus Barcode")
58hist(dnConsensus$n, breaks=1500, xlim=c(0,10000), ylim=c(0, 100),
    main="Dntags", xlab="Frequency of Consensus Barcode")
59
```

```
#9. Match barcodes to genes based on the barcode consenuses and
    actual barcodes
# Max difference allowed is 3 bp mismatches
upTags <- matchBarcodes(upTags, tag, upConsensus$barcodeConsensus)
dnTags <- matchBarcodes(dnTags, tag, dnConsensus$barcodeConsensus)

#10. Write out these tables as csv
write.csv(upTags, "upTags.csv", row.names = F)
write.csv(dnTags, "dnTags.csv", row.names = F)

#11. Calculate the number of mismatches between consensus & actual
     barcodes
table(upTags$nMismatches)
table(dnTags$nMismatches)

#12. Calculate the frequency for each gene-barcode consensus
upTagsFreq <- calculateFrequency(upTags, tag_BestMatch, gene)
dnTagsFreq <- calculateFrequency(dnTags, tag_BestMatch, gene)

#13. Calculate aggregation for each gene-barcode consensus as
    follows:
# - each unique gene against every best barcode consensus
# - each unique best barcode consensus against every gene
upTagsAgg <- aggregateProportions(upTagsFreq, tag_BestMatch, gene)
dnTagsAgg <- aggregateProportions(dnTagsFreq, tag_BestMatch, gene)

#14. Write out these tables as csv
write.csv(upTagsAgg, "upTagsAgg.csv", row.names = F)
write.csv(dnTagsAgg, "dnTagsAgg.csv", row.names = F)

#15. Create safe gene-barcodes with the highest frequency
    occurance based on:
# - min 10 occurences for each gene linked to that particular
    barcode
# - min 80% of the time for which that gene is linked to that
    barcode
```

```
minFreq <- 10
minProportion <- 0.8

#16. Write a function to visualise these two cut offs
  plotSafes <- function(tagsAgg, y, minFreq, minProportion, ylab,
    main) {
  plot(tagsAgg$freq, tagsAgg[[y]], cex=0.1, col=ifelse(tagsAgg$
    freq>minFreq & tagsAgg[[y]]>minProportion, "Blue", "Black"),
    xlab="Barcode-Gene Frequency", ylab=ylab, main=main)
  abline(h=minProportion, col="Red")
  abline(v=minFreq, col="Red")
}

### Plot to visualise the uptag safe list
par(mfrow=c(1,2))
plotSafes(upTagsAgg, "barcodeProportion", minFreq, minProportion,
    "Proportion of Barcode in Barcode-Gene Pair", main="Up Tags")
plotSafes(upTagsAgg, "geneProportion", minFreq, minProportion, "
    Proportion of Gene in Barcode-Gene Pair", main="Up Tags")

### Plot to visualise the dntag safe list
par(mfrow=c(1,2))
plotSafes(dnTagsAgg, "barcodeProportion", minFreq, minProportion,
    "Proportion of Barcode in Barcode-Gene Pair", main="Dn Tags")
plotSafes(dnTagsAgg, "geneProportion", minFreq, minProportion, "
    Proportion of Gene in Barcode-Gene Pair", main="Dn Tags")

#17. Subset these automatically detected gene-barcode pairs
upTagsSafe <- subset(upTagsAgg, upTagsAgg$freq>=minFreq &
    upTagsAgg$barcodeProportion>=minProportion & upTagsAgg$
    geneProportion>=minProportion)
dnTagsSafe <- subset(dnTagsAgg, dnTagsAgg$freq>=minFreq &
    dnTagsAgg$barcodeProportion>=minProportion & dnTagsAgg$
    geneProportion>=minProportion)

```

```
#18. Save the sutomatically detected genes by writing these tables
    out as csv
upTagsSafe$Method <- "auto"
dnTagsSafe$Method <- "auto"

write.csv(dnTagsSafe, "dnTagsAutoSafe.csv", row.names = F)
write.csv(upTagsSafe, "upTagsAutoSafe.csv", row.names = F)

####### Analysis Quality Check #######
#1. Identify the number of safe genes, unique genes and unique
    barcodes
# uptag
nrow(upTagsSafe)
length(unique(upTagsSafe$barcode))
length(unique(upTagsSafe$gene))

#dntag
nrow(dnTagsSafe)
length(unique(dnTagsSafe$barcode))
length(unique(dnTagsSafe$gene))

####### Gene Browser #######
#1. Load the prerequisites for running the Gene Browser
myTxDb <- makeTxDbFromBiomart(biomart = "fungi_mart", dataset = "
    spombe_eg_gene", host="fungi.ensembl.org")
options(ucscChromosomeNames=FALSE)

#2. Set directory
setwd("~/Desktop/Cat_BionnerV5_Recharc_May18-77598553/FASTQ_
    Generation_2018-05-22_07_28_46Z-96698246/up_dn_UniqueMatches/")

#3. Load the dependency files to visualise the genes of interest
BahlerBarcodeBrowser("upTags.csv", "dnTags.csv", "upTagsAgg.csv",
    "dnTagsAgg.csv", "upTagsAutoSafe.csv", "dnTagsAutoSafe.csv", "
    upTagsManualSafe.csv", "dnTagsManualSafe.csv", myTxDb)
```

Appendix B

Bar-seq analysis pipeline

Barcount, a stand-alone python script used to analyse Bar-seq data was developed in collaboration with Stephan Kamrad, a PhD student in the lab. The package was made available to download and can be accessed from this public Git repository (`https://github.com/Catalina37/Barcount_BarSeqTools_Pipelines/tree/master/BarSeq`).

B.1 Script

The script with the steps used to analyse the Bar-seq sequence data is provided below.

```
#!/bin/bash

#1. Unzip the fastq.compressed files
for f in */*.gz; do echo "$f"; done
for f in */*.gz; do gunzip "${f}"; done


#2. Full overlap R1 & R2 fastq assembly with PEAR
R1=($(ls -d */*R1_001.fastq))
R2=($(ls -d */*R2_001.fastq))

output=(${R1[@]/R1_001.fastq/PEAR})
length=${#R1[@]}
length=$(($length-1))

```

```
for ((i=0;i<=$length;i++)); do echo $i; echo ${R1[$i]}; echo ${R2[$i]}; echo ${output[$i]}; done
for ((num=0;num<=$length;num++)); do echo $num; pear -n 86 -m 86 -f ${R1[$num]} -r ${R2[$num]} -o ${output[$num]} > ${output[$num]}_log; done


#3. Remove every other file from the PEAR assembly analysis except the assembled file
for f in */*001.fastq; do echo "$f"; done
for f in */*001.fastq; do rm "$f"; done

for f in */*PEAR.unassembled*.fastq; do echo "$f"; done
for f in */*PEAR.unassembled*.fastq; do rm "$f"; done

for f in */*discarded.fastq; do echo "$f"; done
for f in */*discarded.fastq; do rm "$f"; done


#4. Run barcount on dntag
for f in */*dn*.assembled.fastq; do barcount --fastq "$f" --flanking_left AGTATC --flanking_right TTTAAA --max_distance_flanks 1 --max_distance_barcode 3 --barcode_table /home/ucbtomi/dntag_reference.csv --debug --save_extracted_barcodes --verbose --umiA_position " 4:8" --umiB_position " -8:-4" --out "${f}_Barcode_filter"; done


#5. Run barcount on uptag
for f in */*up*.assembled.fastq; do barcount --fastq "$f" --flanking_left GATATC --flanking_right TTTAAA --max_distance_flanks 1 --max_distance_barcode 3 --barcode_table /home/ucbtomi/uptag_reference.csv --debug --save_extracted_barcodes --verbose --umiA_position " 4:8" --umiB_position " -8:-4" --out "${f}_Barcode_filter"; done
```

Appendix C

CLS lifespan validation script

C.1 Script

The maximum likelihood script with the steps used to estimate the number of viable cells for the CLS validated mutants.

```
#Following installation of the DeadOrAlive package, load library
library(DeadOrAlive)
library(parallel)

#Initialise parallel computing environment
no_cores <- 2
cl <- makeCluster(no_cores)

#Get the directory to be processed
dir <- "~/Desktop/regrowth/ValidationData/Cat_Validation/20190527_
    catValidation/"

#Get all files
files <- list.files(dir, "\\.jpg")

#Extract image number from the file
imageNo <- unlist(regmatches(files, regexec("\\d+_catValidation\\.
    jpg", files)))
imageNo <- as.numeric(sub("_catValidation\\.jpg", "", imageNo))

#Get reference plates and remove from list of files
```

```
referencePlates <- files[imageNo%in%c(1,2,3,4)]
imageNo <- imageNo[files%in%referencePlates==F]
files <- files[files%in%referencePlates==F]

#Split the images based on scanner position
pos1 <- files[imageNo%%4==1]
pos2 <- files[imageNo%%4==2]
pos3 <- files[imageNo%%4==3]
pos4 <- files[imageNo%%4==0]

#Set up parallel environment
clusterExport(cl, c("dir", "referencePlates", "pos1", "pos2", "
    pos3", "pos4"))
clusterEvalQ(cl, library(DeadOrAlive))

#Run colonyThreshold with each scanner position on a different
    worker
parSapply(cl, 1:4, function(x) {
  colonyThreshold(dir=dir, fileFormat="jpg", files=get(paste0("pos
    ", x)), reference=referencePlates[x], inverse=T)
})

#Get the directory to be processed
dir <- "~/Desktop/regrowth/ValidationData/Cat_Validation/Cat/
    20190527_catValidation/"

#Get all files
files <- list.files(dir, "\\.jpg")

#Extract image number from the file
imageNo <- unlist(regmatches(files, regexec("\\d+_catValidation\\.
    jpg", files)))
imageNo <- as.numeric(sub("_catValidation\\.jpg", "", imageNo))

#Get reference plates and remove from list of files
referencePlates <- files[imageNo%in%c(35,36,37,38)]
```

```
referencePlates <- referencePlates[c(3,4,1,2)]
imageNo <- imageNo[files%in%referencePlates==F]
files <- files[files%in%referencePlates==F]

#Split the images based on scanner position
pos1 <- files[imageNo%%4==1]
pos2 <- files[imageNo%%4==2]
pos3 <- files[imageNo%%4==3]
pos4 <- files[imageNo%%4==0]

#Set up parallel environment
clusterExport(cl, c("dir", "referencePlates", "pos1", "pos2", "
    pos3", "pos4"))
clusterEvalQ(cl, library(DeadOrAlive))

#Run colonyThreshold with each scanner position on a different
    worker
parSapply(cl, 1:4, function(x) {
  colonyThreshold(dir=dir, fileFormat="jpg", files=get(paste0("pos
    ", x)), reference=referencePlates[x], inverse=T)
})

#Stop the cluster when everything is done
stopCluster(cl)

#Set working directory
setwd("~/Desktop/regrowth/ValidationData/")

#Extract colony vectors
myColonyVectors <- extractColonyVectors("Cat_Validation/20190527_
    catValidation/Image_Analysis/", "~/Desktop/regrowth/
    ValidationData/plateReferenceFile.csv", "~/Desktop/regrowth/
    ValidationData/sampleReferenceFile.csv")
myColonyVectors2 <- extractColonyVectors("Cat_Validation/Cat/
    20190527_catValidation/Image_Analysis/", "~/Desktop/regrowth/
    ValidationData/plateReferenceFile2.csv", "~/Desktop/regrowth/
```

```
    ValidationData/sampleReferenceFile.csv")
myColonyVectors <- c(myColonyVectors, myColonyVectors2)

#Perform MLE
CFUsMLE<- analyseColonyVectors(myColonyVectors, dilution=3, table.
    name = "CFUsMLE.csv", markdown.name = "CFUsMLE.html")

#Plot estimated number of live cells per mutant and timepoint
setwd("~/Desktop/regrowth/ValidationData/")
CFUsMLE <- read.csv("CFUsMLE.csv", stringsAsFactors = F)
table(CFUsMLE$TotalExclusions)
CFUsMLE <- CFUsMLE[-which(CFUsMLE$TotalExclusions>1),]

library(ggplot2)
g <- ggplot(CFUsMLE, aes(Time, ColonyFormingUnitsPerDroplet, group
    =Sample, color=Sample)) + geom_point() + geom_line() + scale_y_
    log10() + theme(legend.position = "none")
g
```

Appendix D

List of the long-lived mutants

Provided below is the 168 long-lived gene deletions identified from the CLS Bar-seq screen, including the mutants selected for CLS validation as well as the gene product description as shown in (**Table D.1**).

Table D.1: Bar-seq long-lived gene deletion mutants and the top hits CLS validation.

SystematicID	Gene	Validated	CLS	Product
SPAC1002.17c (#5)	urg2	no		uracil phosphoribosyltransferase (predicted)
SPAC1071.04c	spc2	no		signal peptidase subunit Spc2 (predicted)
SPAC10F6.05c	ubc6	yes	long	ubiquitin conjugating enzyme E2 Ubc6 (predicted)
SPAC1142.08	fhl1	no		forkhead transcription factor Fhl1
SPAC11E3.01c	swr1	no		SNF2 family ATP-dependent DNA helicase Swr1
SPAC1296.02	cox4	no		cytochrome c oxidase subunit IV (predicted)
SPAC12B10.07	acp1	no		F-actin capping protein alpha subunit
SPAC13G6.02c	rps101	no		40S ribosomal protein S3a
SPAC13G6.10c	asl1	no		cell wall protein Asl1 predicted O-glucosyl hydrolase
SPAC144.08	jac1	yes	long	mitochondrial (2Fe-2S) cluster assembly co-chaperone Jac1
SPAC14C4.06c	nab2	no		poly(A) binding protein Nab2 (predicted)
SPAC14C4.12c	laf1	no		Clr6 L associated factor 1 Laf1
SPAC14C4.16	dad3	no		DASH complex subunit Dad3
SPAC167.01	ire1	no		serine/threonine protein kinase
SPAC16A10.02	sub1	no		transcription coactivator PC4
SPAC16A10.04	rho4	no		Rho family GTPase Rho4
SPAC16A10.05c	dad1	yes	long	DASH complex subunit Dad1
SPAC16C9.02c (#5)	mta1	yes	short	S-methyl-5-thioadenosine phosphorylase Mta1
SPAC1783.08c (#5)	rpl1502	no		60S ribosomal protein L15b (predicted)

Table D.1 continued from previous page

SystematicID	Gene	Validated	CLS	Product
SPAC17A2.09c	csx1	no		RNA-binding protein Csx1
SPAC17C9.15c		no		Schizosaccharomyces specific protein
SPAC17G6.15c	fsf1	no		mitochondrial carrier, serine Fsf1 (predicted)
SPAC17H9.13c	pro2	no		glutamate 5-kinase Pro2
SPAC1834.05	alg9	no		mannosyltransferase complex subunit Alg9 (predicted)
SPAC1F3.09 (#1, #3)	mug161	no		CwfJ family protein, splicing factor (predicted)
SPAC227.05 (#1)	gim3	no		prefoldin subunit 4, Gim3 (predicted)
SPAC22A12.07c	ogm1	no		protein O-mannosyltransferase Ogm1
SPAC22E12.05c	rer1	no		Rer1 family protein (predicted)
SPAC22E12.11c	set3	no		histone lysine methyltransferase Set3
SPAC22E12.19	snt1	no		Set3 complex subunit Snt1
SPAC22F3.06c	lon1	no		mitochondrial matrix Lon protease
SPAC22F8.09 (#5)	rrp16	no		rRNA processing protein Rrp16 (predicted)
SPAC23C4.08	rho3	no		Rho family GTPase Rho3
SPAC23E2.01	fep1	yes	short	iron-sensing transcription factor Fep1
SPAC23H4.16c	not11	no		CCR4-Not complex subunit Not11
SPAC24B11.10c	cfh1	no		SEL1/TPR repeat protein Cfh1 (predicted)
SPAC26A3.09c	rga2	no		RhoGAP, GTPase activating protein Rga2
SPAC26F1.10c (#3)	pyp1	yes	long	tyrosine phosphatase Pyp1
SPAC27E2.02	yih1	no		IMPACT homolog, cytoplasmic translational regulator Yih1 (predicted)

Table D.1 continued from previous page

SystematicID	Gene	Validated	CLS	Product
SPAC27E2.03c		no		Obg-like ATPase, human OLA1 ortholog (predicted)
SPAC2C4.08		no		conserved fungal protein
SPAC2F3.02		no		ER protein translocation subcomplex subunit (predicted)
SPAC2F7.03c	pom1	no		DYRK family cell polarity protein kinase Pom1
SPAC2F7.04	med1	no		mediator complex subunit Med1
SPAC2G11.07c	ptc3	no		MAP kinase threonine phosphatase, protein phosphatase 2c homolog Ptc3
SPAC2H10.02c	nas2	no		26S proteasome regulatory particle assembly protein Nas2 (predicted)
SPAC30D11.07	nth1	no		DNA endonuclease III
SPAC328.06	ubp2	no		ubiquitin C-terminal hydrolase Ubp2
SPAC343.11c	msc1	no		Swr1 complex subunit Msc1
SPAC3A11.11c		yes	long	oxidoreductase, implicated in vitamin metabolism, or cellular detoxification (predicted)
SPAC3A11.13	gim1	no		prefoldin subunit 6, Gim1 (predicted)
SPAC3A12.10 (#5)	rpl2001	no		60S ribosomal protein L20a (predicted)
SPAC3C7.04		no		transcription factor (predicted)
SPAC3G9.03 (#5)	rpl2301	no		60S ribosomal protein L23
SPAC3G9.07c	hos2	no		histone deacetylase (class I) Hos2
SPAC3H8.07c (#1)	pac10	no		prefoldin subunit 3 Pac10 (predicted)
SPAC4F10.04 (#2)	ypa1	no		protein phosphatase type 2A regulator, PTPA family Ypa1
SPAC4F10.19c	hit1	no		zf-HIT family C/D snoRNP assembly protein Hit1 (predicted)
SPAC4G8.10	gos1	no		SNARE Gos1 (predicted)

Table D.1 continued from previous page

SystematicID	Gene	Validated	CLS	Product
SPAC4H3.02c (#3)	swc3	no		Swr1 complex subunit Swc3
SPAC56E4.06c	ggt2	no		gamma-glutamyltranspeptidase Ggt2
SPAC56E4.07		no		N-acetyltransferase (predicted)
SPAC57A7.12 (#5)	ssz1	no		heat shock protein Ssz1 (predicted)
SPAC5H10.09c	ecm31	no		3-methyl-2-oxobutanoatehydroxymethyltransferase Ecm31 (predicted)
SPAC631.01c	acp2	no		F-actin capping protein beta subunit Acp2
SPAC631.02	bdf2	no		BET family double bromodomain protein Bdf2
SPAC644.15 (#5)	rpp101	no		60S acidic ribosomal protein A1
SPAC6C3.08	nas6	no		proteasome assembly chaperone, gankyrin
SPAC6G9.09c (#5)	rpl2401	no		60S ribosomal protein L24 (predicted)
SPAC6G9.10c (#1)	sen1	no		ATP-dependent 5’ to 3’ DNA/RNA helicase Sen1
SPAC823.11	sgp1	no		sphingosine-1-phosphate phosphatase
SPAC823.15	ppa1	no		minor serine/threonine protein phosphatase Ppa1
SPAC890.06	nup155	no		nucleoporin, WD repeat Nup155
SPAC8C9.17c	spc34	no		DASH complex subunit Spc34
SPAC8E11.02c	rad24	no		14-3-3 protein Rad24
SPAC926.03	rlc1	no		myosin II regulatory light chain Rlc1
SPAC959.04c	omh6	no		alpha-1,2-mannosyltransferase Omh6 (predicted)
SPAC959.07 (#5)	rps403	no		40S ribosomal protein S4 (predicted)
SPAC9E9.08	rad26	no		ATRIP, ATR checkpoint kinase regulatory subunit Rad26

Table D.1 continued from previous page

SystematicID	Gene	Validated	CLS	Product
SPACUNK4.15 (#5)		yes	short	2',3'-cyclic-nucleotide 3'-phosphodiesterase (predicted)
SPAP8A3.05	ski7	no		Ski complex interacting GTPase Ski7
SPAPB17E12.08	eos1	no		N-glycosylation protein Eos1 (predicted)
SPAPJ696.02	lsb4	no		actin cortical patch component Lsb4 (predicted)
SPBC119.08	pmk1	no		MAP kinase Pmk1
SPBC11B10.10c	pht1	no		histone H2A variant H2A.Z, Pht1
SPBC1683.01		no		inorganic phosphate transmembrane transporter (predicted)
SPBC16A3.10	ale1	no		membrane bound O-acyltransferase, MBOAT Ale1 (predicted)
SPBC16C6.05	tma22	no		translation machinery associated protein ortholog Tma22 (predicted)
SPBC16E9.11c (#1)	pub3	yes	long	HECT-type ubiquitin-protein ligase E3 Pub3 (predicted)
SPBC16E9.19		yes	long	proteasome assembly chaperone 3 (predicted)
SPBC16G5.13	ptf2	no		Mst2 histone acetytransferase acytyltransferase complex subunit
SPBC1706.01 (#3)	tea4	no		tip elongation aberrant protein Tea4
SPBC1734.12c (#3)	alg12	no		dolichyl pyrophosphate Man7GlcNAc2 alpha-1,6-mannosyltransferase Alg12 (predicted)
SPBC17A3.06		no		phosphoprotein phosphatase (predicted)
SPBC17D11.04c	nto1	no		histone acetyltransferase complex PHD finger subunit Nto1 (predicted)
SPBC17G9.10 (#5)	rpl1102	yes	long	60S ribosomal protein L11 (predicted)
SPBC18H10.11c	ppr2	no		mitochondrial PPR repeat protein Ppr2
SPBC1921.04c (#3)		no		dubious
SPBC1D7.03 (#3)	clg1	no		cyclin-like protein involved in autophagy Clg1 (predicted)

Table D.1 continued from previous page

SystematicID	Gene	Validated	CLS	Product
SPBC20F10.05	nrl1	no		RNAi-mediated silencing protein, human NRDE2 ortholog Nrl1
SPBC20F10.10	psl1	no		cyclin pho85 family Psl1 (predicted)
SPBC215.02	bob1	no		prefoldin subunit 5 (predicted)
SPBC215.04	git11	yes	long	heterotrimeric G protein gamma subunit Git11
SPBC21B10.10 (#5)	rps402	no		40S ribosomal protein S4 (predicted)
SPBC21C3.20c	git1	no		C2 domain protein Git1
SPBC21D10.11c (#1)	nfs1	no		mitochondrial [2Fe-2S] cluster asssembly & tRNA modification cysteine desulfurase
SPBC23G7.04c	nif1	no		protein kinase inhibitor, SEL1 repeat protein Nif1
SPBC24C6.10c	dip1	no		WISH/DIP/SPIN90 ortholog, endocytosis protein Dip1
SPBC25B2.04c	mtg1	no		mitochondrial translation factor (GTPase) Mtg1 (predicted)
SPBC26H8.01	thi2	no		thiazole biosynthetic enzyme
SPBC27B12.10c	tom7	no		mitochondrial TOM complex subunit Tom7 (predicted)
SPBC28F2.02 (#5)	mep33	no		translation machinery associated protein Mep33
SPBC29A10.16c		no		cytochrome b5 (predicted)
SPBC29A3.05 (#1)	vps71	no		Swr1 complex subunit Vps71
SPBC29B5.03c	rpl26	no		60S ribosomal protein L26 (predicted)
SPBC2G2.13c	dcd1	no		deoxycytidylate deaminase (predicted)
SPBC2G5.02c	ckb2	no		CK2 family regulatory subunit Ckb2 (predicted)
SPBC2G5.03 (#2, #3)	ctu1	no		cytosolic thiouridylase subunit Ctu1
SPBC30B4.03c	ldb1	no		LIM domain binding protein, transcription co-repressor

Table D.1 continued from previous page

SystematicID	Gene	Validated	CLS	Product
SPBC31F10.12	tma20	no		RNA-binding protein Tma20 (predicted)
SPBC32F12.02	rec14	no		Ski complex subunit Rec14
SPBC32H8.07	git5	no		heterotrimeric G protein beta (WD repeat) subunit Git5
SPBC337.03	rhn1	no		RNA polymerase II transcription termination factor homolog
SPBC36.03c (#5)	mfs3	no		plasma membrane spermidine transmembrane transporter Mfs3
SPBC3B8.04c		no		phosphate (Pi) transmembrane transporter (predicted)
SPBC4F6.12	pxl1	no		paxillin-like protein Pxl1
SPBC725.02	mpr1	yes	long	histidine-containing response regulator phosphotransferase Mpr1
SPBC83.10	ecm7	no		ER membrane protein complex subunit Ecm7 (predicted)
SPBC83.11	pet2	no		Golgi phosphoenolpyruvate transmembrane transporter Pet2
SPBC839.11c	hut1	yes	long	ER uridine diphosphate-glucose transmembrane transporter Hut1
SPBC9B6.03		yes	long	zf-FYVE type zinc finger protein, involved in endosomal transport
SPBP18G5.03	toc1	no		Tor complex Tor2 interacting protein 1
SPBP35G2.10	mit1	no		SHREC complex ATP-dependent DNA helicase subunit Mit1
SPBP35G2.13c	swc2	no		Swr1 complex subunit Swc2
SPBP4H10.13 (#5)	rps2302	no		40S ribosomal protein S23 (predicted)
SPBP4H10.16c (#4)		no		G-patch RNA-binding protein, involved in splicing (predicted)
SPBP8B7.06 (#5)	rpp201	no		60S acidic ribosomal protein A2
SPBP8B7.25	cyp4	yes	long	cyclophilin family peptidyl-prolyl cis-trans isomerase Cyp4
SPBPJ4664.06	gpt1	no		UDP-glucose-glycoprotein glucosyltransferase Gpt1

Table D.1 continued from previous page

SystematicID	Gene	Validated	CLS	Product
SPCC1223.11	ptc2	no		MAP kinase threonine phosphatase, protein phosphatase 2c homolog Ptc2
SPCC126.09	zip2	no		vacuolar zinc exporter, ZIP family, Zip2 (predicted)
SPCC1494.08c		yes	long	cortical variant C2 domain protein, implicated in signalling or endocytosis
SPCC162.10 (#3)	ppk33	no		serine/threonine protein kinase Ppk33 (predicted)
SPCC1753.02c	git3	yes	long	G-protein coupled receptor Git3
SPCC18.15	dph7	no		diphthamide biosynthesis complex WD repeat protein subunit Dph7 (predicted)
SPCC18.17c		yes	short	proteasome assembly chaperone (predicted)
SPCC1827.04	vms1	yes	long	ATPase complex subunit involved in ER ubiquitin-dependent protein catabolic process (predicted)
SPCC191.09c	gst1	no		glutathione S-transferase Gst1
SPCC1919.05	ski3	no		Ski complex TPR repeat subunit Ski3 (predicted)
SPCC23B6.03c	tel1	no		ATM checkpoint kinase
SPCC285.10c	ear1	no		specificity factor required for ubiquitination Ear1 (predicted)
SPCC417.06c	mug27	no		meiosis specific NDR family protein kinase Mug27/Slk1
SPCC4B3.15	mid1	no		medial ring protein Mid1
SPCC550.01c	coa4	no		cytochrome c oxidase assembly protein Coa4 (predicted)
SPCC550.03c	ski2	no		Ski complex RNA helicase Ski2 (predicted)
SPCC550.12	arp6	no		actin-like protein Arp6
SPCC553.03 (#1)	pex1	no		AAA family ATPase Pex1 (predicted)
SPCC576.13	swc5	no		Swr1 complex subunit Swc5
SPCC584.15c		no		arrestin involved in ubiquitin-dependent endocytosis

Table D.1 continued from previous page

SystematicID	Gene	Validated	CLS	Product
SPCC736.09c	tfx1	no		TRAX
SPCC74.06	mak3	no		histidine kinase Mak3
SPCC794.11c	ent3	no		ENTH/VHS domain protein Ent3 (predicted)
SPCC962.04	rps1201	no		40S ribosomal protein S12 (predicted)
SPCC970.05 (#5)	rpl3601	no		60S ribosomal protein L36
SPCC970.06	erv29	no		COP II adaptor Erv29 (predicted)
SPCP1E11.02	ppk38	no		Ark1/Prk1 family protein kinase Ppk38
SPCP1E11.09c (#5)	rpp103	no		60S acidic ribosomal protein Rpp1-3
SPCPB16A4.06c (#5)		no		Schizosaccharomyces specific protein

Note that the #1, #2, #3, #4 and #5 labelled genes refer to (Lie et al., 2018; Doi et al., 2015; Sideri et al., 2015; Rallis et al., 2014, 2013) and denote the CLS genes common to the previously published deletion screens as referenced.

References

Aranda-Anzaldo, A. (2012). The post-mitotic state in neurons correlates with a stable nuclear higher-order structure. *Communicative & integrative biology*, 5(2):134–9.

Aravind, L., Watanabe, H., Lipman, D. J., and Koonin, E. V. (2000). Lineage-specific loss and divergence of functionally linked genes in eukaryotes. *Proceedings of the National Academy of Sciences*, 97(21):11319–11324.

Atkinson, S. R., Marguerat, S., Bitton, D. A., Rodríguez-López, M., Rallis, C., Lemay, J.-F., Cotobal, C., Malecki, M., Smialowski, P., Mata, J., Korber, P., Bachand, F., and Bähler, J. (2018). Long noncoding RNA repertoire and targeting by nuclear exosome, cytoplasmic exonuclease, and RNAi in fission yeast. *RNA (New York, N.Y.)*, 24(9):1195–1213.

Bacun-Druzina, V., Cagalj, Z., and Gjuracic, K. (2007). The growth advantage in stationary-phase (GASP) phenomenon in mixed cultures of enterobacteria. *FEMS Microbiology Letters*, 266(1):119–27.

Bähler, J. and Wood, V. (2006). Probably the best model organism in the world. *Yeast*, 23(13):899–900.

Barbosa, M. C., Grosso, R. A., and Fader, C. M. (2019). Hallmarks of aging: An autophagic perspective.

Barker, M. G. and Walmsley, R. M. (1999). Replicative ageing in the fission yeast *Schizosaccharomyces Pombe*. *Yeast*, 15(14):1511–1518.

Behringer, M. G. and Hall, D. W. (2016). Genome-Wide Estimates of Mutation Rates and Spectrum in *Schizosaccharomyces Pombe* Indicate CpG Sites are Highly Mutagenic Despite the Absence of DNA Methylation. *G3; Genes—Genomes—Genetics*, 6(1):149–160.

Bhullar, K. S. and Hubbard, B. P. (2015). Lifespan and healthspan extension by resveratrol. *Biochimica et Biophysica Acta - Molecular Basis of Disease*, 1852(6):1209–1218.

Bitton, D. A., Schubert, F., Dey, S., Okoniewski, M., Smith, G. C., Khadayate, S., Pancaldi, V., Wood, V., and Bähler, J. (2015). AnGeLi: A Tool for the Analysis of Gene Lists from Fission Yeast. *Frontiers in genetics*, 6:330.

Borbolis, F. and Syntichaki, P. (2015). Cytoplasmic mRNA turnover and ageing. *Mechanisms of Ageing and Development*, 152:32–42.

Botstein, D., Chervitz, S. A., and Cherry, J. M. (1997). Yeast as a model organism. *Science (New York, N.Y.)*, 277(5330):1259–60.

Bridge and May (1984). A Numerical Classification of Fission Yeasts of the Genus *Schizosaccharomyces Lindner*. *Journal of General Microbiology*, 130:1921–1932.

Brooks, L. R. K. and Mias, G. I. (2019). Data-Driven Analysis of Age, Sex, and Tissue Effects on Gene Expression Variability in Alzheimer's Disease. *Frontiers in neuroscience*, 13:392.

Brutinel, E. D. and Gralnick, J. A. (2012). Anomalies of the anaerobic tricarboxylic acid cycle in *Shewanella oneidensis* revealed by Tn-seq. *Molecular Microbiology*, 86(2):273–283.

Buckingham, S. D., Esmaeili, B., Wood, M., and Sattelle, D. B. (2004). RNA interference: From model organisms towards therapy for neural and neuromuscular disorders.

Bühler, M., Spies, N., Bartel, D. P., and Moazed, D. (2008). TRAMP-mediated RNA surveillance prevents spurious entry of RNAs into the *Schizosaccharomyces Pombe* siRNA pathway. *Nature Structural & Molecular Biology*, 15(10):1015–1023.

Buzzini, P., Lachance, M.-A., and Yurkov, A. (2017). *Yeasts in natural ecosystems: ecology*. Springer.

Calvo, I. A., Gabrielli, N., Iglesias-Baena, I., García-Santamarina, S., Hoe, K. L., Kim, D. U., Sansó, M., Zuin, A., Pérez, P., Ayté, J., and Hidalgo, E. (2009). Genome-wide screen of genes required for caffeine tolerance in fission yeast. *PLOS ONE*, 4(8):6619.

Carette, J. E., Guimaraes, C. P., Wuethrich, I., Blomen, V. A., Varadarajan, M., Sun, C., Bell, G., Yuan, B., Muellner, M. K., Nijman, S. M., Ploegh, H. L., and Brummelkamp, T. R. (2011). Global gene disruption in human cells to assign genes to phenotypes by deep sequencing. *Nature Biotechnology*, 29(6):542–546.

Chan, K.-L., Ho, C.-L., Namasivayam, P., and Napis, S. (2007). A simple and rapid method for RNA isolation from plant tissues with high phenolic compounds and polysaccharides. *Protocol Exchange*.

Chen, B.-R., Li, Y., Eisenstatt, J. R., and Runge, K. W. (2013). Identification of a Lifespan Extending Mutation in the *Schizosaccharomyces pombe* Cyclin Gene clg1^{+} by Direct Selection of Long-Lived Mutants. *PLOS ONE*, 8(7).

Chen, H. I., Chiu, Y. C., Zhang, T., Zhang, S., Huang, Y., and Chen, Y. (2018). GSAE: An autoencoder with embedded gene-set nodes for genomics functional characterization. *BMC Systems Biology*, 12(8).

Chiron, S., Gaisne, M., Guillou, E., Belenguer, P., Clark-Walker, G. D., and Bonnefoy, N. (2007). Studying Mitochondria in an Attractive Model: *Schizosaccharomyces Pombe*. 372:91–105.

Cruaud, P., Rasplus, J.-Y., Rodriguez, L. J., and Cruaud, A. (2017). High-throughput sequencing of multiple amplicons for barcoding and integrative taxonomy. *Scientific reports*, 7:41948.

de Magalhães, J. P., Wuttke, D., Wood, S. H., Plank, M., and Vora, C. (2012). Genome-environment interactions that modulate aging: powerful targets for drug discovery. *Pharmacological reviews*, 64(1):88–101.

Delewski, W., Paterkiewicz, B., Manicki, M., Schilke, B., Tomiczek, B., Ciesielski, S. J., Nierzwicki, L., Czub, J., Dutkiewicz, R., Craig, E. A., and Marszalek, J. (2016). Iron-Sulfur Cluster Biogenesis Chaperones: Evidence for Emergence of Mutational Robustness of a Highly Specific Protein-Protein Interaction. *Molecular biology and evolution*, 33(3):643–56.

Delneri, D. (2010). Barcode technology in yeast: application to pharmacogenomics. *FEMS Yeast Research*, 10(8):1083–1089.

Deshpande, G. P., Hayles, J., Hoe, K.-L., Kim, D.-U., Park, H.-O., and Hartsuiker, E. (2009). Screening a genome-wide *S. pombe* deletion library identifies novel genes and pathways involved in genome stability maintenance. *DNA Repair*, 8(5):672–679.

Deutschbauer, A. M., Williams, R. M., Chu, A. M., and Davis, R. W. (2002). Parallel phenotypic analysis of sporulation and postgermination growth in *Schizosaccharomyces cerevisiae*. *Proceedings of the National Academy of Sciences of the United States of America*, 99(24):15530.

Doi, A., Fujimoto, A., Sato, S., Uno, T., Kanda, Y., Asami, K., Tanaka, Y., Kita, A., Satoh, R., and Sugiura, R. (2015). Chemical genomics approach to identify genes associated with sensitivity to rapamycin in the fission yeast *Schizosaccharomyces Pombe*. *Genes to Cells*, 20(4):292–309.

Duina, A. A., Miller, M. E., and Keeney, J. B. (2014). Budding Yeast for Budding Geneticists: A Primer on the *Schizosaccharomyces cerevisiae* Model System. *Genetics*, 197(1):33–48.

Dutkiewicz, R. and Nowak, M. (2018). Molecular chaperones involved in mitochondrial iron-sulfur protein biogenesis. *Journal of biological inorganic chemistry: JBIC: a publication of the Society of Biological Inorganic Chemistry*, 23(4):569–579.

E. Yanos, M., F. Bennett, C., and Kaeberlein, M. (2012). Genome-Wide RNAi Longevity Screens in Caenorhabditis elegans. *Current Genomics*, 13(7):508–518.

Eason, R. G., Pourmand, N., Tongprasit, W., Herman, Z. S., Anthony, K., Jejelowo, O., Davis, R. W., and Stolc, V. (2004). Characterization of synthetic DNA bar codes in *Schizosaccharomyces cerevisiae* gene-deletion strains. *Proceedings of the National Academy of Sciences of the United States of America*, 101(30):11046–51.

Erjavec, N., Cvijovic, M., Klipp, E., and Nyström, T. (2008). Selective benefits of damage partitioning in unicellular systems and its effects on aging. *Proceedings of the National Academy of Sciences of the United States of America*, 105(48):18764–18769.

Fabrizio, P., Hoon, S., Shamalnasab, M., Galbani, A., Wei, M., Giaever, G., Nislow, C., and Longo, V. D. (2010). Genome-Wide Screen in *Schizosaccharomyces cerevisiae* Identifies Vacuolar Protein Sorting, Autophagy, Biosynthetic, and tRNA Methylation Genes Involved in Life Span Regulation. *PLOS Genetics*, 6(7):1001024.

Fabrizio, P. and Longo, V. D. (2003). The chronological life span of *Saccharomyces cerevisiae*. *Aging Cell*, 2(2):73–81.

Fang, Y., Hu, L., Zhou, X., Jaiseng, W., Zhang, B., Takami, T., and Kuno, T. (2012). A genomewide screen in *Schizosaccharomyces Pombe* for genes affecting the sensitivity of antifungal drugs that target ergosterol biosynthesis. *Antimicrobial agents and chemotherapy*, 56(4):1949–59.

Fantes, P. A. and Hoffman, C. S. (2016). A Brief History of *Schizosaccharomyces Pombe* Research: A Perspective Over the Past 70 Years. *Genetics*, 203(2):621–629.

Farlow, A., Long, H., Arnoux, S., Sung, W., Doak, T. G., Nordborg, M., and Lynch, M. (2015). The Spontaneous Mutation Rate in the Fission Yeast *Schizosaccharomyces Pombe*. *Genetics*, 201(2):737–44.

Fontana, L., Partridge, L., and Longo, V. D. (2010). Extending healthy life span–from yeast to humans. *Science (New York, N.Y.)*, 328(5976):321–6.

Forsburg, S. L. (2011a). Fission Yeast Cell Cycle PombeNet Forsburg Lab USC Dana and David Dornsife College of Letters, Arts and Sciences.

Forsburg, S. L. (2011b). Forsburg Lab Recipes: Fission Yeast Media.

Forsburg, S. L. and Rhind, N. (2006). Basic methods for fission yeast. *Yeast*, 23(3):173–183.

Fruhmann, G., Seynnaeve, D., Zheng, J., Ven, K., Molenberghs, S., Wilms, T., Liu, B., Winderickx, J., and Franssens, V. (2017). Yeast buddies helping to unravel the complexity of neurodegenerative disorders. *Mechanisms of Ageing and Development*, 161(Pt B):288–305.

Gallagher, L. A., Shendure, J., and Manoil, C. (2011). Genome-Scale Identification of Resistance Functions in Pseudomonas aeruginosa Using Tn-seq. *mBio*, 2(1):00315–10.

Garay, E., Campos, S. E., González de la Cruz, J., Gaspar, A. P., Jinich, A., and DeLuna, A. (2014). High-Resolution Profiling of Stationary-Phase Survival Reveals Yeast Longevity Factors and Their Genetic Interactions. *PLOS Genetics*, 10(2):1004168.

García, P., Encinar Del Dedo, J., Ayté, J., and Hidalgo, E. (2016). Genome-wide Screening of Regulators of Catalase Expression: ROLE OF A TRAN-

SCRIPTION COMPLEX AND HISTONE AND tRNA MODIFICATION COMPLEXES ON ADAPTATION TO STRESS. *The Journal of biological chemistry*, 291(2):790–9.

Gems, D. and Partridge, L. (2013). Genetics of Longevity in Model Organisms: Debates and Paradigm Shifts. *Annu Rev Physiol.*, (75):621–44.

Giaever, G., Chu, A. M., Ni, L., Connelly, C., Riles, L., Véronneau, S., Dow, S., Lucau-Danila, A., Anderson, K., André, B., Arkin, A. P., Astromoff, A., El Bakkoury, M., Bangham, R., Benito, R., Brachat, S., Campanaro, S., Curtiss, M., Davis, K., Deutschbauer, A., Entian, K.-D., Flaherty, P., Foury, F., Garfinkel, D. J., Gerstein, M., Gotte, D., Güldener, U., Hegemann, J. H., Hempel, S., Herman, Z., Jaramillo, D. F., Kelly, D. E., Kelly, S. L., Kötter, P., LaBonte, D., Lamb, D. C., Lan, N., Liang, H., Liao, H., Liu, L., Luo, C., Lussier, M., Mao, R., Menard, P., Ooi, S. L., Revuelta, J. L., Roberts, C. J., Rose, M., Ross-Macdonald, P., Scherens, B., Schimmack, G., Shafer, B., Shoemaker, D. D., Sookhai-Mahadeo, S., Storms, R. K., Strathern, J. N., Valle, G., Voet, M., Volckaert, G., Wang, C.-y., Ward, T. R., Wilhelmy, J., Winzeler, E. A., Yang, Y., Yen, G., Youngman, E., Yu, K., Bussey, H., Boeke, J. D., Snyder, M., Philippsen, P., Davis, R. W., and Johnston, M. (2002). Functional profiling of the *Schizosaccharomyces cerevisiae* genome. *Nature*, 418(6896):387–391.

Giaever, G. and Nislow, C. (2014). The yeast deletion collection: a decade of functional genomics. *Genetics*, 197(2):451–65.

Goffeau, A., Barrell, B. G., Bussey, H., Davis, R. W., Dujon, B., Feldmann, H., Galibert, F., Hoheisel, J. D., Jacq, C., Johnston, M., Louis, E. J., Mewes, H. W., Murakami, Y., Philippsen, P., Tettelin, H., and Oliver, S. G. (1996). Life with 6000 genes. *Science (New York, N.Y.)*, 274(5287):546, 563–7.

Gómez, E. B. and Forsburg, S. L. (2004). Analysis of the fission yeast *Schizosaccharomyces Pombe* cell cycle. *Methods in molecular biology (Clifton, N.J.)*, 241:93–111.

Grech, L., Jeffares, D. C., Sadée, C. Y., Rodríguez-López, M., Bitton, D. A., Hoti, M., Biagosch, C., Aravani, D., Speekenbrink, M., Illingworth, C. J., Schiffer, P. H., Pidoux, A. L., Tong, P., Tallada, V. A., Allshire, R., Levin, H. L., Bähler, J., and Wilke, C. (2019). Fitness Landscape of the Fission Yeast Genome. *Molecular Biology and Evolution*, 36(8):1612–1623.

Han, T. X., Xu, X.-Y., Zhang, M.-J., Peng, X., and Du, L.-L. (2010). Global fitness profiling of fission yeast deletion strains by barcode sequencing. *Genome Biology*, 11(6):R60.

Henderson, K. A. and Gottschling, D. E. (2008). A mother's sacrifice: what is she keeping for herself? *Current Opinion in Cell Biology*, 20(6):723–728.

Heo, J.-M., Livnat-Levanon, N., Taylor, E. B., Jones, K. T., Dephoure, N., Ring, J., Xie, J., Brodsky, J. L., Madeo, F., Gygi, S. P., Ashrafi, K., Glickman, M. H., and Rutter, J. (2010). A Stress-Responsive System for Mitochondrial Protein Degradation. *Molecular Cell*, 40(3):465–480.

Hobbs, E. C., Astarita, J. L., and Storz, G. (2010). Small RNAs and small proteins involved in resistance to cell envelope stress and acid shock in Escherichia coli: analysis of a bar-coded mutant collection. *Journal of bacteriology*, 192(1):59–67.

Hoffman, C. S., Wood, V., and Fantes, P. A. (2015). An Ancient Yeast for Young Geneticists: A Primer on the *Schizosaccharomyces Pombe* Model System. *Genetics*, 201(2):403–423.

Irelan, J. T., Gutkin, G. I., and Clarke, L. (2001). Functional redundancies, distinct localizations and interactions among three fission yeast homologs of centromere protein-B. *Genetics*, 157(3):1191–203.

Janssens, G. and Veenhoff, L. (2016). Evidence for the hallmarks of human aging in replicatively aging yeast. *Microbial Cell*, 3(7):263–274.

Jeffares, D. C. (2018). The natural diversity and ecology of fission yeast. *Yeast*, 35(3):253–260.

Jeffares, D. C., Rallis, C., Rieux, A., Speed, D., Pevorovský, M., Mourier, T., Marsellach, F. X., Iqbal, Z., Lau, W., Cheng, T. M. K., Pracana, R., Mülleder, M., Lawson, J. L. D., Chessel, A., Bala, S., Hellenthal, G., O'Fallon, B., Keane, T., Simpson, J. T., Bischof, L., Tomiczek, B., Bitton, D. A., Sideri, T., Codlin, S., Hellberg, J. E. E. U., van Trigt, L., Jeffery, L., Li, J.-J., Atkinson, S., Thodberg, M., Febrer, M., McLay, K., Drou, N., Brown, W., Hayles, J., Salas, R. E. C., Ralser, M., Maniatis, N., Balding, D. J., Balloux, F., Durbin, R., and Bähler, J. (2015). The genomic and phenotypic diversity of *Schizosaccharomyces Pombe*. *Nature Genetics*, 47(3):235–241.

Jimenez, J., Bru, S., Ribeiro, M., and Clotet, J. (2015). Live fast, die soon: cell cycle progression and lifespan in yeast cells. *Microbial Cell*, 2(3):62–67.

Jo, M. C., Liu, W., Gu, L., Dang, W., and Qin, L. (2015). High-throughput analysis of yeast replicative aging using a microfluidic system. *Proceedings of the National Academy of Sciences*, 112(30):9364–9369.

Kaeberlein, M. (2010). Lessons on longevity from budding yeast.

Kaeberlein, M., Burtner, C. R., and Kennedy, B. K. (2007). Recent Developments in Yeast Aging. *PLOS Genetics*, 3(5):e84.

Kaeberlein, M., Powers, R. W., Steffen, K. K., Westman, E. A., Hu, D., Dang, N., Kerr, E. O., Kirkland, K. T., Fields, S., and Kennedy, B. K. (2005). Cell biology: Regulation of yeast replicative life span by TOR and Sch9 response to nutrients. *Science*, 310(5751):1193–1196.

Kebschull, J. M. and Zador, A. M. (2018). Cellular barcoding: lineage tracing, screening and beyond. *Nature Methods*, 15(11):871–879.

Kennedy, P. J., Vashisht, A. A., Hoe, K.-L., Kim, D.-U., Park, H.-O., Hayles, J., and Russell, P. (2008). A Genome-Wide Screen of Genes Involved in Cadmium Tolerance in *Schizosaccharomyces Pombe*. *Toxicol Sci.*, 106(1):124–139.

Kenyon, C. J. (2010). The genetics of ageing. *Nature*, 464(7288):504–512.

Kim, D.-M., Kim, H., Yeon, J.-H., Lee, J.-H., and Park, H.-O. (2016). Identification of a Mitochondrial DNA Polymerase Affecting Cardiotoxicity of Sunitinib Using a Genome-Wide Screening on *S. pombe* Deletion Library. *Toxicological Sciences*, 149(1):4–14.

Kim, D.-U., Hayles, J., Kim, D., Wood, V., Park, H.-O., Won, M., Yoo, H.-S., Duhig, T., Nam, M., Palmer, G., Han, S., Jeffery, L., Baek, S.-T., Lee, H., Shim, Y. S., Lee, M., Kim, L., Heo, K.-S., Noh, E. J., Lee, A.-R., Jang, Y.-J., Chung, K.-S., Choi, S.-J., Park, J.-Y., Park, Y., Kim, H. M., Park, S.-K., Park, H.-J., Kang, E.-J., Kim, H. B., Kang, H.-S., Park, H.-M., Kim, K., Song, K., Song, K. B., Nurse, P., and Hoe, K.-L. (2010). Analysis of a genome-wide set of gene deletions in the fission yeast *Schizosaccharomyces Pombe*. *Nature Biotechnology*, 28(6):617–623.

Kim, J.-Y., Kim, E.-J., Lopez-Maury, L., Bähler, J., and Roe, J.-H. (2014). A metabolic strategy to enhance long-term survival by Phx1 through stationary phase-specific pyruvate decarboxylases in fission yeast. *Aging*, 6(7):587–601.

Klass, M. R. (1977). Aging in the nematode Caenorhabditis elegans: Major biological and environmental factors influencing life span. *Mechanisms of Ageing and Development*, 6(C):413–429.

Klass, M. R. (1983). A method for the isolation of longevity mutants in the nematode Caenorhabditis elegans and initial results. *Mechanisms of Ageing and Development*, 22(3-4):279–286.

Lang, B. F., Cedergren, R., and Gray, M. W. (1987). The mitochondrial genome of the fission yeast, *Schizosaccharomyces Pombe*. Sequence of the large-subunit ribosomal RNA gene, comparison of potential secondary structure in fungal mitochondrial large-subunit rRNAs and evolutionary considerations. *European journal of biochemistry*, 169(3):527–37.

Lee, B., Moon, T., Yoon, S., and Weissman, T. (2017). DUDE-Seq: Fast, flex-

ible, and robust denoising for targeted amplicon sequencing. *PLOS ONE*, 12(7):e0181463.

Lee, M., Choi, S.-J., Han, S., Nam, M., Kim, D., Kim, D.-U., and Hoe, K.-L. (2018). Mutation Analysis of Synthetic DNA Barcodes in a Fission Yeast Gene Deletion Library by Sanger Sequencing. *Genomics & Informatics*, 16(2):22–29.

Lee, S. S., Lee, R. Y., Fraser, A. G., Kamath, R. S., Ahringer, J., and Ruvkun, G. (2003). A systematic RNAi screen identifies a critical role for mitochondria in C. elegans longevity. *Nature Genetics*, 33(1):40–48.

Li, J., Yu, Y., Suo, F., Sun, L.-L., Zhao, D., and Du, L.-L. (2014). Genome-wide screens for sensitivity to ionizing radiation identify the fission yeast nonhomologous end joining factor Xrc4. *G3 (Bethesda, Md.)*, 4(7):1297–306.

Li, L., Miles, S., and Breeden, L. L. (2015). A genetic screen for Saccharomyces cerevisiae mutants that fail to enter quiescence. *G3: Genes, Genomes, Genetics*, 5(8):1783–1795.

Li, L.-L., Suo, M., Liu, F., and Shen, X.-M. (2013). Global Analysis of Fission Yeast Mating Genes Reveals New Autophagy Factors. *PLoS Genet*, 9(8):1003715.

Lie, S., Banks, P., Lawless, C., Lydall, D., and Petersen, J. (2018). The contribution of non-essential *Schizosaccharomyces Pombe* genes to fitness in response to altered nutrient supply and target of rapamycin activity. *Open Biology*, 8(5):180015.

Lin, S.-J. and Austriaco, N. (2014). Aging and cell death in the other yeasts, *Schizosaccharomyces Pombe* and *Candida albicans*. *FEMS Yeast Research*, 14(1):119–135.

Lock, A., Rutherford, K., Harris, M. A., Hayles, J., Oliver, S. G., Bähler, J., and Wood, V. (2019). PomBase 2018: user-driven reimplementation of the fission

yeast database provides rapid and intuitive access to diverse, interconnected information. *Nucleic Acids Research*, 47(D1):D821–D827.

Longo, V., Shadel, G., Kaeberlein, M., and Kennedy, B. (2012). Replicative and Chronological Aging in *Schizosaccharomyces cerevisiae*. *Cell Metabolism*, 16(1):18–31.

Longo, V. D. and Finch, C. E. (2003). Evolutionary Medicine: From Dwarf Model Systems to Healthy Centenarians? *Science*, 299(5611):1342–1346.

López-Otín, C., Blasco, M. A., Partridge, L., Serrano, M., and Kroemer, G. (2013). The Hallmarks of Aging. *Cell*, 153(6):1194–1217.

Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, 15(12):550.

Madeo, F., Carmona-Gutierrez, D., Kepp, O., and Kroemer, G. (2018). Spermidine delays aging in humans. *Aging*, 10(8):2209–2211.

Malecki, M. and Bähler, J. (2016). Identifying genes required for respiratory growth of fission yeast. *Wellcome Open Research*, 1:12.

Malecki, M., Bitton, D. A., Rodríguez-López, M., Rallis, C., Calavia, N. G., Smith, G. C., and Bähler, J. (2016). Functional and regulatory profiling of energy metabolism in fission yeast. *Genome Biology*, 17(1):240.

Marguerat, S., Schmidt, A., Codlin, S., Chen, W., Aebersold, R., and Bähler, J. (2012). Quantitative Analysis of Fission Yeast Transcriptomes and Proteomes in Proliferating and Quiescent Cells. *Cell*, 151(3):671–683.

Martini, A. V. (1991). Evaluation of phylogenetic relationships among fission yeast by nDNA/nDNA reassociation and conventional taxonomic criteria. *Yeast*, 7(1):73–78.

Masuda, F., Ishii, M., Mori, A., Uehara, L., Yanagida, M., Takeda, K., and Saitoh, S. (2016). Glucose restriction induces transient G2 cell cycle arrest extending cellular chronological lifespan. *Scientific Reports*, 6(1):19629.

Matecic, M., Smith, D. L., Pan, J., Maqani, X., and Bekiranov, N. (2010). A Microarray-Based Genetic Screen for Yeast Chronological Aging Factors. *PLOS Genetics*, 6(4):1000921.

McCurry, J. (2015). Japan will be model for future super-ageing societies. *Lancet (London, England)*, 386(10003):1523.

Mesarich, C. H., Rees-George, J., Gardner, P. P., Ghomi, F. A., Gerth, M. L., Andersen, M. T., Rikkerink, E. H. A., Fineran, P. C., and Templeton, M. D. (2017). Transposon insertion libraries for the characterization of mutants from the kiwifruit pathogen *Pseudomonas syringae pv. actinidiae*. *PLOS ONE*, 12(3):0172790.

Minois, N., Frajnt, M., Dölling, M., Lagona, F., Schmid, M., Küchenhoff, H., Gampe, J., and Vaupel, J. W. (2006). Symmetrically Dividing Cells of the Fission Yeast *Schizosaccharomyces Pombe* Do Age. *Biogerontology*, 7(4):261–267.

Mitchison, J. M. and Nurse, P. (1985). Growth in cell length in the fission yeast *Schizosaccharomyces Pombe*. Technical report.

Mizuguchi, T., Barrowman, J., and Grewal, S. I. (2015). Chromosome domain architecture and dynamic organization of the fission yeast genome. *FEBS Letters*, 589(20):2975–2986.

Mochida, S. and Yanagida, M. (2005). Distinct modes of DNA damage response in S. pombe G0 and vegetative cells. *Genes to Cells*, 11(1):13–27.

Mun, J., Kim, D.-U., Hoe, K.-L., and Kim, S.-Y. (2016). Genome-wide functional analysis using the barcode sequence alignment and statistical analysis (Barcas) tool. *BMC Bioinformatics*, 17:475.

Nakaoka, H. and Wakamoto, Y. (2017). Aging, mortality, and the fast growth trade-off of *Schizosaccharomyces Pombe*. *PLOS Biology*, 15(6):e2001109.

Neff, F., Flores-Dominguez, D., Ryan, D. P., Horsch, M., Schröder, S., Adler, T., Afonso, L. C., Aguilar-Pimentel, J. A., Becker, L., Garrett, L., Hans, W., Hettich,

M. M., Holtmeier, R., Hölter, S. M., Moreth, K., Prehn, C., Puk, O., Rácz, I., Rathkolb, B., Rozman, J., Naton, B., Ordemann, R., Adamski, J., Beckers, J., Bekeredjian, R., Busch, D. H., Ehninger, G., Graw, J., Höfler, H., Klingenspor, M., Klopstock, T., Ollert, M., Stypmann, J., Wolf, E., Wurst, W., Zimmer, A., Fuchs, H., Gailus-Durner, V., De Angelis, M. H., and Ehninger, D. (2013). Rapamycin extends murine lifespan but has limited effects on aging. *Journal of Clinical Investigation*, 123(8):3272–3291.

Niccoli, T. and Partridge, L. (2012). Ageing as a Risk Factor for Disease. *Current Biology*, 22(17):R741–R752.

Nielsen, O. (2008). Mating type tutorial.

Nielson, J. R., Fredrickson, E. K., Waller, T. C., Rendón, O. Z., Schubert, H. L., Lin, Z., Hill, C. P., and Rutter, J. (2017). Sterol Oxidation Mediates Stress-Responsive Vms1 Translocation to Mitochondria. *Molecular Cell*, 68(4):673–685.e6.

Nurse, P. (1990). Universal control mechanism regulating onset of M-phase. *Nature*, 344(6266):503–508.

Oeppen, J. and Vaupel, J. W. (2002). Demography: Broken limits to life expectancy.

Okuyama, T., Inoue, H., Ookuma, S., Satoh, T., Kano, K., Honjoh, S., Hisamoto, N., Matsumoto, K., and Nishida, E. (2010). The ERK-MAPK pathway regulates longevity through SKN-1 and insulin-like signaling in Caenorhabditis elegans. *Journal of Biological Chemistry*, 285(39):30274–30281.

Paddison, P. J., Silva, J. M., Conklin, D. S., Schlabach, M., Li, M., Aruleba, S., Balija, V., O'Shaughnessy, A., Gnoj, L., Scoble, K., Chang, K., Westbrook, T., Cleary, M., Sachidanandam, R., McComble, W. R., Elledge, S. J., and Hannon, G. J. (2004). A resource for large-scale RNA-interference-based screens in mammals. *Nature*, 428(6981):427–431.

Pan, X., Lei, B., Zhou, N., Feng, B., Yao, W., Zhao, X., Yu, Y., and Lu, H. (2012). Identification of novel genes involved in DNA damage response by screening

a genome-wide *Schizosaccharomyces Pombe* deletion library. *BMC Genomics*, 13(1):662.

Partridge, L. (2010). The new biology of ageing. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 365(1537):147–54.

Paul, L. (1893). *Schizosaccharomyces Pombe* sp. nov., a New Ferment. *Wochenschrift fr Brauerei*, 10:451–65.

Piel, M. and Tran, P. T. (2009). Cell shape and cell division in fission yeast. *Current Biology : CB*, 19(17):R823–7.

Powers, R. W., Kaeberlein, M., Caldwell, S. D., Kennedy, B. K., Fields, S., and Fields, S. (2006). Extension of chronological life span in yeast by decreased TOR pathway signaling. *Genes & Development*, 20(2):174–84.

Rallis, C., Codlin, S., and Bähler, J. (2013). TORC1 signaling inhibition by rapamycin and caffeine affect lifespan, global gene expression, and cell proliferation of fission yeast. *Aging cell*, 12(4):563–73.

Rallis, C., López-Maury, L., Georgescu, T., Pancaldi, V., and Bähler, J. (2014). 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.

Rappaport, N., Twik, M., Plaschkes, I., Nudel, R., InyStein, T., Levitt, J., Gershoni, M., Morrey, C. P., Safran, M., and Lancet, D. (2017). MalaCards: an amalgamated human disease compendium with diverse clinical and genetic annotation and structured search. *Nucleic Acids Research*, 45(D1):D877–D887.

Rebelo-Marques, A., Lages, A. D. S., Andrade, R., Ribeiro, C. F., Mota-Pinto, A., Carrilho, F., and Espregueira-Mendes, J. (2018). Aging hallmarks: The benefits of physical exercise.

Robinson, D. G., Chen, W., Storey, J. D., and Gresham, D. (2014). Design and analysis of Bar-seq experiments. *G3 (Bethesda, Md.)*, 4(1):11–8.

Roche, B., Arcangioli, B., and Martienssen, R. (2017). Transcriptional reprogramming in cellular quiescence. *RNA biology*, 14(7):843–853.

Rousset, F., Cui, L., Siouve, E., Becavin, C., Depardieu, F., and Bikard, D. (2018). Genome-wide CRISPR-dCas9 screens in E. coli identify essential genes and phage host factors. *PLoS Genetics*, 14(11).

Roux, A. E., Chartrand, P., Ferbeyre, G., and Rokeach, L. A. (2010). Fission Yeast and Other Yeasts as Emergent Models to Unravel Cellular Aging in Eukaryotes. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, 65A(1):1–8.

Roux, A. E., Leroux, A., Alaamery, M. A., Hoffman, C. S., Chartrand, P., Ferbeyre, G., and Rokeach, L. A. (2009). Pro-Aging Effects of Glucose Signaling through a G Protein-Coupled Glucose Receptor in Fission Yeast. *PLOS Genetics*, 5(3).

Roux, A. E., Quissac, A., Chartrand, P., Ferbeyre, G., and Rokeach, L. A. (2006). Regulation of chronological aging in *Schizosaccharomyces Pombe* by the protein kinases Pka1 and Sck2. *Aging Cell*, 5(4):345–357.

Sideri, T., Rallis, C., Bitton, D. A., Lages, B. M., Suo, F., Rodríguez-López, M., Du, L.-L., and Bähler, J. (2015). Parallel Profiling of Fission Yeast Deletion Mutants for Proliferation and for Lifespan During Long-Term Quiescence. *G3; Genes—Genomes—Genetics*, 5(1):145–155.

Simpkins, S. W., Deshpande, R., Nelson, J., Li, S. C., Piotrowski, J. S., Ward, H. N., Yashiroda, Y., Osada, H., Yoshida, M., Boone, C., and Myers, C. L. (2019). Using BEAN-counter to quantify genetic interactions from multiplexed barcode sequencing experiments. *Nature Protocols*, 14(2):415–440.

Sinha, A. and Rae, R. (2016). Genome-wide RNAi screens in C. Elegans to identify genes influencing lifespan and innate immunity. In *Methods in Molecular Biology*, volume 1470, pages 171–182. Humana Press Inc.

Sipiczki, M. (2000). Where does fission yeast sit on the tree of life? *Genome biology*, 1(2):1011.

Sipiczki, M., Kucsera, J., Ulaszewski, S., and Zsolt, J. (1982). Hybridization Studies by Crossing and Protoplast Fusion within the Genus Schizosaccharomyces Lindner. *General Microbiology*, 128:1989—-2000.

Slack, C. (2017). Ras signaling in aging and metabolic regulation.

Smith, A. M., Heisler, L. E., Mellor, J., Kaper, F., Thompson, M. J., Chee, M., Roth, F. P., Giaever, G., and Nislow, C. (2009). Quantitative phenotyping via deep barcode sequencing. *Genome Research*, 19(10):1836–1842.

Smith, C. L., Matsumoto, T., Niwa, O., Klco, S., Fan, J. B., Yanagida, M., and Cantor, C. R. (1987). An electrophoretic karyotype for *Schizosaccharomyces Pombe* by pulsed field gel electrophoresis. *Nucleic acids research*, 15(11):4481–9.

Smith, D. L., Maharrey, C. H., Carey, C. R., White, R. A., Hartman, J. L., Hartman, J. L., and IV (2016). Gene-nutrient interaction markedly influences yeast chronological lifespan. *Experimental gerontology*, 86:113–123.

Smith, J. T., White, J. W., Dungrawala, H., Hua, H., and Schneider, B. L. (2018). Yeast lifespan variation correlates with cell growth and SIR2 expression. *PLOS ONE*, 13(7):e0200275.

Spivey, E., Jones, S., Rybarski, J., Saifuddin, F., and Finkelstein, I. (2016). An aging-independent replicative lifespan in a symmetrically dividing eukaryote. *bioRxiv*, page 064832.

Steinkraus, K., Kaeberlein, M., and Kennedy, B. (2008). Replicative Aging in Yeast: The Means to the End. *Annual review of cell and developmental biology*, 24:29.

Stratton, P. (2013). The clinical consequences of an ageing world and preventive strategies. *Best Practice & Research Clinical Obstetrics & Gynaecology*, 27(5):643–659.

Su, S. S., Tanaka, Y., Samejima, I., Tanaka, K., and Yanagida, M. (1996). A nitrogen starvation-induced dormant G0 state in fission yeast: the establishment from uncommitted G1 state and its delay for return to proliferation. *Journal of cell science*, 109 (Pt 6:1347–57.

Takeda, K., Mori, A., and Yanagida, M. (2011). Identification of Genes Affecting the Toxicity of Anti-Cancer Drug Bortezomib by Genome-Wide Screening in *S. pombe*. *PLOS ONE*, 6(7):220–21.

Takeda, K., Yoshida, T., Kikuchi, S., Nagao, K., Kokubu, A., Pluskal, T., Villar-Briones, A., Nakamura, T., and Yanagida, M. (2010). Synergistic roles of the proteasome and autophagy for mitochondrial maintenance and chronological lifespan in fission yeast. *Proceedings of the National Academy of Sciences*, 107(8):3540–3545.

Tissenbaum, H. A. and Guarente, L. (2002). Model Organisms as a Guide to Mammalian Aging. *Developmental Cell*, 2(1):9–19.

Turcotte, B., Liang, X. B., Robert, F., and Soontorngun, N. (2010). Transcriptional regulation of nonfermentable carbon utilization in budding yeast. *FEMS Yeast Research*, 10(1):2–13.

Ucisik-Akkaya, E., Leatherwood, J. K., and Neiman, A. M. (2014). A Genome-Wide Screen for Sporulation-Defective Mutants in *Schizosaccharomyces Pombe*. *G3; Genes—Genomes—Genetics*, 4(6):1173–1182.

Vivancos, A. P., Jara, M., Zuin, A., Sansó, M., and Hidalgo, E. (2006). Oxidative stress in *Schizosaccharomyces Pombe*: different H_2O_2 levels, different response pathways. *Molecular Genetics and Genomics*, 276(6):495–502.

Voisine, C., Cheng, Y. C., Ohlson, M., Schilke, B., Hoff, K., Beinert, H., Marszalek, J., and Craig, E. A. (2001). Jac1, a mitochondrial J-type chaperone, is involved in the biogenesis of Fe/S clusters in *Schizosaccharomyces cerevisiae*. *Proceedings of the National Academy of Sciences*, 98(4):1483–1488.

Wagih, O., Usaj, M., Baryshnikova, A., VanderSluis, B., Kuzmin, E., Costanzo, M., Myers, C. L., Andrews, B. J., Boone, C. M., and Parts, L. (2013). SGAtools: one-stop analysis and visualization of array-based genetic interaction screens. *Nucleic acids research*, 41:591–6.

Wallace, D. C. (2005). A Mitochondrial Paradigm of Metabolic and Degenerative Diseases, Aging, and Cancer: A Dawn for Evolutionary Medicine. *Annual Review of Genetics*, 39(1):359–407.

Walsh, C. and Cepko, C. (1988). Clonally related cortical cells show several migration patterns. *Science*, 241(4871):1342–1345.

Wasko, B. M. and Kaeberlein, M. (2014). Yeast replicative aging: a paradigm for defining conserved longevity interventions. *FEMS Yeast Research*, 14(1):148–159.

Wetmore, K. M., Price, M. N., Waters, R. J., Lamson, J. S., He, J., Hoover, C. A., Blow, M. J., Bristow, J., Butland, G., Arkin, A. P., and Deutschbauer, A. (2015). Rapid quantification of mutant fitness in diverse bacteria by sequencing randomly bar-coded transposons. *mBio*, 6(3):e00306–15.

Wilkinson, M. J., Szabo, C., Ford, C. S., Yarom, Y., Croxford, A. E., Camp, A., and Gooding, P. (2017). Replacing Sanger with Next Generation Sequencing to improve coverage and quality of reference DNA barcodes for plants. *Scientific Reports*, 7(1):46040.

Wood, V., Gwilliam, R., Rajandream, M.-A., Lyne, M., Lyne, R., Stewart, A., Sgouros, J., Peat, N., Hayles, J., Baker, S., Basham, D., Bowman, S., Brooks, K., Brown, D., Brown, S., Chillingworth, T., Churcher, C., Collins, M., Connor,

R., Cronin, A., Davis, P., Feltwell, T., Fraser, A., Gentles, S., Goble, A., Hamlin, N., Harris, D., Hidalgo, J., Hodgson, G., Holroyd, S., Hornsby, T., Howarth, S., Huckle, E. J., Hunt, S., Jagels, K., James, K., Jones, L., Jones, M., Leather, S., McDonald, S., McLean, J., Mooney, P., Moule, S., Mungall, K., Murphy, L., Niblett, D., Odell, C., Oliver, K., O'Neil, S., Pearson, D., Quail, M. A., Rabbinowitsch, E., Rutherford, K., Rutter, S., Saunders, D., Seeger, K., Sharp, S., Skelton, J., Simmonds, M., Squares, R., Squares, S., Stevens, K., Taylor, K., Taylor, R. G., Tivey, A., Walsh, S., Warren, T., Whitehead, S., Woodward, J., Volckaert, G., Aert, R., Robben, J., Grymonprez, B., Weltjens, I., Vanstreels, E., Rieger, M., Schäfer, M., Müller-Auer, S., Gabel, C., Fuchs, M., Fritzc, C., Holzer, E., Moestl, D., Hilbert, H., Borzym, K., Langer, I., Beck, A., Lehrach, H., Reinhardt, R., Pohl, T. M., Eger, P., Zimmermann, W., Wedler, H., Wambutt, R., Purnelle, B., Goffeau, A., Cadieu, E., Dréano, S., Gloux, S., Lelaure, V., Mottier, S., Galibert, F., Aves, S. J., Xiang, Z., Hunt, C., Moore, K., Hurst, S. M., Lucas, M., Rochet, M., Gaillardin, C., Tallada, V. A., Garzon, A., Thode, G., Daga, R. R., Cruzado, L., Jimenez, J., Sánchez, M., del Rey, F., Benito, J., Domínguez, A., Revuelta, J. L., Moreno, S., Armstrong, J., Forsburg, S. L., Cerrutti, L., Lowe, T., McCombie, W. R., Paulsen, I., Potashkin, J., Shpakovski, G. V., Ussery, D., Barrell, B. G., Nurse, P., Nurse, P., and Cerrutti, L. (2002). Erratum: The genome sequence of *Schizosaccharomyces Pombe*. *Nature*, 415(6874):871–880.

Wood, V., Harris, M. A., McDowall, M. D., Rutherford, K., Vaughan, B. W., Staines, D. M., Aslett, M., Lock, A., Bahler, J., Kersey, P. J., and Oliver, S. G. (2012). PomBase: a comprehensive online resource for fission yeast. *Nucleic Acids Research*, 40(47):695–699.

Wood, V., Lock, A., Harris, M. A., Rutherford, K., Bähler, J., and Oliver, S. G. (2019). Hidden in plain sight: what remains to be discovered in the eukaryotic proteome? *Open Biology*, 9(2):180241.

Yanagida, M. (2002). The model unicellular eukaryote, *Schizosaccharomyces Pombe*. *Genome biology*, 3(3):2003.1–2003.4.

Yanagida, M. (2009). Cellular quiescence: are controlling genes conserved? *Trends in Cell Biology*, 19(12):705–715.

Yang, Y., Liu, Q., Jiang, G., Chen, S., Zhou, L., Sakamoto, N., Kuno, T., Fang, Y., and Yao, F. (2018). Genome-wide screen reveals important roles for ESCRT proteins in drug/ion resistance of fission yeast. *PLOS ONE*, 13(6):e0198516.

Yao, G. (2014). Modelling mammalian cellular quiescence. *Interface focus*, 4(3):20130074.

Zhang, X., Fang, Y., Jaiseng, W., Hu, L., Lu, Y., Ma, Y., and Furuyashiki, T. (2015). Characterization of Tamoxifen as an Antifungal Agent Using the Yeast *Schizosaccharomyces Pombe* Model Organism. Technical Report 2.

Zhao, L., Liu, Z., Levy, S. F., and Wu, S. (2018). Bartender: a fast and accurate clustering algorithm to count barcode reads. *Bioinformatics*, 34(5):739–747.

Zimmermann, A., Hofer, S., Pendl, T., Kainz, K., Madeo, F., and Carmona-Gutierrez, D. (2018). Yeast as a tool to identify anti-aging compounds. *FEMS Yeast Research*, 18(6).