The use of high-throughput microscopy in the characterisation of phenotypes associated with the Unfolded Protein Response in *Saccharomyces cerevisiae*

By

Peter William Bircham

A thesis
submitted to the Victoria University of Wellington
in fulfilment of the requirements for the degree of
Doctor of Philosophy
in Cell and Molecular Bioscience

Victoria University of Wellington
2014

ABSTRACT

Proteins traversing the secretory pathway begin their passage in the endoplasmic reticulum (ER) where they must be correctly folded and processed to pass quality control measures. Complications with this process can result in the accumulation of misfolded proteins, commonly referred to as ER-stress, which has been associated with a number of diseases. The unfolded protein response (UPR) is the cell's mechanism of dealing with ER-stress and is activated via the IRE1-HAC1 pathway in yeast. Ire1p is the ER-stress sensor and upon recognising misfolded proteins Ire1 oligomerises and forms local clusters. Activated Ire1p then splices out an inhibitory intron from the UPR specific transcription factor Hac1p which goes on to initiate downstream responses to alleviate ER-stress.

Here we utilise high-throughput microscopy and UPR-specific GFP reporter systems to characterise the UPR in the yeast *Saccharomyces cerevisiae*. High-throughput microscopy and automated image analysis is increasingly being used as a screening tool for investigating genome-wide collections of yeast strains, including the yeast deletion mutant array and the yeast GFP collection. We describe the use of GFP labelled Ire1p to visualise cluster formation as a reporter for early UPR recognition of misfolded proteins, as well as a GFP controlled by a Hac1p responsive promoter to measure downstream UPR activation. These UPR-specific GFP reporter systems were used to screen a collection of non-essential gene deletion strains, identifying gene deletions that induce UPR activation and thus are likely to function in the early secretory pathway. This included well known components such as the ALG members of the glycosylation pathway and various ER chaperones such as LHS1 and SCJ1. Additionally this analysis revealed 44 previously uncharacterised genes, suggesting there are still processes related to the secretory pathway that are yet to be described. Moreover, by inducing ER-stress in this screening system we revealed genes required for the normal activation of the UPR including ribosomal/translation and chromatin/transcriptionally related genes, as well as various genes from throughout the secretory pathway.

Furthermore, we screened a collection of ~4000 strains, each expressing a different GFP fusion protein, under ER-stress conditions to identify protein expression and localisation changes induced by the UPR. Comparison to UPR deficient Δhac1 cells uncovered a set of UPR specific targets including 26 novel UPR targets that had not been identified in

previous studies measuring changes at the transcript level. As part of this work, we developed a dual red fluorescent protein system to label cells for automated image segmentation to enable single cell phenotype measurements. Here we describe the use of texture analysis as a means of increasing automation in the identification of phenotypic changes across the proteome. These novel techniques may be more widely applied to screening GFP collections to increase automation of image analysis, particularly as manual annotation of phenotypic changes is a major bottleneck in high-throughput screening. The results presented here from microscopy based screening compare well with other techniques in the literature, but also provide new information highlighting the synergistic effects of integrating high-throughput imaging into traditional screening methodologies.

ACKNOWLEDGEMENTS

I would like to display my gratitude to all those who have helped me throughout the duration of my thesis work. Firstly, my supervisors Professor Paul Atkinson and Dr David Maass, thank you for giving me the opportunity to extend my postgraduate studies this far. Paul your vision and enthusiasm for science has been a driving inspiration. Thank you for giving me the freedom and encouraging me to operate independently and try new things as I saw fit. Without this encouragement which I would not have been able to breach into the field of computer science and image recognition, this has now grown into a personal passion. David your uncanny ability to understand all aspects of biology has been a continual source of amazement. Without your help and guidance I would not have developed the same level of practical skill nor critical thinking I now have. As supervisors you have both gone beyond what the job called for and become true mentors in my life, I have immensely enjoyed your company and conversations.

I would also like to extend my thanks to other faculty members who have been associated with our lab; Professor John Miller, Associate Professor Paul Teesdale-Spittle and Dr Andrew Munkacsi I have greatly appreciated all your thoughts and discussions throughout lab meetings and lunch time banter. Your advice as undoubtedly helped throughout my research and an extra viewpoint is always welcome.

To all of the Chemical Genetics lab group, you have always provided an enjoyable environment from which genuine friendships have developed. In particular I would like to thank Dr James Matthews for all of the philosophical conversations regarding science we have had in different states of sobriety – these have no doubt influenced my critical thinking. Bede I think we can now both agree that homebrew is indeed the best application for all the yeast skills we have picked up from the lab. Veronica, Chuckles, Natelle, Ploi, Namal, Christina, Seeseei, Yee, Leon and Dini – you have all played a unique role and I value all your friendships. Thank you Dini for also being geeky enough to help me out and talk stats.

Finally I would like to thank my family for all their support. Particularly my parents Terry and Jackie who have always encouraged me to follow my passions, without your support I would simply not have been able to take my studies this far. Katie you have provided immense support throughout this thesis, thank you. Your Kind personality has provided me the best of company. I think my family's opinion of you sums it up best as they now favor you over me. Finally we can take that break we have been talking about.

CONTENTS

Abstract 3
Acknowledgements 5
Contents 6
List of Figures and Tables 10
1 Introduction 1
  1.1 Protein Folding and the Secretory Pathway 2
  1.2 Translocation 3
  1.3 Protein maturation in the ER 5
    1.3.1 Disulphide-bond formation 5
    1.3.2 Glycosylation 7
  1.4 Protein misfolding and disease 9
  1.5 The Unfolded Protein Response (UPR) 11
  1.6 IRE1 (Inositol Requiring Enzyme 1) 12
    1.6.1 Structure 12
    1.6.2 Sensor mechanism 12
  1.7 HAC1 (Homologous to Atf/Creb1) 13
  1.8 Regulation of the UPR 14
  1.9 UPR target genes 15
  1.10 ER-assisted degradation (ERAD) 16
  1.11 *Saccharomyces cerevisiae* - a secretory pathway model 17
  1.12 Yeast gene deletion collections 18
  1.13 Yeast GFP collection 20
  1.14 High-throughput microscopy 21
  1.15 Research aims 23
2 MATERIALS AND METHODS 24
  2.1 Yeast strains 24
  2.2 Growth media 25
  2.3 Plasmids used 26
  2.4 Polymerase Chain Reaction (PCR) conditions 27
  2.5 DNA electrophoresis 29
  2.6 Yeast transformation procedure 29
  2.7 Transformation confirmation by colony PCR 30
  2.8 Random spore selection for strain construction 30

2.9 Construction of red fluorescent marker strains 31
2.10 SGA mass mating procedure 31
2.11 Mating type assessment 34
2.12 Serial spot dilution assay 35
2.13 Image acquisition / analysis 35
2.14 Gene ontology analysis 35
2.15 Transcription factor analysis 36
3 Automated image analysis 37
3.1 Background 38
3.1.1 Subcellular organisation of proteins 38
3.1.2 Fluorescent proteins as markers for live cell imaging 39
3.1.3 Location proteomics with the yeast GFP collection 41
3.1.4 Automated image analysis 43
3.2 Image preparation 44
3.3 Cell Detection without nuclei recognition for Reporter-SGAs 46
3.3.1 Cytoplasmic RFP strain construction 46
3.3.2 Cytoplasm based cell detection algorithm 46
3.3.3 Measuring circularity 48
3.4 Cell detection with nuclei detection 48
3.4.1 Dual RFP marker strain development 48
3.4.2 Nuclei detection 50
3.4.3 Cytoplasm identification 52
3.4.4 Confirmation of cell border recognition scripts 54
3.5 Feature measurements 56
3.5.1 GFP expression 56
3.5.2 Spot detection 56
3.5.3 Texture analysis 56
3.6 Discussion 58
4 screening the yeast deletion set for UPR genes 61
4.1 Background 61
4.1.1 Use of the DMA for genomic screening 61
4.1.2 UPR Reporter selection 62
4.1.3 Ire1p as a GFP reporter 63
4.1.4 Hac1p as a GFP reporter 64

4.1.5 UPRE as a reporter 64
4.2 Ire1p-GFP DMA screen 65
4.2.1 Ire1p-GFP reporter strain construction 65
4.2.2 Ire1p-GFP reporter validation 65
4.2.3 Identification of gene deletions that *induce* Ire1p cluster formation 67
4.2.4 Gene deletions that *prevent* Ire1p cluster formation under ER-stress 70
4.3 Hac1p-GFP as a potential reporter of UPR activation 71
4.4 UPRE reporter DMA screen 73
4.4.1 UPRE-ADE2 reporter construction 73
4.4.2 UPRE-ADE2 reporter testing 74
4.4.3 Construction of 4xUPRE-GFP strain 74
4.4.4 4xUPRE-GFP strain validation 75
4.4.5 Gene deletions that induce UPRE-GFP expression 76
4.4.6 Gene deletions that down-regulate UPRE activity 80
4.4.7 Gene deletions that suppress UPRE-GFP induction 81
4.5 Discussion 83
4.5.1 UPR inducers 86
4.5.2 UPR suppressors 87
4.6 Conclusion 91
5 localisation proteomics of er-stress 92
5.1 Background 92
5.2 Screening for proteome changes induced by the UPR 94
5.2.1 Selection of optimal drug concentrations for ER-stress conditions 94
5.2.2 Preparation and screening of the yeast GFP collection under UPR conditions 95
5.3 Abundance changes across the proteome 96
5.3.1 Measurement of protein abundance 96
5.3.2 Data normalisation and analysis: 98
5.3.3 Proteins up-regulated under ER-stress conditions 104
5.3.4 Proteins repressed under ER-stress conditions 108
5.4 Localisation changes across the proteome 111
5.5 Changes compared to UPR deficient GFP set 115
5.6 Transcription factor analysis 119
5.7 Discussion 121

5.7.1 Comparison of DTT and TM as ER-stress inducers 121

5.7.2 GFP proteomic screening reveals novel UPR target genes 122

5.7.3 Additional stress responses act alongside the UPR 127

5.7.4 ER-stressed repressed proteins are enriched for ribosomal proteins........ 129

5.8 Conclusion 130

6 Discussion 131

6.1 Automated recognition procedures 131

6.2 Characterising UPR phenotypes 135

6.2.1 Deletion array screens and GFP screens reveal different aspects of the UPR 135

6.2.2 GFP screening reveals protein-level UPR effects 136

6.2.3 Ire1p activation appears to be a committal event 136

6.2.4 Downstream UPR activity is blocked by SAGA complex components........ 137

6.3 Limitations and future directions 137

6.3.1 DMA screening 137

6.3.2 GFP set screening 139

Conclusion 141

References 142

7 Appendices 164

7.1 Nuclei recognition script 164

7.2 Cytoplasm recognition script from nuclei 165

7.3 Cytoplasm recognition script without nuclei 171

7.4 Combine stack script (mid-section selection) 175

7.5 Z-score script 179

7.6 Mann-Whitney U test script 179

7.7 Radii ratio script 181

7.8 DTT and TM induced localisation changes 182

7.9 Systematically removed GFP strains 203

LIST OF FIGURES AND TABLES

Figure 1 – Overview of the yeast secretory pathway. 3
Figure 2 - Metazoan UPR. 11
Figure 3 - Overview of the UPR. 13
Figure 4 – Comparison of pixel intensity gradients across Z-planes. 45
Figure 5 - cytoplasm detection without nuclei. 47
Figure 6 – Schematic of the dual RFP labelling used. 50
Figure 7 - Nuclei recognition procedure. 51
Figure 8 - cytoplasm recognition procedure. 54
Figure 9 - Percentage of cells recognised 55
Figure 10 - Visual representation of SER texture features on Bap2p-GFP cells with and without DTT treatment. 57
Figure 11 - Ire1p-GFP foci formation in ER-stress conditions. 66
Figure 12 - Ire1p cluster formation literature comparison 69
Figure 13 - Hac1p-GFP activation in ER-stress conditions. 72
Figure 14 - UPRE-ADE2 reporter 74
Figure 15 - 4xUPRE-GFP DTT dose response 76
Figure 16 - UPRE-GFP inducer literature comparison 79
Figure 17 - Comparison of UPRE suppressors and down regulators 83
Figure 18 –UPR activation as a dose-response to Dithiothreitol treatment 94
Figure 19 - UPR activation as a dose response to tunicamycin treatment 95
Figure 20 - Comparison between cell size and total GFP fluorescence after DTT treatment. 97
Figure 21 –Replicate filtering. 99
Figure 22 - boxplots showing control and treated GFP distributions across replicates before and after normalisation. 100
Figure 23 – Treated-Control intensity plots. 101
Figure 24 - Loess normalisation and local Z score calculation. 102
Figure 25 - Local Z-score for GFP expression changes. 103
Figure 26 - Overlap between DTT and TM up-regulated proteins 105
Figure 27 - Overlap between DTT and TM repressed proteins 108
Figure 28 - Example localisation changes. 113
Figure 29 - Overlap between ER-stress inducers and DNA damaging agents. 122
Figure 30 - overlap of up-regulated proteins between DTT, TM and UPR deficient cells 123
Figure 31 - UPR independent TF regulation. 128

Table 1 - Yeast strains 24
Table 2 - Plasmids used 27
Table 3 - List of primers used 28
Table 4 - outline of SGA mating selections 34
Table 5 - summary of various large-scale papers that have used the yeast GFP collection to screen the proteome 42
Table 6 - Gene deletions that induce Ire1p-GFP cluster formation 67
Table 7 - Gene deletions that prevent Ire1p clustering after DTT treatment.................... 70
Table 8 - Gene deletions that induce UPRE-GFP expression 77
Table 9 - Gene deletions that down-regulate UPRE activity 80
Table 10 - Gene deletions that suppress UPRE-GFP induction upon ER-stress 81
Table 11 - Proteins up-regulated by DTT treatment.. 105
Table 12 - Proteins up-regulated by TM treatment... 106
Table 13 - Proteins repressed by DTT treatment. 108
Table 14 - Proteins repressed by TM treatment. 109
Table 15 - Protein localisations altered by ER stress induced by TM or DTT treatment .. 114
Table 16 - Proteins up-regulated by DTT in the absence of a functional UPR (Δhac1). ... 115
Table 17 - Proteins repressed by DTT treatment in the absence of a functional UPR (Δhac1). 117
Table 18 - Localisation changes induced by ER-stress in UPR deficient cells.................. 118
Table 19 – UPR target Genes... 124
Table 20 - UPR localisation changes. 124

List of commonly used abbreviations

CSSR - Core stress-sensing region

DMA – Deletion mutant array

DTT – Dithiothreitol

EMC – ER membrane complex

ER – Endoplasmic reticulum

ERAD – ER assisted degradation

FDR – False discovery rate

GFP – Green fluorescent protein

GPI – Glycosylphosphatidylinositol

HAC1i – Induced HAC1 mRNA

HAC1u – Uninduced HAC1 mRNA

NLS – Nuclear localisation signal

PCR – Polymerase chain reaction

PDI – Protein disulphide isomerase

RFP – Red fluorescent protein

R-SGA – Reporter SGA

SER – Spots, edges and ridges

SGA – Synthetic genetic array

SSL – Synthetic sick or lethal

TF – Transcription factor

TM – Tunicamycin

UPR – Unfolded protein response

UPRE – Unfolded protein response element

1 INTRODUCTION

The central dogma of molecular biology describes the flow of information from DNA to RNA to protein. Indeed it is this controlled flow of information that permits the blueprints of life to be transmissible from one generation to the next, allowing life to prosper so well in all its variety. It is the protein products however, that give physical manifestation to this information and carry out the functions and complex interactions required to maintain living cells. The specialised functions of each protein are highly particular to its three dimensional structure and thus a high fidelity must be maintained in the flow of information from DNA through to protein. DNA replication is extremely accurate with error rates estimates on order of 1 in 10^{10}, even the processes of transcription and translation are generally considered to be accurate although to a lesser extent with error rate estimates around 1 in 10^4 (Cochella and Green, 2005; Hebert and Molinari, 2007; Ogle and Ramakrishnan, 2005; Rosenberger and Hilton, 1983). It is the later steps of protein production that are less accurate, in particular initial protein folding is rife with difficulties with estimates as high as 30% of nascent peptides terminally misfolding (Schubert et al., 2000). It is perhaps surprising that protein folding is inherently so erroneous given that precise protein folding is critical for correct protein functionality and cell viability. Failure of accurate protein folding can result in abnormal protein function or aggregation and is the cause of a number of debilitating and proteopathic diseases including Alzheimer's disease and Cystic fibrosis (Hebert and Molinari, 2007). As such, stringent quality control mechanisms are required to ensure protein folding fidelity.

This dissertation focuses on the early secretory pathway and protein folding within the endoplasmic reticulum. Close to a third of all proteins in the yeast *Saccharomyces cerevisiae* are ultimately destined to various endocytic compartments, plasma membrane or to be secreted, and must traverse the secretory pathway (Ghaemmaghami et al., 2003). This requires being translocated, processed and folded within the endoplasmic reticulum (ER), a factory-like organelle specialised in oxidative protein folding. It became apparent in early studies of the secretory pathway that there must be sophisticated quality control mechanisms in place to maintain protein fidelity and that a specific response capable of recognising unfolded proteins must exist (Kozutsumi et al., 1988; Needham and Brodsky, 2013). Since then quality-control in the ER has been a prevalent area of research.

1.1 Protein Folding and the Secretory Pathway

Most nascent peptides destined for the secretory pathway are synthesised at membrane bound ribosomes where they are co-translationally translocated into the ER lumen through the Sec61 translocon. This process is facilitated by use of signal recognition peptides present on the N-terminus of secretory proteins which can be recognised on the nascent protein chain, designating entry into the ER. Once peptides are in the ER they are modified and folded with assistance of a host of chaperones and folding enzymes to guide the process and prevent aggregates forming. These modifications not only assist folding into a stable functional protein but also determine the ultimate destination of the protein, whether for secretion, membrane insertion or organelles further along the secretory pathway.

Protein folding is the process in which a peptide is arranged into its native state, the confirmation with the lowest Gibbs free energy, that is able to function correctly in its biological context. The polypeptide itself can encode enough information to reach this state in optimal conditions, folding down its energy landscape until the lowest free energy is achieved. Indeed this has been achieved *in vitro* (Hebert and Molinari, 2007). However, this process alone takes far longer than acceptable in a native biological setting. The ER further optimises protein folding by decreasing exposure of hydrophobic residues to the outside of the cell assisted by chaperones, folding enzymes and disulphide bond formation (Hebert and Molinari, 2007). Cycling through states of glycosylation allows unfolded proteins time for multiple folding attempts while maintaining them within the ER without premature export as a misfolded protein (Aebi et al., 2010).

Nascent protein flux fluctuates across time and environmental conditions, as such there will be times where a cell cannot maintain protein folding fidelity with the demand for new protein. The result is an ER-stress condition characterised by the accumulation of unfolded and/or misfolded proteins, and is known to be a contributing factor to a number of pathologies (Yoshida, 2007). ER-stress must be swiftly dealt with in order to prevent toxic levels of misfolded protein accumulation and aggregation. A comprehensive set of cellular responses are in place to monitor and maintain the ER folding environment;

namely the Unfolded Protein Response (UPR) and ER-assisted degradation (ERAD). Figure 1 depicts a basic overview of the yeast secretory pathway, which is discussed in detail throughout this chapter.

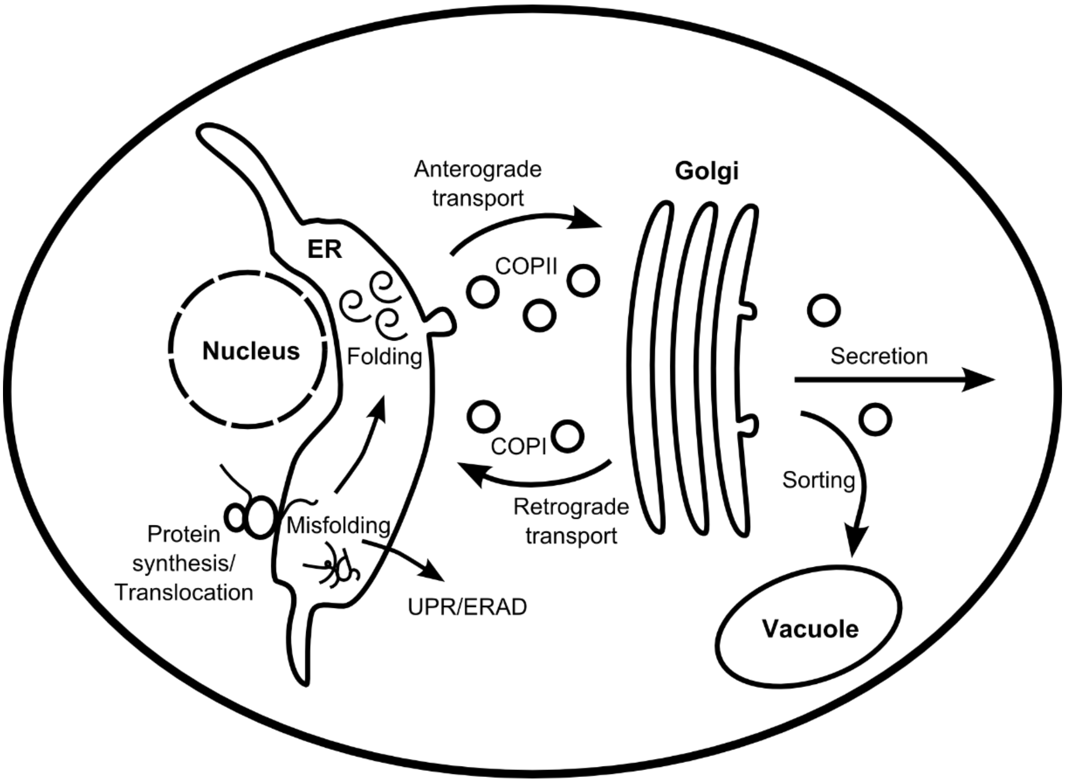


Figure 1 – Overview of the yeast secretory pathway.

Nascent proteins are translocated into the ER lumen either co- or post-translationally. Proteins undergo glycosylation and folding events assisted by chaperones and folding enzymes. Anterograde transport of folded proteins to the Golgi is via COPII vesicles, whereas retrograde transport is via COPI vesicles. Ultimately proteins may be delivered to the cell surface for secretion or sorted to destination organelles. Misfolded proteins initiate an Unfolded Protein Response and are targeted for ER-assisted degradation (ERAD) via the proteasome.

1.2 Translocation

S. cerevisiae has mechanisms for both co- and post-translational translocation of nascent secretory proteins into the ER, including soluble and integral membrane proteins (Delic et al., 2013). Secretory proteins are targeted to the ER either via a hydrophobic N-terminal signal sequence or transmembrane domain (Barlowe and Miller, 2013; Ng et al., 1996). In co-translational translocation as the signal sequence is translated it can be recognised and bound by the signal recognition particle (SRP) forming the ribosome nascent chain (RNC)-SRP complex, halting further translation of the peptide (Walter and Blobel, 1981). The SRP in turn targets the complex to the ER membrane bound SRP receptor (SR) in a

GTP dependant manner, directing and transferring the RNC to the translocon pore. GTP hydrolysis leads to separation of the SRP-SR from the RNC and peptide translation can continue (Keenan et al., 2001). There are two translocon pore complexes in *S. cerevisiae*, the Sec61 complex and its non-essential homologue the Ssh1 complex. Sec61p is a 53 kDa protein with ten trans-membrane spans that forms the major subunit of the heterotrimeric Sec61 complex, along with two smaller proteins Shb1p and Sss1p (Osborne et al., 2005). The Sec61 complex is part of the larger SEC' complex, made up of the Sec61 complex along with Sec63p, Sec71p and Sec72p which is required for co-translational translocation (Jermy et al., 2006). In addition to the SEC' complex the hexameric Ssh1 complex is capable of co-translational translocation comprising; Ssh1p, Sbh2p and Sss1, analogous to the Sec61 complex, along with Sec63p, Sec71p and Sec72p (Finke et al., 1996). Removal of the signal sequence occurs during translocation and is an essential part of protein maturation. The sequence is removed by the signal peptidase complex consisting of Spc1p, Sp2p, Spc3p and Sec11p (Böhni et al., 1988; YaDeau et al., 1991), of which Sec11p contains the active protease site (Mullins et al., 1995). The ER luminal Hsp70 chaperone Kar2p, along with co chaperones Lhs1p and Sil1p, is thought to provide the means of pulling the peptide through the translocon in a ratcheting mechanism (Matlack et al., 1999).

For SRP-independent post-translational translocation the nascent peptide must be protected from misfolding and aggregation upon release from the ribosome. This is achieved by the actions of the cytosolic Hsp70 chaperones Ssa1p-Ssa4p along with Hsp40 co-chaperone Ydj1p; which help maintain a loosely folded conformation suitable for translocation into the ER (Becker et al., 1996). The heptameric SEC complex is required for post-translational translocon and is made up from an association of the Sec61 complex along with the Sec63 complex (Sec62p, Sec63p, Sec71p and Sec72p) (Brodsky and Schekman, 1993; Panzner et al., 1995). The Sec63 complex is involved in recognition and initial binding of the signal peptide region facilitating transferral to the Sec61 translocon.

Different substrate specificity between co and post translation translocation differs between organisms. Simpler organisms such a yeast tend to have a greater dependence on post-translational translocation than mammalian cells, possibly due to their high growth rates preventing protein translocation keeping up with synthesis (Rapoport, 2007). However specific substrates for both co- and post-translational translocation have been identified in yeast and mammalian cells (Lakkaraju et al., 2012), indicating that requirement differences are more likely rather than volume of throughput alone. Highly hydrophobic signal peptides and integral membrane proteins tend to use co-translational translocation whereas soluble proteins tend to depend on post-translational mechanisms (Ng et al., 1996).

A third post-translational translocation mechanism exists for membrane proteins targeted to the ER known as the guided entry of tail-anchored proteins (GET) pathway (Schuldiner et al., 2005, 2008). This pathway operates independently of the Sec61 and Sec63 translocon complexes, and instead includes the proteins Get1-5 and Sgt2 (Battle et al., 2010; Costanzo et al., 2010; Jonikas et al., 2009; Schuldiner et al., 2008). A cytosolic GET pre-targeting complex consisting of Sgt2p, Get4p and Get5p loads nascent tail-anchored proteins onto the targeting factor complex (Get3p), which then delivers the protein to an ER-membrane receptor complex made up of Get1p and Get2p. Finally the Get1/2 complex facilitates insertion of the protein into the ER-membrane in an ATP dependant mechanism (Shao and Hegde, 2011).

1.3 Protein maturation in the ER

1.3.1 Disulphide-bond formation

Once proteins have translocated, the endoplasmic reticulum lumen provides an oxidative environment that encourages the formation of disulphide bonds intrinsic to protein folding and stabilising the correct protein confirmation. This process is especially important for proteins destined for the cell surface or to be secreted that will eventually be exposed to an oxidising environment as opposed to reducing conditions such as in the cytoplasm. The major pathway for disulphide bond formation in yeast is mediated by ER oxidoreductin (Ero1p) and protein disulphide isomerases (PDI) (Barlowe and Miller, 2013). Ero1p provides the major source of de novo disulphide bonds in *S. cerevisiae*, which are eventually shuttled onto protein folding substrates via PDIs in a series of dithiol-disulphide exchange reactions (Frand and Kaiser, 1998, 1999; Pollard et al., 1998).

There are five PDI proteins in *S. cerevisiae* only one of which, Pdi1p, is essential and constitutes the main PDI in yeast (Farquhar et al., 1991; Laboissière et al., 1995). PDIs contain thioredoxin-like domains that in their active oxidised state catalyse the direct formation and rearrangement of disulphide bonds within nascent peptides during protein folding. Inactive reduced PDI is regenerated by re-oxidation in a dithiol-disulphide exchange with Ero1p (Sevier et al., 2007). Ero1p is an ER membrane bound flavoprotein containing 14 cysteines with two catalytically active cysteine pairs; the active-site pair Cys352-Cys355, and the shuttle pair Cys100-Cys105 (Kim et al., 2012b). The shuttle cysteine pair directly oxidises PDI and is itself re-oxidised by internal dithiol-disulphide exchange with the active site cysteine pair. The active site is in turn re-oxidised by electron transfer to molecular oxygen as the preferred acceptor (Tu and Weissman, 2004) via the redox cofactor FAD in a process generating reactive oxygen species (ROS) (Bader et al., 1999; Tu and Weissman, 2002, 2004; Tu et al., 2000).

Ero1p catalytic activity is modulated through an additional two cysteine pairs Cys90-Cys349 and Cys150-Cys295 in a redox dependant manner (Sevier et al., 2007). When the regulatory cysteine pairs are in a reduced state Ero1p is catalytically active and vice versa. The sophistication of this system is in the maintenance of oxidative folding homeostasis by the feedback loop created as the regulatory cysteines compete with substrate cysteines for Ero1p activity. Upon increased demand for protein folding and disulphide formation the regulatory cysteines are exposed to reducing conditions and activate Ero1p to provide more oxidising equivalents. As demand is fulfilled the ER becomes increasingly oxidising shutting down Ero1p activity by autonomous inactivation in both cis and trans preventing hyper-oxidising conditions. PDI is consequently the main substrate and also main regulator of Ero1p activity.

In this respect it can be noted that under conditions of high protein turnover cells are often subject to significant levels of oxidative stress. Glutathione may play a role in the detoxification of the ER-generated ROS by using up reduced glutathione (GSH) generating the oxidised disulphide form (GSSG), this is likely to be a contributing factor to the relatively high levels of GSSG in the ER compared to the cytoplasm (Chakravarthi et al., 2006).

The ER redox status is also important for regulating the balance of PDI activity and largely determined by glutathione the major small molecule redox buffer in the ER (Kim et al., 2012b). In this respect it should be noted that Pdi1 but not Ero1 responds to the relative glutathione redox state (Delic et al., 2012) and in this manner Ero1p activity is regulated through the redox conditions of the ER through Pdi1 (Appenzeller-Herzog, 2011; Sevier et al., 2007). The ER ratio of reduced glutathione (GSH) to oxidised glutathione disulphide (GSSG) is much lower than found in the cytosol (~3:1 compared to ~100:1 respectively) hence the ER is less reducing (Hwang et al., 1992). This ratio is optimal for regulating a steady state activation of Ero1p and disulphide bond formation (Kim et al., 2012b; Lyles and Gilbert, 1991). Glutathione is not required for disulphide bond formation itself, but instead acts as a net reductant helping to maintain an optimal level of reduced PDI required for activation of the regulatory bonds in Ero1p (Cuozzo and Kaiser, 1999; Frand and Kaiser, 1998; Kim et al., 2012b). PDI reductase activity is also required to reshuffle misoxidised disulphide bonds or reduction of a terminally misfolded protein prior to ERAD (Fassio and Sitia, 2002). As such GSH may have a role in ensuring some of the ER oxidoreductases are maintained in a reduced state so that they can catalyse the rearrangement of non-native disulphides (Chakravarthi and Bulleid, 2004; Jessop and Bulleid, 2004; Molteni et al., 2004).

1.3.2 Glycosylation

Most proteins traversing the secretory pathway are glycosylated to some level. Glycosylation is essential for a number of protein functions and has importance for such processes as maintaining cell wall integrity and osmotolerance. Glycosylation is also generally considered to be important in protein folding dynamics. Addition of oligosaccharides increases solubility of a protein and provides a means of signalling to target proteins to the Golgi, retaining them in the ER for further folding or for targeting terminally misfolded proteins to the proteasome.

Protein glycosylation occurs parallel to translocation and signal sequence removal. As the nascent peptide emerges from the translocon it is scanned for consensus N-glycosylation sites (Asn-X-Ser/Thr). The enzyme oligosaccharylytransferase (OST) catalyses the addition of the preassembled oligosaccharide precursor $Glc_3Man_9GlcNAc_2$ from dolichyl pyrophosphate donor to a nitrogen on the asparagine residue. The OST complex is made of eight integral membrane proteins (Ost1p, Ost2p Ost3p or Ost6p, Ost4p, Ost5p, Wbp1, Swp1p and the catalytically active subunit Stt3p), whereas the assembly of the oligosaccharide precursor is carried out by the ALG family of enzymes. Although the first step of N-linked glycosylation occurs as the protein is translocated, signal sequence cleavage must occur before oligosaccharides can be attached to sites near the signal sequence (Chen et al., 2001). After the initial N-glycosylation the two terminal glucose residues are sequentially removed by glucosidase I and II resulting in $GlcMan_9GlcNAc_2$ which is specifically recognised and bound by the ER membrane bound lectin chaperone calnexin (Cne1). Calnexin disassociates from the protein when the remaining α1,3-linked glucose residue is removed by glucosidase II. Higher eukaryotes including mammalian cells have a quality control system involving a cycle of calnexin binding and disassociation. If the protein has not folded correctly UDP-glucose:glycoprotein glucosyltransferase (UGGT) adds another α1,3-linked glucose residue allowing re-binding of calnexin (Caramelo and Parodi, 2008). This process is thought to allow more time for proteins to fold properly, however unlike other yeast, *S. cerevisiae* has no functional homologue of UGGT (Fernández et al., 1994). After the final glucose removal, glycans can be acted on by ER mannosidase I (Mns1p) which trims a single mannose residue generating a specific isomer $Man_8GlcNAc_2$ (Aebi et al., 2010). Correctly folded proteins can then be directed to ER-exit sites for transport to the Golgi. Terminally misfolded proteins may undergo further mannosyl trimming by Mnl1p to generate $Man_7GlcNAc_2$ exposing an α1,6-linked mannose. These mannosyl trimmings are thought to be the first step in directing terminally misfolded protein to the ERAD pathway for degradation (discussed in more detail below).

O-linked glycosylation in *S. cerevisiae* involves the attachment of mannose to ser/thr residues by protein O-mannosyltransferases (PMT, Pmt1-7) using dolichol phosphate mannose as the mannosyl donor. The role of these additions is less well understood that that for N-linked glycans, however conflicting roles have been proposed for the O-mannosylation in labelling of misfolded proteins, both for protection from degradation or alternatively leading to ERAD elimination (Harty et al., 2001; Hirayama et al., 2008).

Glycosylphosphatidylinositol (GPI) anchoring is a type of C-terminal glycolipid addition to a protein, the lipid component of which ultimately provides a means for the protein to attach to the cell membrane. Between ten and twenty percent of proteins processed through the secretory pathway receive a GPI anchor addition (Orlean and Menon, 2007). The GPI core itself is preassembled in a complex multistep process involving greater than twenty gene products. Biosynthesis of the GPI core begins on the cytosolic side of the ER membrane where phosphatidylinositol receives an addition of glucosamine which is then deacetylated. The GPI intermediate is then flipped onto the luminal side of the ER where it receives additions of an acyl chain, four mannose residues and two ethanolamine phosphates (Pittet and Conzelmann, 2007). Substrate proteins are bought to the ER via traditional N-terminal signal peptide and upon completion of translocation into the ER is further recognised by a C-terminal GPI anchor signal sequence. The GPI core is transferred to the designated ω residue via the five membered transamidase complex (Gpi8p, Gaa1p, Gpi17p, Gpi16p, and Gab1p) and the GPI signal sequence is removed. Further processing includes lipid remodelling of diacylglyerol and the addition of a ceramide moiety in yeast (Sipos et al., 1994). Additionally extra sugar side chains can be added at later stages and often in yeast GPI-proteins ultimately end up cross-linked to the cell wall and associated with lipid rafts (Orlean and Menon, 2007). Interestingly GPI attachment is an essential process in yeast, yet there is only one known GPI containing protein, Rot1p, that is itself essential (Leidich et al., 1994; Machi et al., 2004). GPI proteins have various roles on the cell surface including maintenance of cell wall structure and cell morphology, enzymatic actions and cell adhesion (Pittet and Conzelmann, 2007).

1.4 Protein misfolding and disease

Protein folding within the ER is a complex process and there are numerous points at which it could go wrong. As mentioned above, up to a third of nascent peptides are initially misfolded (Schubert et al., 2000). Correct protein structure is absolutely critical to function, and indeed for protecting the protein from inappropriate interactions or aggregation. Unsurprisingly there are numerous human pathological disorders associated with protein folding defects (Hebert and Molinari, 2007; Yoshida, 2007).

The classic example is that of cystic fibrosis, caused by defects in the folding of the cystic fibrosis transmembrane conductance regulator (CFTR) preventing enough functional protein from being produced. CFTR is a notoriously poor folding protein, with estimates as low as 25% of nascent proteins folding into mature protein (Ward and Kopito, 1994). When combined with mutations that cause further folding defects this leads to pathological levels of protein misfolding, the most common of which is ΔF508, present in up to 90% of cystic fibrosis cases and results in retention of CFTR in the ER (Kopito, 1999). Encouragingly, CFTR is a good example of how targeting the ER and protein folding processes can lead to the development of potential therapeutics such as chemical chaperones to assist in folding dynamics (Brown et al., 1996; Roomans, 2003).

Several other protein folding diseases arise from the formation of toxic protein aggregates that commonly form due to exposed hydrophobic residues in misfolded proteins (Yoshida, 2007). This is problem in a number of neurodegenerative disorders including Parkinson's, Huntington's and Alzheimer's disease (Hebert and Molinari, 2007; Yoshida, 2007). In the case of Alzheimer's disease – one of the most common neurodegenerative disorders, disease pathology is associated with accumulation of amyloid-β plaques and Tau proteins (Ozcan and Tabas, 2012). Mutations in the presenilin genes PS1 and PS2 have been associated with familial Alzheimer's disease. This is an example where defects in ER machinery can lead to disease onset, in particular mutations in PS1 have been shown to down-regulate ER quality control responses leading to a build-up of the toxic aggregates (Katayama et al., 1999). As a final neurodegenerative example, Parkinson's disease is associated with aggregation of the protein α- synuclein, named lewy bodies (Cooper et al., 2006). Parkinson's disease has been associated with mutations in parkin, a gene encoding a ubiquitin ligase enzyme involved in the degradation of misfolded proteins (Shimura et al., 2000). This leads to the accumulation of misfolded parkin substrates in the ER and ultimately leads to ER-stress induced cell death (Imai et al., 2000).

It is clear that defects in ER-machinery and accumulation of misfolded proteins within the ER is an issue of major cellular concern. It is unsurprising then that cells have evolved a number of ER quality control mechanisms to deal with typical levels of protein misfolding. Of particular importance to this thesis is the UPR that acts as the sensing mechanism for misfolded proteins, and further initiates the response to cope with and eliminate the ER-stress.

1.5 The Unfolded Protein Response (UPR)

The UPR is a stress monitoring system for protein folding conditions within the ER lumen comprising an ER to nucleus signalling system. Activated in times of ER-stress, the UPR controls expression of a host of gene products to protect the cell from and deal with the accumulation of misfolded proteins. This highly conserved process is present across all eukaryotes, with a very highly conserved sensor protein, Inositol requiring enzyme 1 (IRE1), as well as two additional metazoan systems (Figure 2). Metazoans also use activating transcription factor 6 (ATF6) and Protein Kinase-RNA like ER kinase (PERK).

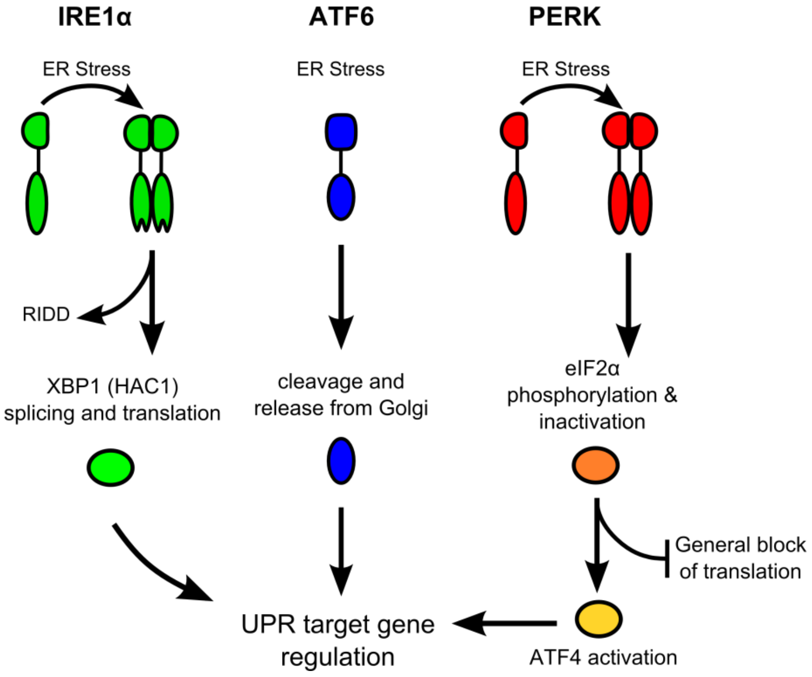


Figure 2 - Metazoan UPR.

Three separate signalling pathways make up the metazoan UPR. The IRE1 pathway is highly conserved between yeast and metazoans. Metazoan IRE1α responds to ER stress in a similar manner to yeast Ire1p and ultimately splices out an inhibitory intron from the UPR specific transcription factor XBP1 (homologous to yeast HAC1). Additionally the RNase function of IRE1α is also involved in the regulated Ire1-dependant decay (RIDD) of specific mRNAs. RIDD substrates include a number of secretory protein mRNAs and thus RIDD acts to reduce the protein folding load in the ER. PERK shows structural homology to IRE1 in the ER-stress sensing domain. Upon activation, PERK phosphorylates the eukaryotic translation initiation factor eIF2α resulting in a general block of translation also decreasing the protein folding load in the ER. Additionally, phosphorylation of eIF2α results in the preferential translation of the transcription factor ATF4 which in turn upregulates specific UPR target genes. Under conditions of chronic ER-stress PERK can ultimately lead to activation of pro-apoptotic genes including the transcription factor CHOP. The ATF6 branch of the UPR shows no structural similarity to IREα or PERK and senses ER stress through a different mechanism. Upon activation ATF6 is transported to the Golgi where the cytosolic domain is cleaved off and released allowing it to translocate to the nucleus and act as a transcription factor for specific UPR genes.

The UPR was first characterised in *Saccharomyces cerevisiae* and yeast has remained an invaluable tool to the field owing to genomic scale tools unavailable in metazoan systems (Ghaemmaghami et al., 2003; Giaever et al., 2002; Huh et al., 2003; Winzeler et al., 1999). The yeast *Saccharomyces cerevisiae* system has only the most conserved of the three sensing systems, the IRE1 - HAC1 signalling pathway as described below.

1.6 IRE1 (Inositol Requiring Enzyme 1)

1.6.1 Structure

Ire1p is a trans-ER-membrane serine/threonine kinase comprising a luminal ER-stress sensing domain and cytosolic RNase and kinase effector domains. In yeast the ER luminal domain has been proposed to consist of five subregions (Kimata et al., 2004). Subregions I and V are thought to be loosely folded while II-IV form a tightly folded core stress-sensing region (CSSR), also referred to as the core luminal domain cLD (Credle et al., 2005; Kimata et al., 2004). The CSSR forms two interfaces, I and II, that facilitate Ire1p oligomerisation; interface I allows dimerization of Ire1p monomers creating a major histocompatibility complex (MHC) like groove thought to be involved in direct binding of unfolded proteins, while interface II facilitates higher-order oligomerisation of Ire1p clusters (Credle et al., 2005; Gardner and Walter, 2011; Kimata et al., 2007). Subregion III sticks out as a flexible section of the CSSR and is essential for Ire1p activation. Subregion V is a binding site for the Hsp70 chaperone Kar2p (BiP in Mammalian cells) juxtaposed with the transmembrane domain (Kimata et al., 2004).

1.6.2 Sensor mechanism

The exact mechanism of unfolded protein sensing and Ire1p activation has been a subject of some controversy with arguments put forward for direct unfolded peptide binding via the MHC-like groove in Ire1p dimers or alternatively competitive binding of Kar2p with unfolded peptides. However recent compelling evidence would suggest it is likely a combination of these with direct peptide binding being the activation event and Kar2p having a modulating role for the response (Credle et al., 2005; Gardner and Walter, 2011; Kimata et al., 2004, 2007; Pincus et al., 2010). The likely sequence of events seems to be a two-step process initially involving the dissociation of Kar2p from subregion V, probably through competitive binding with misfolded proteins (Bertolotti et al., 2000; Kimata et al., 2003), and withdrawal of repression by subregion I. This is followed by oligomerisation, direct binding of unfolded peptides via the CSSR and activation of the cytosolic domains (Gardner and Walter, 2011; Kimata et al., 2007). Transduction of the UPR signal is facilitated by an Ire1p trans-autophosphorylation event, activating the cytosolic RNase effector domain and the unconventional splicing of its sole target in yeast HAC1 (Niwa et al., 2005; Sidrauski and Walter, 1997). Available evidence suggests activation of the endonuclease domain is unlikely to be a direct consequence of phosphorylation and more like to be due to a conformational change induced by the phosphorylation and clustering of Ire1p (Credle et al., 2005; Korennykh et al., 2009).

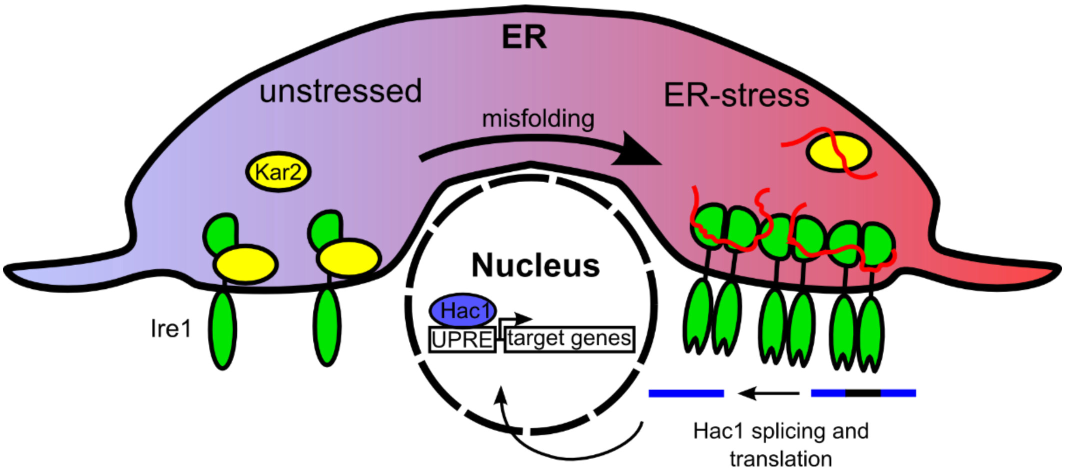


Figure 3 - Overview of the UPR.

In unstressed conditions Ire1p is associated with the chaperone Kar2p. Under ER-stress conditions unfolded proteins competitively bind to Kar2p causing a dissociation of Kar2p from Ire1p. Activated Ire1p forms homo-dimers through interface I forming a MHC-like groove capable of binding unfolded proteins. Ire1p undergoes high-order oligomerisation through interface II and a conformational shift in the cytosolic domain induces trans-autophosphorylation. This conformational change supports efficient splicing of an inhibitory intron from the UPR transcription factor HAC1. Hac1p protein can then translocate into the nucleus where it binds UPRE upstream activator sequences of various UPR target genes for transcriptional upregulation.

1.7 HAC1 (Homologous to Atf/Creb1)

HAC1 encodes a basic leucine zipper transcription factor with homology to Atf/Creb1 (Nojima et al., 1994). Hac1p upregulates a number of UPR genes upon activation of the UPR as described above. Facilitated by the tRNA ligase Trl1p, splicing of *HAC1* mRNA by Ire1p removes a 252 nucleotide inhibitory intron (Cox and Walter, 1996; Gonzalez et al., 1999; Sidrauski et al., 1996). The uninduced HAC1 mRNA (HAC1^u) is constitutively transcribed but not translated due to a stalling event at the ribosome (Rüegsegger et al., 2001). The induced HAC1^i is efficiently translated into the transcription activator Hac1p – initiating transcriptional induction of a broad array of UPR target genes (Kimata et al., 2006; Travers et al., 2000). Hac1p is translocated to the nucleus where it binds an upstream non-coding activating sequence, termed the UPR element (UPRE) that occurs in various target genes including itself. The UPRE was initially identified as a cis-acting 22 base pair element in the promoter region of the well characterised UPR target gene *KAR2* (Mori et al., 1992). This binding element was then described in numerous other UPR target gene promoters including genes for chaperones, ERAD components and phospholipid synthesis, although it is not ubiquitous across all target genes. Bioinformatic approaches further characterised an additional two UPR elements designated UPRE-2 and 3, although collectively all three elements are only present in less than half of all known UPR target genes (Patil et al., 2004).

1.8 Regulation of the UPR

Although the basic mechanism of IRE1-HAC1 UPR signalling appears to be simple at first glance, there is a surprising amount of sophistication in the fine-tuned regulation of the response. Firstly there is the level of gain control provided by the positive regulation of HAC1 which itself has a UPRE Hac1p binding domain in its promoter region (Ogawa and Mori, 2004) and secondly the Hsp70 chaperone Kar2p (BiP) provides another level of regulation. Kar2p has been shown to bind inactive monomeric Ire1p but not activated Ire1p dimers (Bertolotti et al., 2000; Okamura et al., 2000). Additionally mutations of the Kar2p binding region that prevent Kar2p binding do not constitutively activate the UPR, which still requires binding of misfolded protein (Oikawa et al., 2007; Pincus et al., 2010). Moreover mutations in KAR2 that disrupt the dissociation of Kar2p from Ire1p significantly lower the level of UPR activity as does overexpression of Kar2p (Kimata et al., 2003, 2004; Kohno et al., 1993). These data suggest that rather than initiating UPR activation, Kar2 modulates the level of response and deactivation dynamics, and may provide a buffer to physiological fluctuations in protein folding fidelity preventing premature UPR activation (Pincus et al., 2010).

Finally there is a level of control provided through formation of high-order Ire1p oligomers that form dense local clusters across the ER (Credle et al., 2005; Kimata et al., 2007). This is thought to promote efficient HAC1 splicing by providing dense clusters of Ire1p to which HAC mRNA is targeted to (Aragón et al., 2009). This high-order oligomerisation also induces a conformational shift in the cytosolic domain of Ire1p dimers, optimising the Ire1p trans-autophosphorylation event that significantly increases splicing activity (Korennykh et al., 2009). This cluster formation of Ire1p is associated with strong short-term activation of the UPR. In conditions of chronic ER-stress Ire1p no longer forms clusters but instead exists as still active Ire1p dimers providing another level of regulation for lower but sustained UPR activation (Ishiwata-Kimata et al., 2013a). This is important as Hac1p itself has a short half-life of around two minutes, thus it is continued Ire1p activity that must sustain the UPR in long-term ER-stress (Kawahara et al., 1997).

1.9 UPR target genes

A number of genome-wide approaches have been employed to define the set of UPR target genes. The gold standard for referring to UPR target genes is a set of cDNA microarray based studies by Travers et al. In these experiments Travers et al. used dithiothreitol and tunicamycin as ER-stress inducing agents to activate the UPR and measure mRNA regulation of potential UPR target genes (Travers et al., 2000). By using the UPR knockout strains, Δire1 and Δhac1, Travers et al. were able to eliminate non-UPR effects and identified a set of 381 potential UPR-specific target genes. It is noteworthy that two previously known UPR target genes, KAR2 and INO2 did not pass their statistical selection criteria. What is particularly interesting about this is that KAR2 was the gene from which the UPRE Hac1p binding element was originally identified – suggesting that this may not represent the full scope of UPR targets, a point this thesis examines in detail. The Travers study identified a coordination between the UPR and ER-assisted protein degradation, as well as a range of UPR target gene functions including translocation, glycosylation, protein folding, vesicle transport, cell wall biogenesis and lipid metabolism.

Kimata et al. followed up with a similar cDNA microarray approach, however they made use of a constitutively active HAC1^{i} mutant lacking the inhibitory intron to define UPR target genes. This study identified additional anti-oxidative stress genes as UPR targets, suggesting that the UPR also responds to the ROS produced in protein folding reactions (Kimata et al., 2006). Furthermore Kimata and colleagues identified a set of genes down-regulated by the UPR, mainly cell wall proteins.

1.10 ER-assisted degradation (ERAD)

The UPR regulates a number of ER quality control (ERQC) measures to ensure protein folding fidelity is maintained. A fundamental aspect of protein folding is the ability to allow correctly folded proteins to be separated out and transported to their final destination, while still retaining actively folding proteins giving them time to fold. Importantly the system differentiates terminally misfolded proteins that must be prevented from accidental export and must be targeted for degradation. One such mechanism of degradation is the highly conserved process of ER assisted protein degradation (ERAD), which relies on involvement of chaperones and glycosylation processing markers for recognition of misfolded substrates before exporting them into the cytosol for degradation by the 26S proteasome (Xie and Ng, 2010).

ERAD systems are present across all eukaryotes and while metazoan cells have a more complex system with an unknown number of pathways, yeast have three well described pathways. ERAD-C for proteins with a misfolded cytosolic domain, ERAD-L for misfolded soluble proteins or those with a misfolded domain in the ER lumen and ERAD-M for a misfolded transmembrane domain (Carvalho et al., 2006). There are two specialised E3 ubiquitin ligases in yeast which are central to directing ERAD substrates to the proteasome; Doa10p for ERAD-C and Hrd1p for ERAD-L and M. There is only a small degree of homology between Doa10p and Hrd1p although their complexes share some of the same members including the E2 ubiquitin conjugating enzyme Ubc7p attached via Cue1p, as well as the Cdc48 complex (Cdc48p, Ufd1p and Npl4p) attached via Ubx2p. The Hrd1 E3 complex is, however, more intricate and also includes Der1p attached via Usa1p, both of which are required for ERAD-L. Also Hrd3p and Yos9p are both required and are involved in substrate recognition. Associated with the Hrd1 E3 complex is Sec61p, which is possibly used in translocation of ERAD-L substrates (Xie and Ng, 2010).

The glycan detection system in ERAD-L is by far the best characterised system of substrate recognition. After N-linked glycosylation, release from calnexin and mannose trimming by Msn1p (described above in 1.3.2) a terminally misfolded protein can be designated to ERAD-L rather than exported to the Golgi. The mannosidase-like protein Mnl1p is able to trim another mannose generating $Man_7GlcNAc_2$ (Clerc et al., 2009), this is specifically recognised by the lectin-like protein Yos9p and along with Hrd3p binding the unfolded structure of the protein, is sufficient to target the protein to the proteasome (Gauss et al., 2006; Xie et al., 2009). The misfolded protein is then translocated back to the cytosol likely through a Sec61p pore via the Cdc48 complex, ubiquitinated by Ubc7p and Hrd1p and finally presented to the 26S proteasome (Xie and Ng, 2010). The ERAD-M pathway differs from ERAD-L in that recognition is independent of Hrd3p/Yos9p, instead Hrd1p itself is able to directly recognise substrates with misfolded transmembrane domains (Sato et al., 2009). Finally, the ERAD-C pathway likely works in a manner similar to ERAD-M, but instead utilising Doa10p as the ubiquitin ligase (Carvalho et al., 2006). Recognition of the misfolded domain in yeast makes use of the cytosolic Ssa-family of Hsp70 chaperones as well as their Hsp40 co-chaperones Ydj1p and Hlj1p (Huyer et al., 2004).

1.11 *Saccharomyces cerevisiae* - a secretory pathway model

There is a great deal of history behind the usage of the model yeast *Saccharomyces cerevisiae*. It is probably the oldest of domesticated organisms and has for thousands of years been exploited for its fundamental role in fermentation for brewing and leavening of bread. For this reason it is colloquially known as brewers or baker's yeast. It is perhaps no surprise then that yeast has long been the canonical model organism for a plethora of studies on eukaryote cell biology including the workings of the secretory pathway (Schekman Nobel address 2013).

Yeast is unique in its scope of traits that make it so suited to use as a model organism. Like bacteria, yeast has the benefit of a rapid growth rate, ease of culturing conditions and is a unicellular organism. Yeast is genetically very malleable, with very efficient homologous recombination allowing for precise genetic modifications including gene deletion or insertion. Importantly yeast are eukaryotic organisms and have a remarkably high level of homology to higher eukaryotes including humans, with estimates of around 30% of yeast genes having human homologues (Botstein et al., 1997; Gilbert, 2002) and an order of 1000 human disease related genes (Botstein and Fink, 2011; Heinicke et al., 2007). It has become increasingly well documented that inter-species homology generally holds true at a gene function level as well (Dolinski and Botstein, 2007). These factors have led to the commonplace use of yeast in studies that lay the groundwork for our understanding of complex biological phenomena. In fact early landmark studies defining the genes involved in protein secretion and organising them into a pathway were carried out by Schekman and colleagues using yeast as a model system, and would simply not have otherwise been feasible in metazoan systems (Novick et al., 1980, 1981). Classical genetic screens utilising random mutagenesis, such as those of Schekman and colleagues, involve isolation of a huge amount of colonies to be tested in order to comprehensively cover the genome and require lengthy complementation and cloning assays before individual genes can be identified. However recent advances in genomics have paved the way for more precise and systematic methodologies for genomic screening.

1.12 Yeast gene deletion collections

The benefits as a model organism as well as having a small genome size, ~12 million bp, led to *S. cerevisiae* being the first eukaryote organism to have its complete genome sequenced and published (Goffeau et al., 1996) identifying approximately 6000 genes. The manageable size of this genome enabled development of a number of genomic tools in yeast that have opened the road for development of systems biology and functional genomics, fields in which yeast still drives forefront technologies. There are a number of yeast clone collections that comprise a set of mutant strains that cover the entire genome; prominent amongst these are the yeast knockout collections. These collections comprise a set of precise knockout mutants for each gene, systematically generated by PCR mediated transformation and replacement with a kanamycin resistance gene and a uniquely identifiable 'molecular barcode' DNA sequence (Giaever et al., 2002; Winzeler et al., 1999). Construction of the Deletion Mutant Array (DMA) revealed quite surprisingly that roughly 80% of yeast genes were not individually required for growth on rich medium and dubbed ‘non-essential’ genes. Although these non-essential gene knockout strains show no viability defect on rich medium, their function often becomes apparent when they become conditionally required upon exposure to various stress conditions (Giaever et al., 2002; Hillenmeyer et al., 2008).

The non-essential deletion mutant array has been extensively used for synthetic genetic array (SGA) analysis, a technique in which a query gene knockout strain is mass mated against the DMA in order to comprehensively map any epistatic interactions indicative of functional relationships between genes (Costanzo et al., 2010; Tong et al., 2001, 2004). SGA analysis measures colony size as a means to infer synthetic sick or lethal (SSL) interactions, a form of negative genetic epistasis whereby the combined effect of two individually non-essential gene knockouts results in either a reduced growth ‘sick’ or lethal phenotype. In addition to SSL interactions, positive genetic interactions can be measured in the case where each gene deletion may individually show a growth defect that is non-additive in the double mutant. The generalised interpretation of this is that negative genetic interactions tend to imply a buffering relationship between two genes potentially involved in two related pathways for some shared essential downstream function, while a positive genetic interaction likely implies two genes lie within the same pathway or protein complex (Boone et al., 2007). A good example showcasing the power of this approach is the work by Schuldiner and colleagues, where they mapped both positive and negative genetic interactions across a large array of secretory pathway genes from which they identified the GET complex, functioning in Golgi to ER trafficking and tail-anchored protein translocation to the ER (Costanzo et al., 2010; Schuldiner et al., 2005, 2008). The GET pathway was further expanded by applying the same epistasis principles to phenotypes other than growth measurements, namely UPR activation in knockout mutant strains (Jonikas et al., 2009). This is also a good example of the DMA being used in what has become to be known as a reporter-SGA (R-SGA) in which a reporter system for a phenotype of interest is mass mated to the DMA in an SGA-like methodology. This approach allows for measurement of phenotypic changes as a consequence of non-essential gene knockouts in the DMA that otherwise show no impact on cell viability (Burston et al., 2009; Fillingham et al., 2009; Kainth et al., 2009; Proszynski et al., 2005; Vizeacoumar et al., 2010; Wolinski et al., 2009). Previously we have used this approach to identify non-essential gene knockouts that cause a defect in the secretory pathway processing of a GFP tagged variant of the protein Mrh1p, a multi trans-membrane spanning protein of the plasma membrane. By using high-content microscopy we were able to screen for a build-up of intracellular Mrh1p-GFP levels in an automated high-throughput fashion. This identified 24 gene knockouts causing a defect in Mrh1p-GFP processing including all six members of the ER-membrane complex (EMC), all of which showed EMC-unique localisation in an ER-like retention pattern (Bircham et al., 2011). The EMC had only recently been identified as a complex in a complementary UPR reporter based screen of the DMA, where the authors suggested a probable function in protein folding (Jonikas et al., 2009), although taken in the context of our work a role in protein trafficking seems likely.

1.13 Yeast GFP collection

The versatility of yeast has facilitated the development of further genome-wide strain collections including protein over-expression and two-hybrid and tandem affinity purification systems (Ghaemmaghami et al., 2003; Sopko et al., 2006; Uetz et al., 2000). Of particular importance for the studies presented here is the yeast green fluorescent protein (GFP) tagged collection of 4,156 strains comprising ~75% of the yeast proteome (Huh et al., 2003). In this collection each strain expresses a different GFP fusion protein tagged at the carboxyl terminal with the *Aequorea Victoria* GFP variant S65T (Heim et al., 1995). Expression remains driven by the endogenous gene promoter and accordingly fluorescence can be measured proportionally to gene expression (Newman et al., 2006; Soboleski et al., 2005). Huh et al used this collection to classify proteins into 22 localisation patterns and provided localisation data for 70% of previously unclassified proteins. The Huh et al. study identified >300 ER localised proteins almost a third of which had not previously been described.

It is important to keep in mind the possible effects that adding a GFP tag could have on protein functionality. One consideration is that protein stability or turnover rates could be affected. However there is no known folding problems associated with GFP and overall there is no evidence of any effect on protein interactions with the proteasome that constitutes the principal means of protein turnover (Newman et al., 2006). Another consideration is the possibility of protein mislocalisation due to the C-terminal tag. This is likely to be the case where proteins contain a C-terminal signal such as in tail-anchored and GPI-anchored proteins or those with a HDEL ER retention signal, such as the ER chaperone Kar2p, which was annotated with an ambiguous localisation in the GFP dataset. It should be noted however that only a small portion of proteins, around 4% of the GFP collection, are likely to mislocalise due to C-terminal tagging (Breker et al., 2013) and for the most part there is a high correspondence, around 80% agreement, between GFP localisations and previously annotated localisations (Huh et al., 2003). Additionally 87% of essential genes in the collection were successfully tagged indicating that in general gene function is likely to be maintained in the presence of a C-terminal GFP tag.

1.14 High-throughput microscopy

Over the past decade advances in technology have progressed development of high-throughput imaging platforms while maintaining a high level of image quality. These systems are often referred to as high-content screening (HCS) due to the plethora of information that can be extracted from images on a single cell basis, in comparison to traditional growth assays (Li et al., 2011). Because of the ease of growth conditions of yeast and their small cell size the yeast GFP collection pairs perfectly with HCS enabling live cell imaging of subcellular features across the proteome (Chong et al., 2012; Vizeacoumar et al., 2009). Combined with the mass mating procedures available in yeast this system is open to rapid modifications such as introduction of additional fluorescent cell markers for colocalisation or cell identification. Deletion mutants can be used to visualise non-growth phenotypes, or alternatively individual strains provide a rich source for GFP reporters in R-SGAs.

There are now a number of commercial HCS imaging platforms available including the Perkin Elmer Opera™ system used here. The Opera system is a high-throughput spinning disk confocal microscope, fitted with a 60X water immersion lens and a dual camera setup with 488 nm and 561 nm excitation lasers for imaging GFP and RFP simultaneously. A motorised plate holder enables automated image acquisition across 384 well microtitre plates with automatic focusing achieved via an infrared laser. This system is capable of imaging the entire yeast proteome in a day. Screening the GFP set with both control and treated cells along with RFP cell markers, two fields of view and three Z-stacks will generate almost 100,000 images per replicate. The enormous quantity of images produced has led to a demand for automated image analysis software. Manual scoring of images by eye is not suitable at this level of throughput, and has the added complication caused by subjective rather than quantitative measurements, variation between individuals and possible bias.

Because of this many of the commercial imaging platforms also come with image analysis software. Acapella™ is the Perkin Elmer offering and has been used for the studies presented in this thesis. Most of these commercial software packages offer out-of-the-box drag and drop functionality for building basic scripts with the ability to customise for purpose built features. There are also a number of open-source software packages that are increasing in popularity such as ImageJ, CellProfiler and Cell-ID (Abramoff et al., 2004; Carpenter et al., 2006; Gordon et al., 2007). Recently there have been a number of advances in image recognition procedures for yeast screening systems, particularly in the recognition of single cells for protein abundance measurements (Breker et al., 2013; Dénervaud et al., 2013; Handfield et al., 2013; Mazumder et al., 2013; Nadler-Holly et al., 2012; Tkach et al., 2012). More sophisticated algorithms for specific measurements of reporter proteins show the versatility of image recognition procedures. Examples are the detection of spindle morphologies (Vizeacoumar et al., 2010), plasma membrane retention (Bircham et al., 2011), nuclear-cytoplasm translocations (Mazumder et al., 2013) and even proteome-wide localisation assessment (Dénervaud et al., 2013; Handfield et al., 2013). The application of high-throughput microscopy and automated image analysis for yeast screening is discussed in more detail in later chapters.

1.15 Research aims

The overall aim of this dissertation was to investigate the proteome-wide changes initiated by the UPR under conditions of ER-stress in the yeast *Saccharomyces cerevisiae*. For this we chose to utilise high-throughput live cell microscopy of genome-wide yeast collections, paired with automated image analysis. The work was split into three specific goals:

1. Firstly, to develop a system for the automated recognition and analysis of yeast cells. Specifically to design a reliable labelling system for cells to be accurately identified in confocal images, and furthermore to develop a method of automating identification of localisation changes across the proteome.

2. Secondly, to apply the image recognition procedures for the analysis of specific reporters of UPR activity as a means to systematically assess the consequences of single gene deletions on UPR activity. This will be assessed under both normal growth conditions and ER-stress conditions.

3. Finally, to apply the image recognition procedures in the systematic identification of UPR-specific localisation and abundance changes throughout the yeast proteome in response to ER-stress.

2 MATERIALS AND METHODS

2.1 Yeast strains

All yeast strains used in this study are derived from S288c and are maintained as glycerol stocks stored at -80°C. Strains are displayed in Table 1 below.

Table 1 - Yeast strains

Strain	Name	Description	Genotype
Y7092		SGA starting strain from Boone Lab	*MATalpha*; can1Δ::STE2pr-Sp_his5; lyp1Δ; his3Δ1 leu2Δ0 ura3Δ0 met15Δ0 LYS2+
Y9230		SGA starting strain from Boone Lab	*MATalpha*; can1Δ::STE2pr_URA3 lyp1Δ his3Δ1 leu2Δ0 ura3Δ0 met15Δ0 LYS2+
yCG307	BY4742 cenLEU2	BY4742 transformed with pRS315 (cenLEU2) plasmid	*MATalpha*; his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0
yCG215	NLS-RedStar2-nat	BPSV40 NLS fused to the N-terminus of DsRed2 with the Nat MX cassette at the 3' end intergrated into the ura3 locus	*MAT* apha; can1Δ::STE2pr_URA3 lypΔ1 leu2Δ0 his3Δ1 leu2Δ0 met15Δ0 ura3Δ0::NLS-DsRed2-NAT LYS2+
yCG251	NLS-RedStar2-HPH	yCG215 marker switch to HPH	*MAT* apha; can1Δ::STE2pr-URA3; lyp1Δ; his3Δ1; leu2Δ0; ura3Δ0; met15Δ0:
yCG253	NLS-RedStar2 mCherry	yCG251 transformed with mCherry; NATMX4-TEFpr-mCherry	*MATalpha*; can1Δ::STE2pr-Sp_URA; lyp1Δ::mCherry-Nat; his3Δ1; leu2Δ0; ura3Δ0::NLS-RedStar2-HPH; LYS2+;
yCG261	UPRE-GFP	4xUPRE-GFP integrated into met17 locus of Y7092	*MATalpha*; can1Δ::STE2pr-Sp_his5; lyp1Δ; his3Δ1 leu2Δ0 ura3Δ0 met15Δ::4xUPRE-GFP-URA3 LYS2+
yCG262	mCherry	TEF2pr_mCherry-NAT integrated into lyp1 locus of Y7092	*MATalpha*; can1Δ::STE2pr-Sp_his5; lyp1Δ::TEF2pr_mCherry-NAT; his3Δ1 leu2Δ0 ura3Δ0 met15Δ0 LYS2+

yCG266	UPRE-GFP mCherry	yCG261 transformed with mCherry into lyp1 locus	*MATalpha*; can1Δ::STE2pr-Sp_his5; his3Δ1 leu2Δ0 ura3Δ0 met15Δ::4xUPRE-GFP-URA3 lyp1Δ::TEF2pr_mCherry-NAT LYS2+
yCG458	Ire1-GFP NLS-RedStar2 mCherry	ire1-gfp from the NLS/mCherry GFPset mated against YCG307 (BY4742 cen::leu) and selected for by random spore	*MATalpha*; IRE1-GFP_HIS3, can1Δ::STE2pr-Sp_URA; lyp1Δ::mCherry-Nat; his3Δ1; leu2Δ0; ura3Δ0::NLS-RedStar2-HPH; LYS2+;

2.2 Growth media

All cultures were grown using the following media and standard growth procedures detailed below (Amberg et al., 2005). Yeast strains were incubated at 30°C whereas bacteria strains were grown at 37°C. All media was autoclaved at 121°C for 20 min and allowed to cool to 65°C before addition of glucose from a 40% (w/v) stock solution, as well as any necessary drug or antibiotics. Broth media was made following the same recipes with the omission of agar.

Yeast-extract peptone dextrose (YPD) Media

Yeast extract	1% (w/v)
Bacto-peptone	2% (w/v)
Adenine	0.012% (w/v)
Glucose	2% (w/v)
Agar	2% (w/v)

Synthetic Complete (SC) or Synthetic Dropout (SD) Media

Yeast nitrogen base (without amino acids or ammonium sulphate)	0.17% (w/v)
Monosodium glutamate	0.1% (w/v)
Amino acid mixture to suit*	0.2% (w/v)
Glucose	2% (w/v)
Agar	2% (w/v)

* synthetic complete amino mixture is made up as follows:
3 g adenine, 2 g uracil, 2 g inositol, 0.2 g para-aminobenzoic acid, 2 g alanine, 2 g arginine, 2 g asparagine, 2 g aspartic acid, 2 g cysteine, 2 g glutamic acid, 2 g glutamine, 2 g glycine, 2 g histidine, 2 g isoleucine, 10 g leucine, 2 g lysine, 2 g methionine, 2 g phenylalanine, 2 g proline, 2 g serine, 2 g threonine, 2 g tyrosine, 2 g tryptophan, and 2 g valine.

Dropout variants follow the mixture above with the omission of any desired amino acid(s)

Glucose Nutrient Agar (GNA) Pre-Sporulation Media

Yeast extract	0.8% (w/v)
Bacto-peptone	0.3% (w/v)
Glucose	10% (w/v)
Agar	2% (w/v)

Enriched Sporulation Media

Potassium acetate	1% (w/v)
Yeast extract	0.1% (w/v)
Amino acid supplement (histidine, leucine, lysine, uracil)	0.01% (w/v)
Glucose	0.05% (w/v)
Agar	2% (w/v)

Luria-Bertani (LB) Media

Bacto-tryptone	1% (w/v)
Yeast extract	0.5% (w/v)
Sodium chloride	1% (w/v)
Agar	2% (w/v)

Antibiotic supplement	**Stock solution**	**Working concentration**
Nourseothricin (ClonNat, Werner BioAgents)	100 mg/mL	100 µg/mL
Geneticin (G418, Invitrogen)	200 mg/mL	200 µg/mL
Canavanine	50 mg/mL	50 µg/mL
Thialysine	50 mg/mL	50 µg/mL
Hygromycin B (HPH, InvivoGen)	100 mg/mL	200 µg/mL
Ampicillin	100 mg/mL	100 µg/mL

2.3 Plasmids used

All plasmids used in this study were maintained in bacterial cultures frozen at -80°C as glycerol stocks using standard procedures (Amberg et al., 2005). Competent DH5α *E. coli* cells (Invitrogen) were transformed with plasmid DNA following the manufacturer's instructions with minor alterations. Briefly, 25 µL cell aliquots were thawed on ice before addition of 1-10 ng plasmid DNA, followed by further incubation on ice for 30 min. Cells were then heat shocked at 42°C for 20 s and placed back on ice for a further 2 min. After addition of 475 µL of pre-warmed LB broth, cells were incubated at 37°C for 1 hr with shaking at 225 rpm. Aliquots of 20, 100 and 200 µL were then plated onto pre-warmed LB plates containing ampicillin and were incubated overnight at 37°C. Single colonies were picked and grown in 3 mL LB broth containing ampicillin overnight at 37°C. Finally, cells were harvested to be frozen down as 15% glycerol stocks at -80°C. Plasmid DNA for use in PCR amplification reactions was isolated from bacteria cultures and purified using the Zippy™ plasmid miniprep kit (Zymo Research) following the manufacturer's instructions.

Table 2 - Plasmids used

Plasmid name	Description	Source
pYM-N19	natNT2 TEF2pr	Janke et al, 2004
pYM43	Redstar2 natNT2	Janke et al, 2004
pFA6a-HPH	HPHNT1	Janke et al, 2004
pYM-N22	KanMX4	Janke et al, 2004
pMJ002	natMX4 TEF2pr_mCHerry	David Breslow, UCSF
P4339	natMX4	Tong et al, 2001
pPM47	URA3 4XUPRE-RFP	(Addgene #20132, Merksamer et al 2008

2.4 Polymerase Chain Reaction (PCR) conditions

PCR products for use in yeast transformations were all amplified using HotStar™ taq DNA polymerase (Qiagen). All primers were resuspended in ddH_2O to a concentration of 100 pM and stored at -20°C. Reaction mixtures were made up to 25 µL volumes following the manufacturer's guidelines as follows:

ddH_2O	19.6 µL
10X buffer	2.5 µL
5 mM dNTPS	1.25 µL
HotStar Taq	0.15 µL
Template DNA	0.5 µL
5' primer	0.5 µL
3' primer	0.5 µL

We used robust PCR cycling conditions that work well with difficult PCR reactions (Janke et al., 2004). This was found to work well in reactions using the long primers (>70 bp) required for efficient homologous recombination during transformation. The conditions used for this general PCR cycle are as follows:

Hot start activation :	95°C	15 min
3 step cycling x 10 repeats:		
Denaturation	95°C	1 min
Annealing	54°C	30 s
Extension	72°C	2 min 40 s
3 step extension cycling x 20 repeats:		
Denaturation	95°C	1 min
Annealing	54°C	30 s
Extension	72°C	2 min 40 s + 20 s / cycle

PCR products were checked for correct product size by electrophoresis on a 1% agarose gel with comparison to a 1 kb+ ladder. Confirmed PCR products were purified using a Geneaid PCR clean-up kit as per manufacturer's instructions.

Table 3 - List of primers used

Primer number	Description	Sequence
144	5' promoter-NLS	TGCGAGGCATATTTATGGTGAAGGATAAGTTTTGACCATCAAAGAAG GTTCGTACGCTGCAGGTCGAC
145	3' promoter-NLS	CTTTCTCTTTTTCTTTGGAGATTCAAATTCAGAACCATCAGCAGTTCTTT TACCACCAGTCATAGAAGCCATGTCCGGGGGGGATCCACTAG
146	5' NLS-RFP	GAATTTGAATCTCCAAAGAAAAAGAGAAAGGTTGAAGCTTCTGGTTT GGTTCCTAGAGGTTCTGCTTCTTCTGAAGATGTCATC
147	3' NLS-RFP	CCATGAAGCTTTTTCTTTCCAATTTTTTTTTTTTTCGTCATTATAGAAATC CGCTGGCCGGGTGACCCGGCGGGGAC
148	5' Cyc promoter	ATGATACATTTCTTACGTCATGATTGATTATTACAGCTATGCTGACGTA CGCTGCAGGTCGAC
149	3' cyc promoter	ACGTCCCAATTGTCCCCCTCCTAATATACCAACTGTTCTAGAATCCATC GATGAATTCTCTGTCG
176	UPRE Kan MX	GGCCATCCACGCTATATATACACGCCTGGCGGATCTGCTCTTTCGACA CGCTGTCCAGTTCCGTTTTCGACACTGGATGGCGGC
240	HPH marker switch	CGAGAAAATCTGGAAGAGTAAAAAAGGAGTAGAAACATTTTGAAGCT ATGAGCTCCGAGCTCGTTAAAGCCTTCGAG
224	mCherry fw	AATTGCATTGTCTATAACGATAACAAAAGACATCGTATATATATATAT ATCGTACGCTGCAGGTCGAC
252	mCherry rv	TCTATTTTTTTATTTTTTTCTATTTTGAAGGCATGCAAGAGGTTCTGTGA ACTATAGGGAGACCGGCAGA
257	MET15_URA3_fw	TCGAATCCCTTAGCTCTCATTATTTTTTGCTTTTTCTCTTGAGGTCACAT TCAGCGGGTGTTGGCGGGT
258	MET15_plasmid_rv	AAACTTTGTTGAATGTTGAGCAAGTTAACATCTTATAGGACATATTAA ACGTGAATGTAAGCGTGACATA

259	UPRE-GFP overlap	AACACCAGTGAATAATTCTTCACCTTTAGACATTATTAATTTAGTGTGTGTAT
260	GFP-MET15	AAACTTTGTTGAATGTTGAGCAAGTTAACATCTTATAGGACATATTAAACTTATTTGTACAATTCATCCATACC
261	UPRE overlap-GFP	AACACAAATACACACACTAAATTAATAATGTCTAAAGGTGAAGAATTATTCA

2.5 DNA electrophoresis

Electrophoresis was performed using 1% agarose gels made with TBE buffer (89 mM Boric Acid, 2 mM EDTA disodium dihydrate, 89 mM Tris Base [pH 8]). Prior to casting, ethidium bromide was added to the gels for DNA staining to a concentration of 0.5 µg/ml. Gels were run in the same TBE buffer and loaded with samples of 5 µL of PCR product and 2 µL loading dye (glycerol 30% (v/v), bromophenol blue 0.25% (w/v)) beside 2uL 1 kb+ DNA ladder (Invitrogen) in loading dye for size comparison. Electrophoresis was run at 120 V until the loading dye had migrated a satisfactory distance. DNA bands were visualised and images captured using an Alpha Imager mini transilluminator (Alpha Innotech) at 365 nm.

2.6 Yeast transformation procedure

Yeast transformations were carried out following a high efficiency lithium acetate/single-stranded carrier DNA/PEG method with minor modification (Gietz and Schiestl, 2007). In brief, 50 mL of YPD broth was inoculated with yeast cells picked from a single colony and grown overnight at 30°C with shaking at 225 rpm. Once the culture had reached an OD in the range of 1 – 2, cells were harvested by centrifugation at 3,000 g for 5 min and washed twice by repeated steps of resuspension in sterile H_2O and centrifugation. Cells were then resuspended to a concentration of 10^9 cells/mL of which 100 µL samples were placed into micro-centrifuge tubes for each transformation. Cells were then centrifuged at 13,000 g for 30s and the supernatant removed. Cells were resuspended in a transformation mix of the following components:

50% (w/v) PEG 3350	240 µL
1M Lithium acetate	36 µL
2 µ/mL Denatured salmon sperm DNA	50 µL
PCR product	10 µL
ddH_2O	24 µL

Cells were then heat shocked at 42°C for 40 min, centrifuged at 13,000 g for 30 s and the supernatant removed. Cells were resuspended in 3 mL YPD and incubated at 30°C in a 15 mL Falcon tube with rotation for 3 hrs. Finally aliquots of 2, 20 and 200 µL were plated onto selective media and incubated at 30°C for two days. Transformed cells were re-streaked again onto selective media and single colonies were isolated to be stored in 15% glycerol stocks at -80°C.

2.7 Transformation confirmation by colony PCR

Colony PCR was used to confirm that the transformant yeast cells had indeed integrated the desired product into the correct locus. PCR primers were designed within the internal antibiotic cassette that had been introduced and ~200 bp upstream or downstream of the target ORF or integration site. Confirmation PCR product amplification should then only be possible if the transformation had occurred within the correct locus. PCR products were amplified and visualised using standard procedures as above.

2.8 Random spore selection for strain construction

Random spore selection was used to create strains of a desired mating type and traits from the genetic background of the two haploid parental strains. By using strains with the genotype background of the SGA starting strains ([can1Δ::STE2pr-Sp_his5; lyp1Δ] or [can1Δ::STE2pr_URA3; lyp1Δ]), haploid progeny can be selected in the same manner as the SGA mass mating procedure described in detail below. The desired haploid parental strains of opposite mating type were mated on YPD plates by streaking cells harvested from a single colony toward each other in a chevron pattern. Cells were mixed at the point of the chevron to ensure efficient mating and the plates were incubated overnight at 30°C. Diploids were selected by replica plating onto media selective for traits from both parental strains and incubated overnight at 30°C. Diploid cells were replica plated onto GNA pre-sporulation media and incubated for no more than 16 hours at 30°C. Cells were then harvested from a patch ~1 cm^2 and used to inoculate 1.5 mL of enriched sporulation broth in a 15 mL Falcon tube. These cells were left to sporulate for a minimum of 5 days at 20°C on a rotator. Following this, 200 µL of the sporulation culture was centrifuged at 13,000 g for 30 s and the cell pellet resuspended in 100 µL of 2.5% (w/v) zymolyase 20T. This was then incubated for 30 min at 30°C before vortexing for 1 min to release spores from the ascus. Samples were diluted 100, 1000 and 10,000 fold in water and plated onto selective media to isolate single colonies. A series of replica plating onto selective media allowed for genotypic selection and colonies containing the desired traits were re-streaked and frozen down in 15% (w/v) glycerol stocks at -80°C.

2.9 Construction of red fluorescent marker strains

A dual red fluorescent protein marker strain was created to visualise the cytoplasm and nucleus for the efficient automated identification of a cell in image analysis. A bipartite nuclear localisation signal from SV40 (Hodel et al., 2006, 2001) was attached to the red fluorescent protein Redstar2(Bevis and Glick, 2002; Janke et al., 2004) with expression controlled by the TEF2 promoter (Janke et al., 2004). The construct was created by PCR amplification of two products with homologous flanking regions to the yeast URA3 locus and an overlapping region between them. The TEF2 promoter was amplified from plasmid pYM-N19 with PCR primers 144 and 145, while the RedStar2 RFP along with the cloneNAT antibiotic resistance marker was amplified from plasmid pYM43 with primers 146 and 147. An overlapping region containing the NLS signal was introduced from primers 145 and 146 allowing for concurrent transformation of both products into the genome of the yeast strain Y9230 resulting in the strain yCG215. This strain was then marker switched from cloneNAT to hygromycin B (HPH) by transformation with the PCR product amplified from plasmid pFA6a-HPH with primers 144 and 240 resulting in the strain yCG251. This strain was further transformed with the PCR product amplified from plasmid pJM002 (A kind gift from David Breslow, University of California, San Francisco) with primers 224 and 252. This introduced the cytoplasmic RFP mCherry under control of the TEF2 promoter into the lyp1 locus, resulting in the strain yCG253.

2.10 SGA mass mating procedure

Mass mating of query strains to the yeast libraries was achieved using the Singer ROTOR HDA (Singer Instruments, Somerset, UK) colony arraying robot. This allowed for the rapid introduction of a query deletion or reporter into these libraries. Query strains were derived from the SGA starting strains Y7092 and Y9230 that contain a *MAT***a** specific selection system (can1Δ::STE2pr-Sp_his5 or can1Δ::STE2pr_URA3 respectively) for haploid selection. Mating procedures were carried out using standard protocol (Tong and Boone, 2007). The general procedure is summarized below with specific strains and their selection media indicated in Table 4:

1. **Yeast Library**
    a. **DMA**

       The DMA was a kind gift from Charles Boone (University of Toronto) and consists of ~4,500 *MAT***a** non-essential deletion strains arrayed across 14 plates. Plates were arrayed with 384 strains per plate with a *Δhis3* control strain along the borders. The DMA collection is maintained on YPD + G418 media.

    b. **GFP Collection**

       The GFP collection was obtained from Invitrogen and consists of 4159 *MAT***a** strains each with a different GFP-fusion protein arrayed across 11-384 colony plates. The GFP collection is maintained on SD-His media.

2. **Mating of the Query strain and Yeast libraries**

   Query strains for reporter SGAs were derived from the appropriate SGA starting strains (see row two, Table 4). A 10 mL overnight culture of the *MATα* query strain was poured into an empty singer plate as a liquid source to pin onto agar plates in 384 format to match the density of the yeast library being used. Plates were incubated overnight at 30°C.

   These initial query plates were then pinned onto fresh YPD plates and the *MAT***a** yeast library pinned directly on top to allow mating to occur. The mating plates were incubated at 30°C overnight.

3. **Diploid selection**

   *MAT***a**/*α* diploids were selected by pinning onto YPD media containing the appropriate selections as seen in row 4 of Table 4. This double selection media allowed for conjoint selection of the desired genetic markers from both parent strains (query & library array), thus eliminating unmated haploids. Diploid selection plates were then incubated for two days at 30°C.

4. **Sporulation**

   Diploid cells were then pinned onto enriched sporulation media as described in section 1.2. These plates were incubated for 7 days at 20°C. These plates are deficient in nitrogen which encourages diploid cells to undergo meiosis and produce haploid spores.

5. ***MAT*a selection**

   After sporulation cells were pinned onto *MAT***a** specific selection media (see row 5 Table 4) and plates were incubated for two days at 30°C. This *MAT***a** specific selection works on the basis of the *STE2* promoter (specific to haploid *MAT***a** cells) which drives the expression of a selectable marker (Histidine or Uracil biosynthesis). Additional haploid selection markers are present in these cells as they carry gene deletions of the arginine and lysine permeases *CAN1* and *LYP1*. When grown on media containing toxic analogues of arginine and lysine (Canavanine and Thialysine) any diploid cells still carrying the functional wild type permeases will be selected against as they take up these toxic drugs. For SGAs using Y7092 derived strains the selection media was SD-His/Arg/Lys + CAN/THIA or for Y9230 derived strains SD-Ura/Arg/Lys + CAN/THIA. Plates were then incubated for two days at 30°C.

   A second *MAT***a** selection was repeated following this and incubated overnight at 30°C.

6. **Allele selection**

   Following *MAT***a** selection, a series of pinning procedures were carried out maintaining the *MAT***a** genotype and selecting one more allele each time until all desired alleles had been accounted for (See row 6 Table 4). For each selection step cells were allowed to grow for two days at 30°C.

Table 4 - outline of SGA mating selections

Library created	UPRE-GFP DMA	Ire1p-GFp DMA	GFP collection with RFP markers
Query strain	yCG266 (from Y7092)	yCG458 (from Y9230)	yCG253 (from Y9230)
Yeast library	DMA	DMA	GFP collection
Diploid selection	YPD + NAT/G418	YPD + NAT/G418	SD – His + NAT
MATa selection	SD – His/Arg/Lys + CAN/THIA	SD – Ura/Arg/Lys + CAN/THIA	SD – Ura/Arg/Lys + CAN/THIA
Final selection	SD – His/Arg/Lys + CAN/THIA/NAT/G418	SD – URA/His/Arg/Lys + CAN/THIA/NAT/HPH	SD – URA/His/Arg/Lys + CAN/THIA/NAT/HPH

2.11 Mating type assessment

To verify reliability of the R-SGA selection procedures, colonies were selected for PCR-based mating type testing using a combination of three primers as described in Huxley et al. (1990). One primer corresponded to the *MAT* locus while the other two are mating type specific primers that produce a different product size for each mating type; a 544 bp product for *MAT***a** and a 404 bp product for *MATα* or both products in the case of a diploid. The protocol in brief is as follows: strains of interest were streaked on agar plates and grown for two days at 30°C. Single colonies were harvested and suspended in 50 µL of 1 mg/mL zymolyase 20T for 30 min at 30°C to extract template DNA from cells. PCR reactions of 25 µL were setup with the reaction mix listed below:

ddH$_2$O	14.6 µL
10X buffer	2.5 µL
5 mM dNTPS	1.25 µL
HotStar Taq	0.15 µL
Template DNA	5 µL
MAT locus primer	0.5 µL
*MAT***a** specific primer	0.5 µL
MATα specific primer	0.5 µL

A thermal cycle consisting of a 15 min 95°C hot start followed by 30 cycles of 95°C for 1 min denaturation, 58°C for 2 min annealing and 72°C for 2 min extension. PCR products were analysed by DNA electrophoresis as described above (section 2.5).

2.12 Serial spot dilution assay

Cells were harvested from single colonies grown on YPD agar plates and used to inoculate a 3 mL YPD culture. Cultures were grown overnight until an OD_{600} of 1 (~$1x10^7$ cell/mL) was reached from which serial 10-fold dilutions were set up in a 96 well microtitre plate. Five µL spots were pipetted onto agar drug plates using an 8-channel pipette giving a range from 50,000 cells per spot down to single cells. Plates were grown for 24 or 48 hours and scored for growth sensitivity.

2.13 Image acquisition / analysis

Cells were grown on agar plates in 384 colony format overnight at 30°C to obtain fresh cells. These cells were transferred into 384 well clear bottomed microtitre plates (Perkin Elmer cell carrier) containing 50 µL of media using the Singer RoToR HAD colony arraying robot. Cells were incubated for four hours at 30°C and placed in the microscope 5 minutes prior to imaging to allow cells to settle to the bottom. Cells were imaged using the Perkin Elmer Opera high-throughput confocal microscope using the following setup: 60X water immersion lens $NA = 1.2$. GFP and RFP were simultaneously imaged by splitting light through a 568 nm detection dicroic mirror to two peltier cooled 1.3 megapixel CCD cameras. GFP was excited using a 488 nm laser and captured through a 520/35 bandpass filter. RFP was excited using a 561 nm laser and captured through a 600/40 nm bandpass filter. Images were exposed for 400 ms and a binning of 2 was used for screening conditions.

2.14 Gene ontology analysis

Gene ontology analysis was conducted using the web-based software tools on YeastMine (Balakrishnan et al., 2012. http://yeastmine.yeastgenome.org). Gene ontology enrichments were calculated for gene hit lists by comparison to custom background population sets that only represented the strains present in the yeast collection used. Benjamini–Hochberg (False discovery rate) was used as a test correction with p-value cut-offs as described in the text. Redundant gene ontology terms were ignored after comparison using amiGO (Carbon et al., 2009. http://amigo.geneontology.org).

2.15 Transcription factor analysis

Transcription factor analysis was conducted using a comprehensive web-based database of all known yeast transcription factor (TF) interactions in YEASTRACT (Abdulrehman et al., 2011; Monteiro et al., 2008; Teixeira et al., 2006, 2013). TF interactions were identified using the online search tool ‘rank by TF’ available on the YEASTRACT website. The gene hit list of expression changes being analysed was used as input for the target ORFs. Potential TFs were either limited to the same set of input genes to search for interactions within the hits, or alternatively all known TFs were used to search for interactions outside of the hit list. A p-value of 0.005 as calculated by YEASTRACT was used as a significance cutoff for TF interactions unless otherwise stated.

3 AUTOMATED IMAGE ANALYSIS

Commercial availability of automated microscopy screening platforms – such as the PerkinElmer Opera microscope used here - provide an efficient means to capture thousands of images that has enabled measures of proteome-wide responses. It is commonly said a picture is worth a thousand words, although the challenge now appears to be extracting information from the wealth of data we are able to capture. Whilst the Opera microscope – like most commercial platforms – provides its own software for phenotypic analysis, we found that the existing cell recognition procedures were better suited to larger mammalian cells than the *Saccharomyces cerevisiae* cells used here. One of the major goals of this thesis was therefore to develop an automated system for analysing phenotypic data from proteome-wide yeast screens. Developing a standard platform that could be broadly applied, but also customized for each experiment would enable the identification of a range of possible phenotypes under different growth conditions. This chapter describes the innovative development of a consistent RFP labelling system and custom yeast-optimised image recognition software scripts for cell identification. This was developed in the Acapella programming language that comes with the Opera microscope using custom written as well as built in procedures. Because we chose to image live cells that were freely suspended in growth media, one of the first challenges that we had to overcome was to automate selection of optimal focal planes for image analysis. The aims of this chapter were thus:

1. To define a procedure for automated selection of images containing optimal yeast mid-sections.
2. To develop a cytoplasm based cell recognition procedure, suitable for reporter-SGA screens with specific reporter proteins.
3. To develop a high fidelity nucleus-cytoplasm based cell recognition procedure, suitable for the analysis of the wide range of highly variable GFP proteins in the yeast GFP collection.
4. Design these algorithms to be easy to use and highly customisable as a screening platform for future experiments.

3.1 Background

3.1.1 Subcellular organisation of proteins

In order to function appropriately, proteins not only require accurate synthesis and folding but must also localise to the correct subcellular compartment or organelle. The subcellular arrangement of proteins within a cell is highly structured but also highly dynamic. At any given time a cell expresses thousands of different proteins; all of which must be spatiotemporally managed to maintain normal biological function. Organelles can provide the environmental context in which a protein functions, and can vary in conditions such as pH, redox state and ionic concentration. Furthermore, localisation of a protein can ensure that interacting proteins can efficiently find each other, for instance the assembly of protein complexes, or conversely to prevent inappropriate interactions from taking place. In this sense, localisation can control the post-translational machinery and microenvironment conditions necessary for protein function, or control the physical interaction partners a protein is exposed to. Evidencing the essential role of location is research showing that the controlled mislocalisation of proteins can negate protein function analogous to loss-of-function mutants (Geda et al., 2008). Indeed abnormal localisation of proteins has been implicated in a number of wide-ranging human diseases including several cancers and neurological disorders (Hung and Link, 2011).

Living cells must respond dynamically to changes in growth conditions and adapt their intracellular components correspondingly to suit. Control of protein localisation is one means of regulating protein function in response to environmental conditions. Of note is Ire1p, a regulator of the UPR, which splices HAC1 mRNA more efficiently when it oligomerises in the presence of unfolded proteins, and changes from a dispersed ER protein to form localised foci. This dynamic change is used as a means of regulating levels of UPR activity (Kimata et al., 2007). Another instance of regulation is the localisation of transcription factors which are often held in an inactive state in the cytoplasm. When required, they are rapidly translocated into the nucleus where they exert their actions. An example of this is the response of the transcription factor Crz1p. Upon cellular stress conditions, cytoplasmic Crzp1 is dephosphorylated by calcineurin resulting in translocation into the nucleus (Stathopoulos-Gerontides et al., 1999). Another example is the ribosomal RNA processing element-binding protein, Stb3p, which in glucose starved quiescent cells is maintained in the nucleus, suppressing numerous growth related genes. Upon glucose repletion, Stb3p is translocated into the cytoplasm, permitting the expression of genes required for cell growth (Liko et al., 2010). This approach, where mature proteins are held in check until needed, can offer a much faster response than if a protein must be synthesized first.

Increasingly, there are reports of proteins 'moonlighting' or having multiple unrelated functions (Jeffery, 1999). Often these functions occur at differing locations within the cell. An example is Hxk2p which is the predominant hexokinase used for glucose metabolism in the cytoplasm in *S. cerevisiae*. Additionally, in cells growing on glucose, a proportion of Hxk1p localises to the nucleus where it plays a secondary role of regulating the glucose-dependent repression of various genes including invertase and galactose metabolic genes (Moreno and Herrero, 2002; Randez-Gil et al., 1998). It has been noted that these additional functions may often be overlooked or overcautiously disregarded as possible artefacts; however there is increasing evidence that it is likely to be a common phenomenon (Butler and Overall, 2009; Copley, 2012). Unexpected changes in localisation have been proposed as a means to identify potential moonlighting proteins (Gancedo and Flores, 2008). In fact, there is evidence for a significant number of well-described cytoplasmic proteins potentially 'moonlighting' with secondary functions on the cell surface (Nombela et al., 2006).

3.1.2 Fluorescent proteins as markers for live cell imaging

With the importance of the interplay between protein localisation and function in mind, it is apparent that live cell imaging of localisation dynamics can give insight into the function of uncharacterised proteins, or provide additional functions to expand existing annotations. Green Fluorescent Protein (GFP) has widely been used for its versatility as a means to label proteins for live cell imaging. Originally isolated from the jellyfish *Aequorea victoria*, GFP was first cloned in 1992 (Prasher et al., 1992) and subsequently demonstrated as a useful *in vivo* fluorescent tag (Chalfie et al., 1994). The open reading frame encoding GFP can be integrated genomically in yeast for either N- or C-terminal tagging of endogenous proteins. This provides a number of advantages over other labelling approaches. Firstly as GFP is expressed by the cells themselves there is no need for multi-step staining procedures that may be impractical in a high-content screening process. Additionally, GFP has very low toxicity related side-effects making it ideal for live-cell imaging. Because of these factors; GFP has been adopted as a preferred means of labelling proteins in live cells for both visual inspection and as a reporter of gene expression (Tsien, 1998). Various other coloured fluorescent proteins have since been identified, both naturally occurring such as the red fluorescent protein dsRed isolated from the coral *Discosoma sp*. (Matz et al., 1999), as well as mutational variants created to cover a broad spectrum of colours (Chudakov et al., 2010). There are a number of mutation enhancements that can be made to improve the performance of fluorescent protein tags *in vivo*. An enhanced GFP variant was used in the creation of the yeast GFP collection (Huh et al., 2003). This is one of the most widely used GFP variants and contains Ser65→Thr and Phe64→Leu mutations that dramatically increase maturation rate and brightness. Additionally, these mutations shift the excitation and emission wavelengths to a more useful 488 nm and 510 nm respectively, which is suited to commonly used fluorescein isothiocyanate (FITC) filter sets (Cormack et al., 1996; Heim et al., 1995). Pairing GFP with RFP is a common approach for dual labelling systems due to their well separated excitation and emission spectra. However, wild-type RFP proteins such as dsRed are often unsuitable for protein tagging due to their low brightness and multimeric nature that can lead to protein aggregation and toxicity. A large number of mutational enhancements were therefore required to create a fast maturing and monomeric version of dsRed (Bevis and Glick, 2002; Campbell et al., 2002). This has since been further enhanced to create a number of dsRed variants with much enhanced brightness and photostability including the two used here, RedStar2 and mCherry (Janke et al., 2004; Shaner et al., 2004). The RFP mCherry is a highly optimised variant with an especially good maturation rate and is also a true monomer, which are good attributes for a protein tag. RedStar2 has a combination of mutations that make it ~6-10 fold brighter than mCherry, and a fast maturing RFP variant (Janke et al., 2004). However Redstar2 is also a dimer and hence has the potential to cause localisation artefacts (Kaufmann, 2009). Nevertheless in the context of our studies this has not been a problem.

3.1.3 Location proteomics with the yeast GFP collection

Systems biology is concerned with the understanding of cell biology on an all-inclusive level. Sequencing of the yeast genome has seen the development of a number of genome scale tools that have led to an escalating amount of powerful genome-wide screens that produce an abundance of data. However a wealth of data does not always correspond to a wealth of knowledge. Bioinformatics is becoming increasingly important to interpret datasets, especially to comprehend some of the often, rather abstract relationships behind data such as the synthetic lethal interactions of an SGA. Over the past decade or so there has been an increase in excitement regarding high-content microscopy for genomic screens – a technology that potentially offers more direct insight into protein function. As discussed above, the location or re-localisation of a protein has great implications for function. Thus by directly visualising proteins and their phenotypes on a subcellular level, we are able to gain insights into function and/or their involvement in biological processes of interest. 'Location proteomics' has been termed to describe high-content imaging to measure protein localisation as means to screen for protein function on a genome-wide level (Chen et al., 2003). Current microscopy technologies have enabled high-throughput image acquisition of live cells and the availability of increasingly powerful computers for automated image analysis has laid the grounds for microscopic analysis of the entire proteome. Forefront in these studies is the use of the yeast GFP collection, the most comprehensive fluorescent protein library covering ~70% of the yeast proteome. A number of studies using this collection are listed in Table 5. Initial construction of the yeast GFP library allowed for manual annotation of protein localisations into 22 localisation classifiers, although a significant portion of these were classed as ambiguous (Huh et al., 2003). The GFP collection may be used to screen for protein changes in localisation and/or abundance in response to various conditions, the combination of which has been used very successfully to study proteome changes in response to both drug treatment and mutations (Breker et al., 2013; Nadler-Holly et al., 2012; Tkach et al., 2012).

Table 5 - summary of various large-scale papers that have used the yeast GFP collection to screen the proteome

Publication	Conditions	Live cells	Strains	Analysis	Notes
(Breker et al., 2013)	Stress conditions (DTT, H_2O_2, nitrogen starvation)	Yes	5,330	Automated cell recognition and abundance analysis. Manual inspection of localisation	Wide-field 60X air. Cytoplasmic RFP to identify cell borders.
(Handfield et al., 2013)	N/A	Yes	4,004	Automated cell recognition, protein localisation and cell cycle stage	Confocal 60X water. Cell cycle analysis associated proteins with specific stages of growth
(Herzig et al., 2012)	Mutations in secretory cargo receptors	Yes	~150 Cargo proteins	Manual inspection of localisation	Paired mutations in cargo receptors with GFP-tagged cargo proteins to match receptors and substrates
(Nadler-Holly et al., 2012)	Mutations in subunits of the cytosolic chaperonin Cct ring complex	Yes	~5,100	Automated cell recognition and abundance analysis. Manual inspection of localisation	Wide-field 60X air. Cytoplasmic RFP to identify cell borders.
(Narayanaswamy et al., 2009a)	Alpha factor mating pheromone	No	~4,200	Automated cell recognition. Manual and machine learning methods to identify proteins localising to the mating projection	Fixed cells printed onto slide micro-arrays
(Narayanaswamy et al., 2009b)	Nutrient depletion	No	~800 Cytoplasmic proteins	Manual inspection of foci formation	Fixed cells printed onto slide micro-arrays
(Newman et al., 2006)	Rich and minimal media	Yes	>2,500	Flow cytometry to monitor protein abundance	
(Noree et al., 2010)	Filament formation in cytoplasmic proteins	Yes	1,632 Cytoplasmic proteins	Manual inspection of filament formation	
(Shin et al., 2009)	Rapamycin	Yes	4159	Manual inspection of localisation	Wide-field 100X oil
(Tkach et al., 2012)	DNA damage (MMS, HU)	Yes	4,148	Automated cell recognition and abundance analysis. Manual inspection of localisation	Confocal 60X water. Nup49p-RFP to identify nuclei, estimated cytoplasm region rather than identify cell borders
(Dénervaud et al., 2013)	DNA damage (MMS)	Yes	4,085	Automated cell recognition and abundance analysis. Automated and manual inspection of localisation	Chemostat array system to maintain cultures while imaging. Used texture analysis for localisation changes

(Mazumder et al., 2013)	DNA damage (MMS)	No	>4000	Automated cell recognition, abundance analysis and nuclear/cytoplasm expression ratios	Used DAPI and Alexa 647 conjugated Concanavalin A to label cells for image detection

3.1.4 Automated image analysis

Since the construction of the Yeast GFP collection there has been a push toward automated analysis of protein localisation and abundance at the single cell level. Automated identification of single cells in micrographs has been a successful technique in a number of applications and usually requires fluorescent labelling of cell components such as nuclei and cytoplasm (Bircham et al., 2011; Breker et al., 2013; Nadler-Holly et al., 2012; Narayanaswamy et al., 2009a; Tkach et al., 2012). When screening for a specific type of localisation change, as in the case of Reporter-SGA screening, efficient recognition procedures can be created to measure differences. We have previously had success measuring internal accumulation of a GFP-tagged membrane protein in response to gene deletions by measuring the ratio of internal to cell surface fluorescence (Bircham et al., 2011), Others have also had good success with specific reporters measuring spindle morphology (Vizeacoumar et al., 2010). However in the context of screening the entire GFP collection, a large number of organelles must be considered and every protein has the potential to behave differently.

Most image recognition applications in the context of the GFP collection have been to measure protein abundance on a single cell level. However, there has been significant progress made toward automated localisation analysis as well (Chen et al., 2007; Handfield et al., 2013; Huh et al., 2009). One approach is to classify localisations based on a large number of parameters including morphological features (e.g. cell size and shape), as well as descriptors such as Haralick texture features, Zernike moment features or wavelet features (Murphy et al., 2003). Classification approaches can use a supervised approach where control images for each pre-determined localisation are used to train the algorithm (Chen et al., 2007; Huh et al., 2009; Narayanaswamy et al., 2009a), or alternatively a less biased, unsupervised approach uses clustering methods to discriminate into an undetermined number of groups based on similarity (Handfield et al., 2013). Although success of such techniques has been shown, and in some cases can match or even exceed manual human annotation between proteins (Murphy et al., 2003), a vast majority of studies comparing protein changes to a treatment still rely on manual visual inspection of images to determine localisation changes (Breker et al., 2013; Tkach et al., 2012). One of the major limitations lies in the classification system itself, which may be inadequate to describe subtle localisation changes, or proteins that fall under multiple categories (Murphy, 2005; Newberg et al., 2009). For the studies presented here, we decided not to question where a protein was located within the cell, but rather whether there was any change in localisation pattern between treatments, without regard for its similarity to other proteins. To achieve this, accurate cell segmentation scripts were developed to measure textural features at a single cell level and discriminate protein localisation patterns between control and treated cells.

3.2 Image preparation

Prior to analysis, one image must be chosen from the stack of z-planes that have been imaged. Depending on the trait being measured, one can add the values of all planes together to get a total value, generate a maximum projection image from the brightest points across all z-planes, or as in most cases, select the z-plane which most accurately portrays the cell mid-section. For localisation analysis the mid-section provides the highest level of intracellular detail. It is therefore essential that the mid-section is accurately selected to avoid identifying false localisation changes that can occur when comparing images of different heights. For instance, cell wall proteins appear only on the periphery of a cell in a mid-section, but may appear blurred and more cytoplasmic-like if imaged at z planes approaching the top or bottom of the cell. One of the challenges with our screening setup was the variability in optimal focal height across wells. This problem was twofold, firstly the Opera is setup with a high-speed auto-focus laser that measures focal height from the bottom of the well; rather than slower image-based focus systems. Consequently image accuracy in terms of yeast mid-sections is not automated at time of image acquisition. Secondly we imaged live yeast cells freely suspended in media. Thus, depending on how they had settled in the well, cells could be at different heights between wells. These two factors contributed to significant well-to-well variation in optimal focal height, therefore we imaged five Z-planes per well that encompassed this variation. Z-planes were spaced one micron apart, altogether spanning the approximate height of a yeast cell. In order to automate mid-section identification from these z-planes, the gradient of pixel intensity changes was measured across all images and the plane with the greatest standard deviation in values was assumed to represent the mid-section (Figure 4). This is because the mid-section usually has a clear change in contrast at nuclei and cell borders and thus a much greater variance across the gradient. This method is sensitive to bright auto-fluorescent artefacts that often appear in images that are taken too close to the bottom of a well. These can be removed by first filtering out any small or very large objects that are unlikely to be cells. To test the accuracy of the automated selection, images from a 384 well plate were analysed using our automated scripts and compared to z-plane selection by manual visual inspection with 98% agreement. The few differences were due to bright artefacts in the well, outlier cells with extremely bright nuclei, or cases where a choice was made between two z-planes that both presented different cells in mid-section.

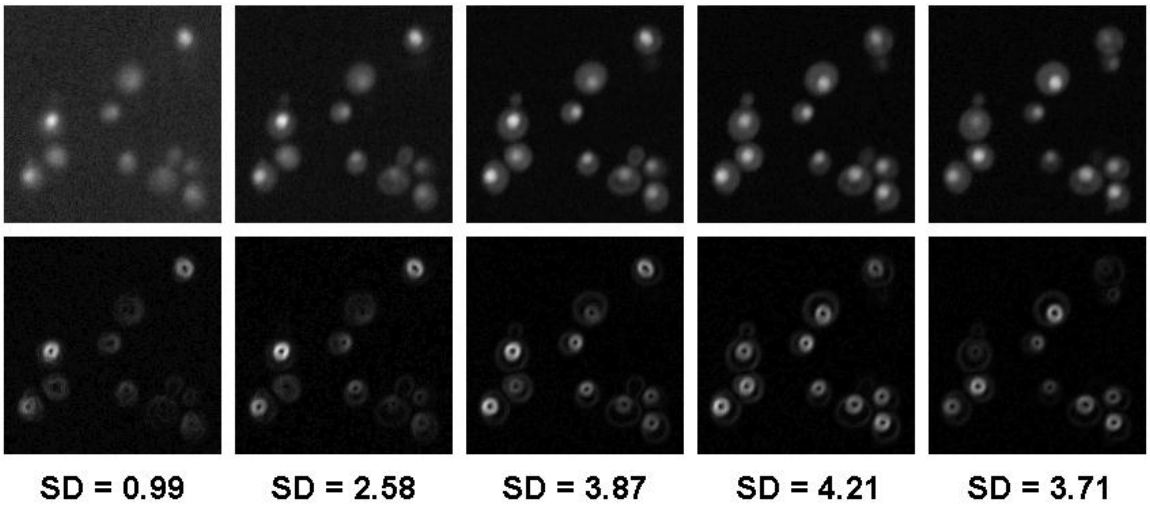


Figure 4 – Comparison of pixel intensity gradients across Z-planes.

The top row is a series of RFP images in greyscale taken at various Z-heights. Below is the corresponding pixel intensity gradient map represented as an image where brighter pixels correspond to a higher gradient slope. The fourth plane from the left has the best defined nuclei and cell borders and highest corresponding gradient standard deviation (SD) used for automated selection.

3.3 Cell Detection without nuclei recognition for Reporter-SGAs

For experiments using consistently well expressed reporter strains such as the 4xUPRE-GFP reporter, we used an automated cell recognition algorithm that only requires cytoplasmic labelling for detection. Because reporter expression is the primary measurement in these screens, we chose to use a constitutively expressed cytoplasmic marker for cell detection that also doubles as a control for protein expression. Thus, mCherry RFP under the control of a constitutively expressed TEF2 promoter was used for cytoplasmic labelling. This system has proved effective in the literature as a control protein to which reporter expression can be normalised (Jonikas et al., 2009).

3.3.1 Cytoplasmic RFP strain construction

The mCherry strain was created by transformation of the usual laboraotory SGA starting strain Y7092 with the PCR product amplified from plasmid pJM002 (a kind gift from David Breslow, UCSF) with primers 224 and 252. This introduced cytoplasmic RFP mCherry under the control of the TEF2 promoter – along with a NatMX4 nourseothricin resistance cassette – into the lyp1 locus of Y7092 resulting in the R-SGA starting strain yCG262.

3.3.2 Cytoplasm based cell detection algorithm

The identification of the cell boundaries based on cytoplasmic labelling was achieved using custom Acapella scripts (Appendix 7.3). Cytoplasmic staining is achieved from the expression of the TEF2pr_mCherry cassette which has been genomically integrated. Firstly a mask of all the pixels that are brighter than a surrounding ring of pixels in the image is created to define potential cell objects. A mask like convolution kernel – representing a ring with an internal diameter of 24 pixels which is approximately equal to 1.3x the width of a yeast cell – was used to convolve the image to infer the surrounding reference intensity. The convolved image was then subtracted from the original image leaving only pixels brighter than their surroundings. These pixels were then used to define a binary mask of bright pixels. This produces a mask of well-defined cell objects, but a high level of noise in the surrounding image background (Figure 5b). The mask is then converted into individual object stencils by defining each separate mask feature as a new object (Figure 5c). The high background noise in the mask produces a number of inappropriate small objects which are removed by setting a minimum size filter to below the size of a yeast cell (Figure 5d). The remaining objects were then checked to see if any are actually multi-cell objects ‘stuck’ together (Figure 5c, arrows) and split into individual cells (Figure 5d).

Next the cell’s object borders are individually fine-tuned using a local threshold for each cell. Local minimum and maximum brightness values are calculated for each cell from the background area surrounding the cell, and the internal region of the cell respectively. These values are defined by rings or ‘zones’ that expand out from, and also within the cell (Figure 5e-f). Pixels within each zone are then added or subtracted from the cell border as they meet or fail the local threshold (Figure 5g).

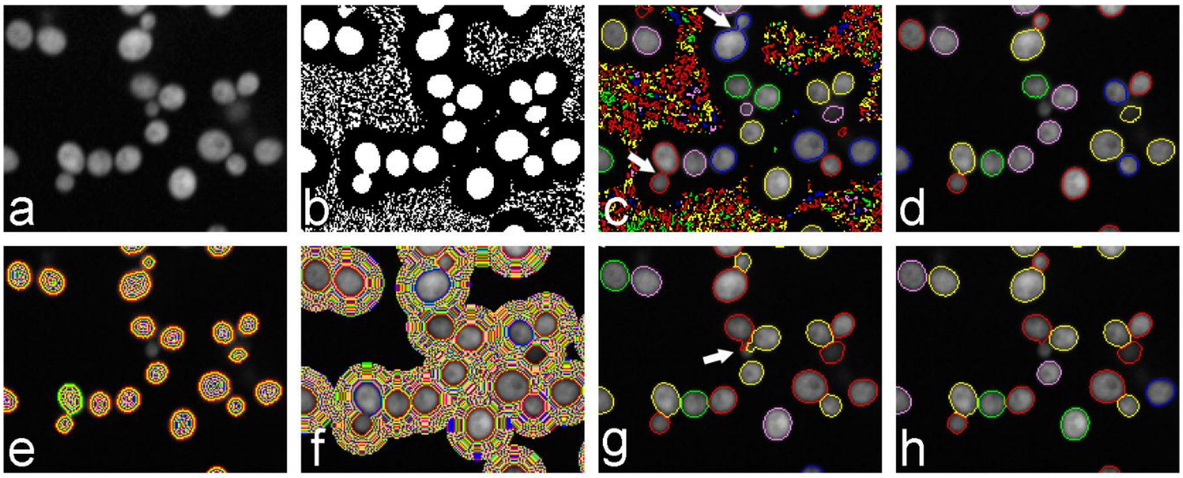


Figure 5 - cytoplasm detection without nuclei.

a) mCherry RFP labelling of cytoplasm; b) Mask of bright pixels highlighting potential cells ; c) Initial cell objects. Different colours indicate individual objects, the arrow points to two groups of ‘stuck’ cells; d) splitting of ‘stuck’ cells and removal of small non-cell objects; e) Inner zones used for individual cell contrast measurements, different colours represent individual zones; f) Outer zones used for individual cell contrast measurements; g) Cells after individual threshold adjustment, arrows point to erroneous projections from the cell border; h) Cell trimming and filtering of cells.

To correct imperfections in the cell recognition, cell borders are smoothed to more closely match the ellipsoid shape expected of a yeast cell. We found an efficient way to achieve this was to 'trim' off any small erroneous projections from the cell border (Figure 5g, arrowed). This can be achieved simply by shrinking the cell border down in size and lowering the resolution of the objects, thus losing any small details along the cell border. The objects can then be expanded back up to the original size but will have a smoother surface (Figure 5h). Objects are filtered based on size and circularity (described below) – to remove any spurious objects; and by intensity – to remove any dead cells which often become highly fluorescent (Breker et al., 2013). Finally any partial cells touching the border of the image are removed.

3.3.3 Measuring circularity

Yeast cells have smooth contours and range from circular to ellipsoidal in shape (de Carvalho et al., 2007). Therefore, a measure of circularity was used as a filter to remove erroneous shapes from the analysis. A perfect circle has a centre equidistant to all points along its perimeter. As a quick and efficient method to assess circularity a measure of the radius ratio (RR) was used. The distance from the mass centre of a cell to each point along it's perimeter was calculated, and circularity defined as the ratio of the minimum and maximum distances (Gordon et al., 2007; Ritter and Cooper, 2009):

$$RR = \frac{radius\ min}{radius\ max}$$

The radius ratio provides a well performing method of measuring circularity that is computationally inexpensive, resolution independent and matches well with human perception of circularity (Ritter and Cooper, 2009). Because RR is a ratio, it measures on a convenient scale from 0 to 1, where 1 would be a perfect circle, as the shape of yeast cells ranges from circular to ellipsoid a radius ratio cut-off of 0.5 was found to be effective for selecting cells. There are however limitations in the case where a cell body is essentially round but the boundary has been incorrectly identified with a small section either sunken or projecting out exaggerating the radius minimum or radius maximum respectively. This was seen commonly in cells that show a large negative space corresponding to the vacuole. For the analysis described in this thesis, accuracy was favoured over cell count and thus, elimination of cells with this problem was not deemed to be significant.

3.4 Cell detection with nuclei detection

3.4.1 Dual RFP marker strain development

Using the mCherry RFP alone for cell border recognition, as described above, is fine for automating relatively simple measurements such as expression changes of a cytoplasmic reporter-GFP. However, for measurement of more complex phenotypes such as localisation changes, cell border accuracy is of the highest importance. For example inaccuracies along cell borders can greatly effect measurement of cell periphery proteins and could lead to false positive or negatives in localisation change. Moreover, whilst using only a cytoplasmic protein marker is possible for accurate cell recognition in some situations, it often leads to inaccurate cell separation when cell density in an image is too high. As we do not have the ability to rigorously control cell counts or distribution of cells in our images we needed a more accurate method of cell separation. In order to facilitate automated cell recognition to identify individual yeast cells with high fidelity, we developed a consistent labelling system to label cells throughout cell cycle stages and/or growth conditions. To this end, we created a dual RFP system specifically labelling the cytoplasm and nucleus of each cell, which could be integrated into the genome of any reporter strain (Figure 6). In addition to the mCherry used to label the cytoplasm above, we used a RFP protein Redstar2 (Bevis and Glick, 2002; Janke et al., 2004), fused to a nuclear localisation signal (NLS) to target the nucleus. In order to ensure as much consistency between individual cells as possible we used the constitutive promoter region of TEF2 to control expression of each RFP reporter (Janke et al., 2004). The use of this simultaneous nuclear and cytoplasmic labelling system allows each object to be centred on the nucleus which will always be well separated from adjacent cells, thereby improving detection accuracy.

A bipartite nuclear localisation signal from SV40 (Hodel et al., 2006, 2001) was used to target the RFP Redstar2 to the nucleus. The construct was created by PCR amplification of two products with homologous flanking regions to the yeast URA3 locus and an overlapping region between them. The TEF2 promoter was amplified from plasmid pYM-N19 with PCR primers 144 and 145, while the RedStar2 RFP along with the nourseothricin (clonNAT) antibiotic resistance marker was amplified from plasmid pYM43 using primers 146 and 147. An overlapping region containing the NLS signal was introduced from primers 145 and 146, and allowed for simultaneous transformation of both products into the genome of the yeast strain Y9230 resulting in the strain yCG215. This strain was then marker switched from a clonNAT to a hygromycin B (HPH) resistance cassette by transforming with the PCR product amplified from plasmid pFA6a-HPH with primers 144 and 240 resulting in the strain yCG251. This strain was further transformed with the PCR product amplified from plasmid pJM002 with primers 224 and 252, introducing the cytoplasmic RFP mCherry under control of the TEF2 promoter into the lyp1 locus, resulting in the strain yCG253.

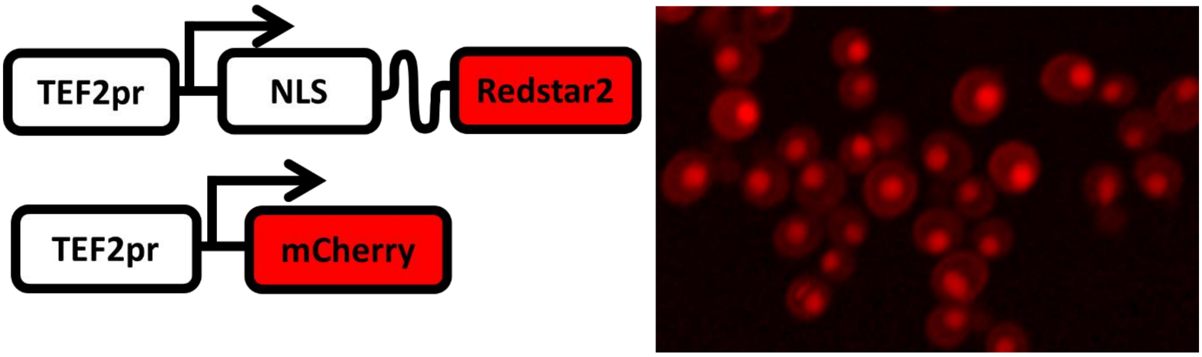


Figure 6 – Schematic of the dual RFP labelling used.

NLS-Redstar2 marks the nucleus and mCherry marks the cytoplasm. Both of these proteins are under control of a ‘neutral’ constitutively expressed promoter TEF2.

3.4.2 Nuclei detection

The first procedure in cell recognition is the identification of cell nuclei. Genomic integration of the TEF2pr_NLS-RS2 construct provided a consistently bright nuclear signal across all cells. This also avoids the need for any additional staining steps and the problems associated with dyes such as DAPI including; loosing cells from washing steps, over or under staining, mitochondrial staining and exposure to more harmful UV laser excitation. Once the optimal z-plane has been selected, a lenient global threshold is applied to the mid-section image to create a binary mask that can be converted into initial nuclei objects (Figure 7a-c). This initial threshold is rather liberal and over estimates the nuclear size (sometimes to the size of a whole cell) to increase the likelihood of encompassing all nuclei that can be further refined in later stages. The initial objects are then checked to see if any larger objects are likely to be multiple nuclei objects ‘stuck’ together from being in close proximity. These compound objects were split up into individual nuclei using built in Acapella procedures (Figure 7d). This step also includes a minimal size filter to remove any non-nuclear objects caused by background noise in the image. Once objects have been identified as likely nuclei, they are further refined into more accurate nuclei objects by applying an individual local threshold to each object and smoothing the final nuclei border (Figure 7e-g). Firstly, zones are created extending both inward and outward of the initial nuclei borders. These zones, and individual pixels within these zones are used to measure local contrast for each nucleus with an individual threshold applied to each nucleus to optimise the boundary. Results of this are generally very accurate since there is such a large contrast difference between the nuclear and cytoplasmic fluorescence. A final filter step removes any objects that are too large to be nuclei (such as those produced by the auto-fluorescence of dead cells or debris in the well), nuclei that fail a circularity test (Figure 7h, see section 3.3.3), and any nuclei touching the border of the image.

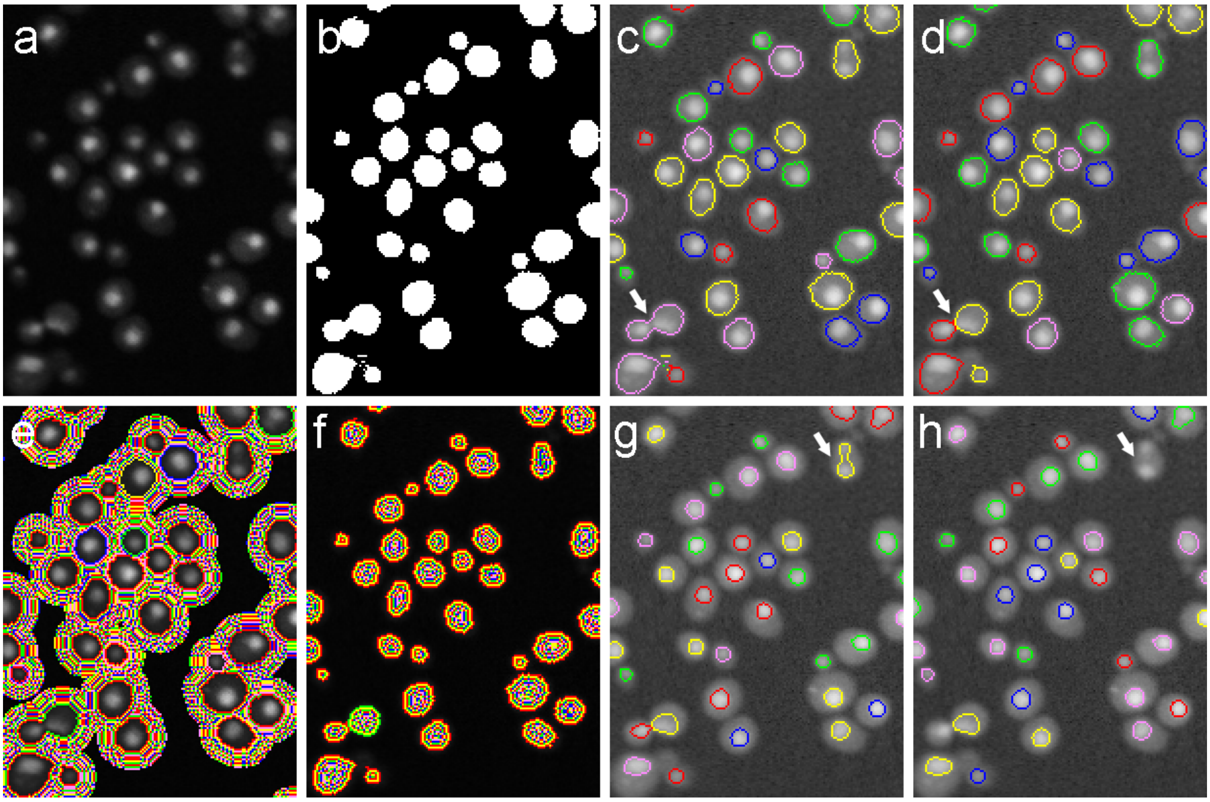


Figure 7 - Nuclei recognition procedure.

a) RFP labelling of nuclei and cytoplasm; b) Mask of potential nuclei from initial global thresholding; c) Initial nuclei objects. Different colours indicate individual objects, the arrow points to two 'stuck' nuclei; d) Splitting of 'stuck' nuclei (arrowed) and removal of small non-nuclei objects (lower left corner); e) Outer zones used for individual nuclei contrast measurements – different colours represent individual zones; f) Inner zones used for individual nuclei contrast measurements; g) Nuclei after individual threshold adjustment; h) Nuclei filtering based on size and shape. The arrow in g and h points to a nucleus that was filtered based on circularity.

3.4.3 Cytoplasm identification

The second procedure in cell recognition is identification of the cell boundaries based on cytoplasmic labelling. Cytoplasmic staining is achieved from genomic integration of the TEF2pr_mCherry cassette. Although both the nuclei and cytoplasmic markers are expressed using the same promoters, mCherry produces a fainter cytoplasm in comparison to nuclear RedStar2. This is due to its fluorescent properties and the fact that it is distributed over a larger area of the cell. However, the large contrast difference can be problematic for accurate cytoplasm detection described above as the thresholding technique tends to pull the cell border in toward the brighter nucleus, underestimating the size of cells. To circumvent this problem, the first stage of the cytoplasm detection algorithm is removal of nuclei fluorescence (Figure 8c). For each identified nucleus, the average cytoplasm fluorescence is measured in a two pixel wide ring, one pixel away from the nucleus border. The nuclei region is then expanded by one pixel and used as a stencil on the RFP image to replace the nuclei staining with the average measured cytoplasm intensity. The whole image then has a mean filter applied to smooth the transition into cytoplasm. The resulting image is similar to that which would be produced if only cytoplasm had been labelled but is suitable for accurate cytoplasm and cell border detection.

Once nuclei have been removed from the image, a mask defining bright points in the image is created to determine initial cell objects (Figure 8d). To increase accuracy of initial cell objects, the image is separated into areas of high or low cell density and an optimised bright mask algorithm used for each. The image is split into 35 subdivisions and the number of nuclei counted in each, any section with greater than 10 nuclei was considered to be high density. The bright mask for low cell density areas was calculated as described in section 3.3.2. The bright mask for high density areas was defined as the top 60% bright pixels within the section. These bright masks produce a set of well-defined cell objects, but a high level noise in the surrounding image background. The mask is then converted into individual object stencils (Figure 8e), from which the majority of noise can be removed by setting a minimum size filter that will remove most spurious objects. The remaining objects are then checked to see if any objects are multiple cells clumped together and if so split into individual cells (Figure 8f).

The next step is to individually adjust the threshold for each cell in order to define more accurate cell borders. To do this, average local minimum and maximum brightness values are calculated for each cell from the background area and the internal region of the cell respectively. Zones are then created that expand out from and also within the cell (Figure 8g-h). These zones and individual points of each zone are then added or subtracted from the cell border as they meet or fail the local threshold (Figure 8i).

Because yeast should have a smooth ellipsoid shape we can further refine the cell border by 'trimming' off any small erroneous projections from the cell border as described in section 3.3.2 (Figure 8i-j). Each cell object is then checked to ensure that there is one nucleus per cell (Figure 8k-l). Any cells with more or less than one nucleus are removed, and any nuclei without a corresponding cell are also removed. The objects are then filtered based on size, intensity and circularity (see section 3.3.2) and any cells touching the border of the image are removed (Figure 8l-m). Finally regions of interest can be added including cytoplasm, membrane and internal regions (Figure 8n-p).

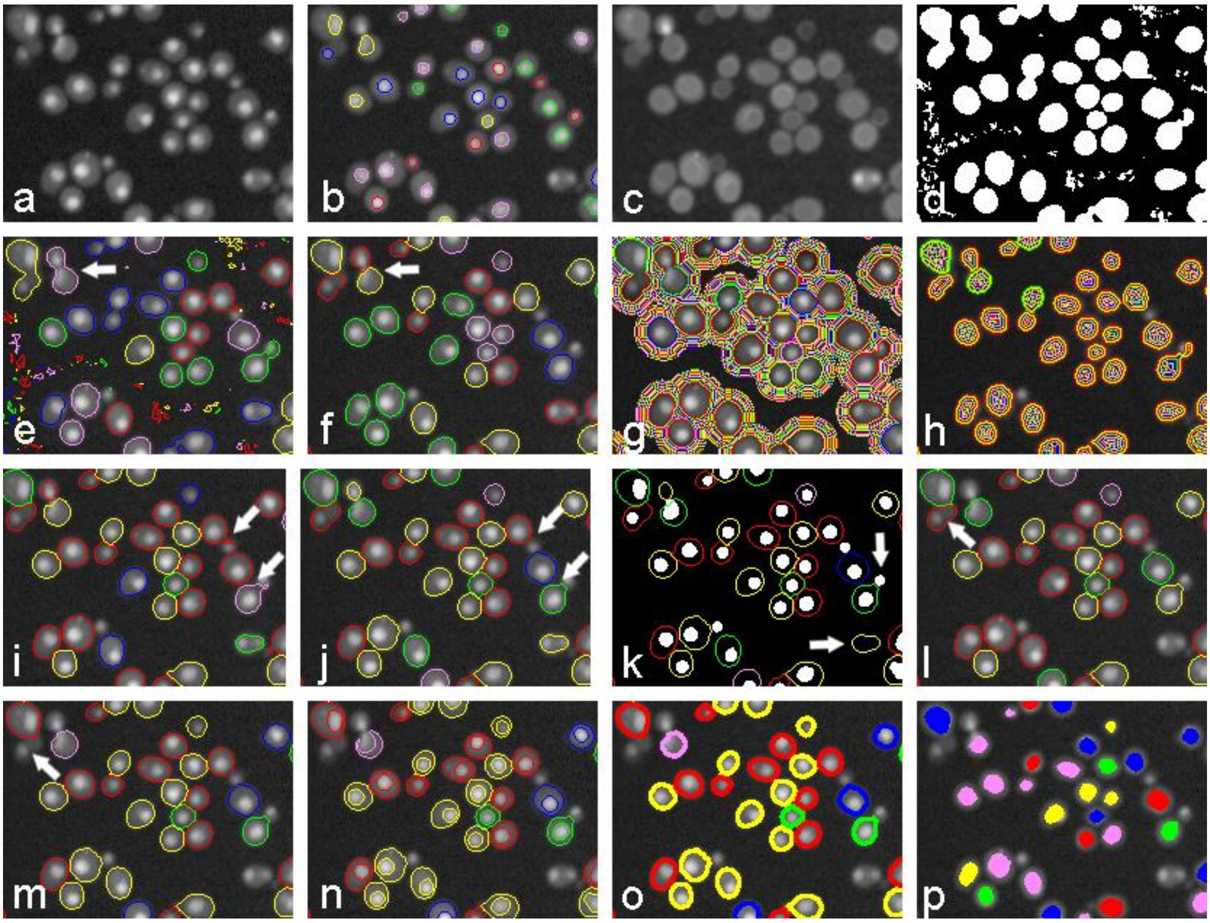


Figure 8 - cytoplasm recognition procedure.

a) **RFP labelling of nuclei and cytoplasm; b) Nuclei as recognised from the nuclei recognition procedure (Figure 7); c) RFP labelling of cytoplasm after nuclear staining has been removed; d) Mask of bright pixels highlighting potential cells ; e) Initial cell objects. Different colours indicate individual objects, the arrow points to two 'stuck' cells; f) splitting of 'stuck' cells (arrowed) and removal of small non-cell objects; g) Outer zones used for individual cell contrast measurements with different colours representing individual zones; h) Inner zones used for individual cell contrast measurements; i) Cells after individual threshold adjustment, arrows point to erroneous projections from the cell border; j) Cell trimming, arrows point to cell borders that have been corrected; k-l) Filtering for cells with single nuclei, arrows in k point to a cell without nuclei and nuclei without a cell; m) Filtering based on size and shape. Arrow in l and m points to a cell that was filtered based on circularity; n) Final cell borders with nuclei outlined representing the cytoplasm region; o) Membrane region; p) internal region.**

3.4.4 Confirmation of cell border recognition scripts

To evaluate the performance of the cytoplasm identification procedure, cell borders were manually drawn for ~1000 example cell images and compared to the corresponding automatically identified cells. Firstly the object centres were identified for the manual and automated cell bodies and the distance between these centres calculated, showing that 95% were no further than 2 pixels away from each other and 30% of cells had perfectly matching centres. The average distance was 0.858 pixels with a standard deviation of 0.655. The areas of manual and automated cells were calculated and showed a correlation of 0.912 in area and 0.889 in perimeter indicating high similarity in shape and size. Coverage of the total percentage of cells identified was estimated from a sample of 1000 images as the percentage of cells identified from the initial nuclei count. An average of 85.2% of cells were recognised that met selection criteria with a standard deviation of 6.2. It was noted that cell coverage was density dependent, as the total cell number increased (estimated from total nuclei) the percentage of cells identified decreases (Figure 9). At low cell density (<250 cells/image) the average coverage was 90.0% ± 3.7, at medium cell density (250-500 cells/image) with the average coverage 84.1% ± 4.5, whilst at high cell density (>500 cells/image) the average coverage dropped to 77.7% ± 4. Even at the highest cell densities the cell recognition procedures still provided a more than adequate sample collection for further image analysis.

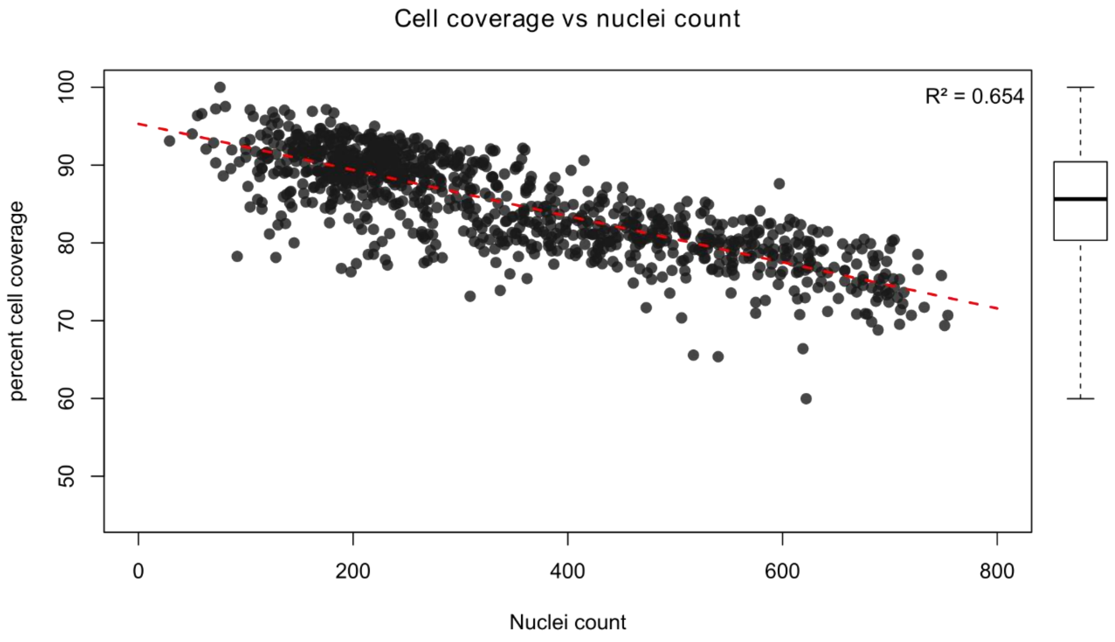


Figure 9 - Percentage of cells recognised

3.5 Feature measurements

3.5.1 GFP expression

GFP expression performs well as a measure of fusion protein abundance and is comparable/correlates to other methodologies such as western blotting and mass spectrometry (Breker et al., 2013; Newman et al., 2006). When measuring GFP-protein fluorescence for expression changes, the intensity values are often normalised against a control RFP protein (Handfield et al., 2013; Jonikas et al., 2009). As such, GFP expression measurements in this study were normalised against the constitutive mCherry RFP expression where possible. GFP expression levels were measured by calculating the pixel intensities within each cell body and normalising the expression as Log_2(GFP/RFP). By averaging pixel intensities within each cell the intensity measurement is independent of cell size, and measuring in Log_2 space means that an increase or decrease of expression will be depicted on an even scale.

3.5.2 Spot detection

Detection of punctate spot patterns of GFP-protein localisation within the cell was achieved using the built-in Acapella spot detection algorithms. The spot detection algorithms were optimised to identify features on order of 2-3 pixels in diameter (~0.5 µm). This size was empirically determined to be appropriate for recognising Ire1p-GFP cluster formation, and thus considered appropriate for general in spot detection in screening the GFP collection for localisation changes (see later chapters).

3.5.3 Texture analysis

Texture analysis offers a means to automate assessment of protein localisation changes without having to manually inspect images by eye and presents a number of advantages. Firstly, the increase in automation provides a more efficient work-flow for identification of localisation changes. Manual inspection of images may be fine for one-off experiments but is unproductive when assessing multiple genomic screens with replicates. Secondly, texture analysis is an unbiased measurement of localisation change, thus removing any human bias or judgment errors and can in some cases perform better than human vision (Murphy et al., 2003). Lastly since texture analysis measures changes in fluorescent intensity patterns, there is no requirement to specifically define subcellular regions or co-localise against organelle markers. This also means one is not limited to any specific organelle which is advantageous as localisation shifts often result in proteins being distributed across multiple locations. An example can be seen in published work of ours, where gene deletions caused a partial blockage in delivery of the plasma membrane protein Mrh1p to the cell surface (Bircham et al., 2011). In this case protein was still delivered to the plasma membrane, but also caused an accumulation in other organelles including the ER, Golgi and vacuole.

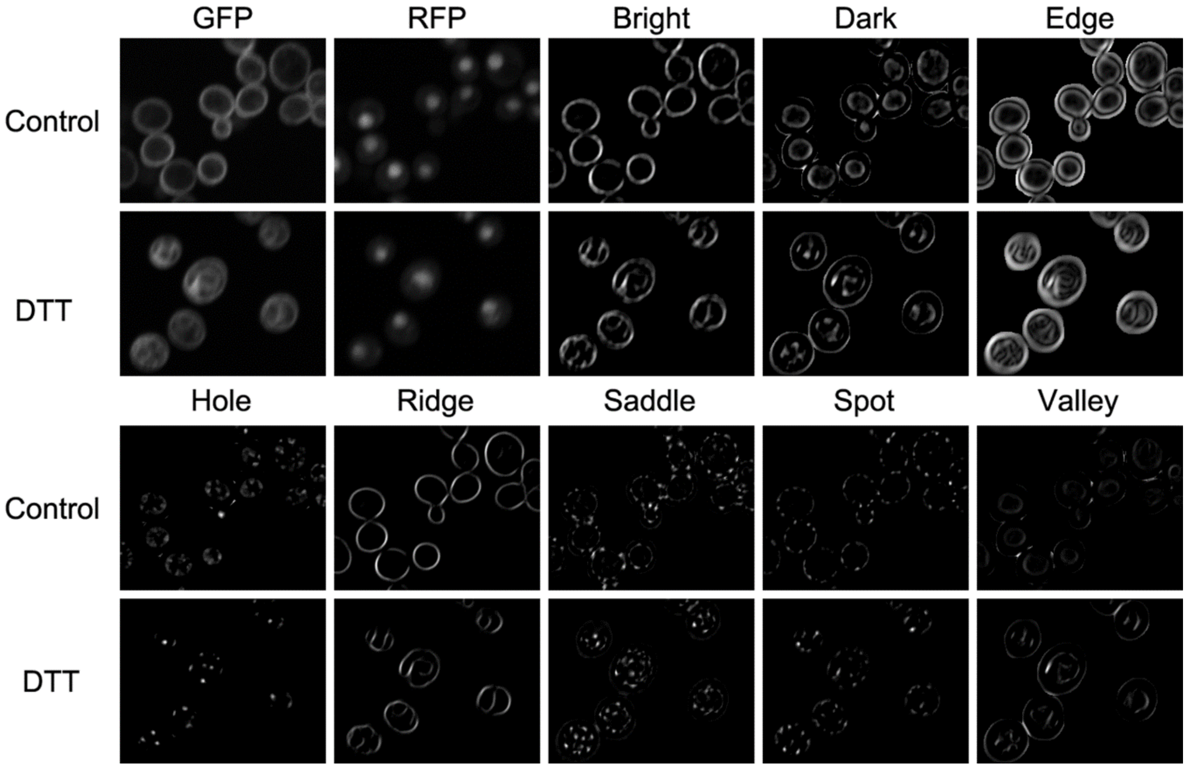


Figure 10 - Visual representation of SER texture features on Bap2p-GFP cells with and without DTT treatment.

Here we used the spots edges and ridges (SER) texture analysis procedures available in Acapella. SER features defines a set of eight characteristic fluorescence patterns, namely: bright, dark, edge, hole, ridge, saddle, spot and valley. These feature names are self-explanatory in terms of the features identified (shown in Figure 10), the size of which is determined by a scale factor. In our case we used two scale factors of 1 and 2. The SER features are examples of textural energy features (Laws, 1980a, 1980b), generated by applying a set of second order Gaussian derivatives filters to the original image (van Vliet et al., 1998). The corresponding texture feature is characterized by mean intensity of the filtered image in that cell region (P. Kask, personal communication, Oct 2012). We have found SER features to perform well with the small size and limited resolution of yeast cells in images, importantly SER features are rotation-invariant and intensity independent and so should be unaffected by variation in fluorescence or cell orientation. Furthermore, because SER features produces only a small set of eight measurements, there is no need for complex dimensionality reduction as required by other textural analyses (Dénervaud et al., 2013; Liu et al., 2004; Singan et al., 2012).

3.6 Discussion

In this chapter we have described the development of automated procedures to identify optimal images containing yeast cell mid-sections and recognise individual cells for further analysis. One of the initial problems early on in these studies was the variation in image quality due to differences in focal planes across wells. Yeast cells were found to settle down to the bottom of the well within ten minutes, yet we still noticed small differences of 1-2 microns between the optimal focal heights across wells leading to significant inaccuracies when measuring textural and localisation changes. Other studies have adhered cells to the bottom of wells to form a monolayer for imaging (Breker et al., 2013), but this could cause alteration in membrane dynamics. We therefore chose to image multiple Z-planes and choose the optimal section per sample during image analysis. To this effect we developed a highly accurate procedure capable of automatically selecting the optimal mid-sections of yeast cells from multiple image stacks based on fluorescence intensity gradients across the images. Testing of the procedure showed a high degree of accuracy and was found to be as effective as manual inspection by eye.

In addition, we also encountered a number of problems with the image recognition procedures supplied with the Opera microscope, which had been optimised for much larger cell types. This often resulted in cells with jagged borders, and poor cell separation in areas of high density. Because a high-level of accuracy is needed for detailed measurement of sub-cellular features we needed to develop a set of yeast-optimised procedures. Here we have described the development of two automated image recognition algorithms that are able to accurately separate individual yeast cells - even in densely clumped regions - and produce a precise mapping of cell borders. The first algorithm is able to rapidly identify yeast cells from cytoplasmic labelling alone and is appropriate for the measurement of cytoplasmic proteins such as the UPRE-GFP reporter discussed in Chapter 4. The second algorithm utilises both cytoplasmic and nuclear labelling for identification of yeast cells with high precision; ideal for texture and expression measurements of a wide variety of GFP proteins such as those in the yeast GFP collection. Both algorithms also exclude outlier cells that are not typical of the overall population to improve measurement reliability.

One of the goals of this project was to set up the image recognition procedures developed in this study as easy to use, fully customisable yeast screening platforms for future experiments. Thus, each procedure was created as a stand-alone module in Acapella, complete with basic usage instructions. These modules have been designed in such a way that they are optimised for the experiments described in this thesis, but also have a set of straight-forward customisation parameters for future users with limited programming knowledge. This should allow easy manipulation of these modules to incorporate any future phenotypic assays. The modules that were developed include:

- **Combine_stack** – takes a stack of images and selects the optimal focus plane, combines all planes, and produces a max projection of the brightest points across all planes. Also has an option for a rolling ball background correction (Sternberg, 1983).
- **Radii_ratio** – measures the circularity of an object
- **Cytoplasm_without_nuclei**
- **Nuclei_identification**
- **Cytoplasm_from_nuclei**
- **Analyse_cells** – measures a number of cellular features including: spots, GFP and RFP intensities, membrane-cytoplasm ratio, nucleus-cytoplasm ratio, and texture features including SER, Haralick and TAS features (Hamilton et al., 2007; Haralick et al., 1973).
- **Z-score** – Calculates the Z score for a feature between two populations
- **MannWhit** – carries out a Mann Whitney U test for a feature between two populations

The code for each scripting module can be found in Appendices 7.1-7.7

Overall we have found the cell recognition procedures described here to be highly effective. There are however a number of enhancements that could be integrated into future versions. Firstly, the intensity-gradient based approach to optimal image selection performed extremely well in selecting the overall best mid-section image. However this could be further improved by implementing the algorithm on a per-cell basis, as cells are not a strict monolayer there is always a portion of cells that have an optimal mid-section up or down one Z-plane. Additionally, it could be beneficial to use image planes directly above and below the mid-section to gain volumetric measurements within the cell, for example of total protein abundance. Using these outer images could also provide additional localisation data, for example yeast mitochondria display complex tubular networks in these outer planes that only appear as small spots in the mid-section. Another appealing feature to add would be cell-cycle analysis based on bud size. This has previously proven to be an effective technique in revealing cell-stage specific responses that may not be apparent when analysing the entirety of the cell population (Handfield et al., 2013). Finally, although the analysis is automated and faster than manual inspection, the analysis can take over 24 hours for data processing to analyse a complete screen of the yeast GFP collection. This is partly due to Acapella being a high-level programming language, which although easy to program is known to cause performance issues in similar applications (Mursalin, 2013). High-level programming languages such as Acapella are processed at run-time by an interpreter which converts the script line-by-line into machine code for execution, thereby slowing the processing speed. Additionally implementing these scripts in an open source programming language would make them more distributable and accessible for other applications.

4 SCREENING THE YEAST DELETION SET FOR UPR GENES

4.1 Background

4.1.1 Use of the DMA for genomic screening

The yeast deletion mutant array (DMA) is a genome-wide set of gene knockouts (Winzeler et al., 1999) that has been widely utilised as a screening platform to elucidate gene function (Boone et al., 2007; Costanzo et al., 2010). Early screens using the DMA typically involved either drug treatment or introduction of a second mutation to define chemical-genetic (Stockwell, 2000) or epistatic genetic interactions respectively. DMA screens such as these have proved successful in identifying novel gene functions in processes of interest, including studies of the secretory pathway (Schuldiner et al., 2005). An example is some of our earlier published work (Bircham et al., 2011) which combined secondary gene mutations and drug treatment of the DMA, to identify additional genes required for secretory pathway function under conditions of unfolded protein stress. Although studies such as these have proved fruitful, the actual assay measurements – that is growth rate based on colony size – is a relatively simplistic indicator of gene function. It seems reasonable then to suggest that cell-location behaviour of specific reporter proteins may offer insight into some of the less obvious effects caused by gene deletions. This has been shown to be the case by both ourselves and others using reporter-SGAs (Bircham et al., 2011; Jonikas et al., 2009; Vizeacoumar et al., 2010; Wolinski et al., 2009). In our case we introduced a GFP labelled plasma membrane protein, Mrh1p-GFP, into the DMA to identify genes required for correct delivery of this reporter to the cell surface. This screen revealed that deletions within the six-membered ER membrane complex (EMC) resulted in an accumulation of Mrh1p-GFP within the cell, despite the fact that none of these genes caused any visible growth defect. Continuing along this rationale, we decided to introduce specific UPR reporters into the DMA to test for gene deletions that induced the UPR, and define genes required for this induction under conditions of induced unfolded proteins stress.

The aims of this chapter are thus:

1. To develop a sensitive reporter system for measuring activity of the unfolded protein response in both standard growth conditions and chemically induced ER-stress conditions

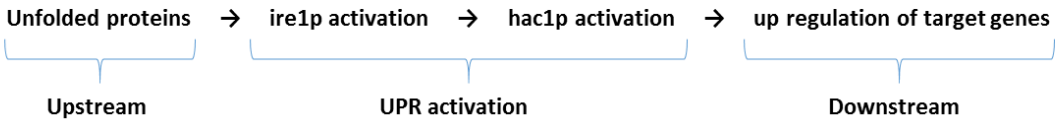
2. To introduce the UPR reporter into the yeast DMA and screen for gene deletions that induce the UPR under standard growth conditions

3. To screen for gene deletions that prevent normal UPR induction under conditions of chemically induced ER-stress

4.1.2 UPR Reporter selection

There are a number of desirable characteristics that need to be considered when deciding on a suitable reporter system. Firstly the reporter needs to have low background levels and be easily measureable in high-throughput assays. The reporter system should have a large dynamic range so the level of reporter activation can be accurately quantified. The reporter should be fast responding so as to provide real-time measurements of the condition being tested. Finally and most importantly the reporter gene should be related to the biological function of interest to limit false-positive results. GFP tagged proteins are often used as they provide a rapid real-time reporter that can be visualised with automated fluorescence microscopy. Additionally, the ability to label native proteins with a GFP-tag can often provide a simple means for developing specific reporters.

To choose suitable candidate reporter proteins we looked at the signalling pathway of the UPR, which at its simplest level can be broken down to:



Upstream of the UPR, an accumulation of unfolded proteins is recognised by Ire1p which in turn activates the response (Cox et al., 1993; Mori et al., 1993). However mis-folding of proteins in a broad sense is not something that can be readily assayed and is therefore not suitable for a reporter system.

The UPR itself requires both Ire1p and Hac1p to be sequentially activated, each of which involves a phenotypic change. As such, Ire1p and Hac1p were both considered as potential reporter candidates for UPR activation.

Downstream of Ire1p and Hac1p activation, UPR target genes are transcriptionally up-regulated following the binding of Hac1p to the UPR-element (UPRE) upstream activating sequences found in the promoter of the target genes. This provides two possible candidates for a reporter system. Firstly, various target genes themselves should change in expression proportionate to UPR activation. This idea has been used in recent studies in our lab by a colleague Yee Low, who used the GFP labelled variants of the UPR target genes Orm2p, Yip3p and Erv29p to measure the downstream UPR effects from overexpression of a misfolded protein CPY* (Low, 2013). The second candidate is using UPRE sequences to drive the expression of a reporter itself rather than tagging a native protein (Mori et al., 1992). This method has the benefit of being UPR specific, without the possibility of non-UPR gene regulation affecting protein expression as is possible when using native proteins.

4.1.3 Ire1p as a GFP reporter

Ire1p functions directly to recognise unfolded proteins in the ER (Credle et al., 2005; Gardner and Walter, 2011), making it the pivotal component in initiating the UPR in yeast. The ER luminal domain of Ire1p contains a core stress sensing domain, which facilitates self-dimerization and formation of an MHC-like groove capable of directly binding unfolded proteins (Credle et al., 2005; Gardner and Walter, 2011). Binding to unfolded proteins results in the clustering of Ire1p into high-order oligomers that function in regulating the activation levels of the UPR (Kimata et al., 2007). By using GFP tagged Ire1p, these clusters can be visualised to determine activation levels of the UPR by exogenous ER-stress agents, such as dithiothreitol (DTT) (Aragón et al., 2009; Ishiwata-Kimata et al., 2013b; Kimata et al., 2007; Promlek et al., 2011).

Although Ire1p cluster formation is a consequence of direct binding to unfolded proteins, it was recently reported that actin plays a role in facilitating efficient cluster formation

(Ishiwata-Kimata et al., 2013b). It is unclear to what extent, if any, other proteins play in cluster formation and thus we hypothesised that Ire1p-GFP cluster formation could be used as a visual reporter to screen the yeast DMA and identify proteins involved in Ire1p activation and cluster formation. Additional proteins required for this process may be identified by screening the DMA for gene-deletions that prevent cluster formation. Additionally Ire1p-GFP cluster formation caused by gene deletions under non-stressed conditions should identify genes involved in protein folding and ER homeostasis.

4.1.4 Hac1p as a GFP reporter

Hac1p is the transcription factor responsible for executing the up-regulation of UPR target genes. Under normal unstressed conditions, HAC1 mRNA (HAC1^{u}) is not translated due to an inhibitory intron. Once activated, Ire1p splices HAC1^{u} mRNA, removing this intron, to form induced HAC1^{i} mRNA. HAC1^{i} is then translated and translocated into the nucleus where it binds to UPRE sequences in the promoters of target genes. This provides a possible reporter system whereby GFP labelled Hac1p would not be present in unstressed cells, however expression would increase and localise to the nucleus under stress conditions. By screening the DMA for changes in Hac1p activity it should be possible to identify gene deletions that affect UPR activation post-Ire1p activation and pre-UPR target gene up-regulation. Screening in ER-stress conditions could potentially discriminate gene deletions that block UPR activation from those that prevent target gene expression.

4.1.5 UPRE as a reporter

UPRE elements are upstream activating sequences present in a majority of known UPR target genes. Thus far three distinct UPRE elements have been identified (Patil et al., 2004). The first element identified is a 22 bp DNA sequence that was originally found in the promoter region of KAR2 (Mori et al., 1992) and is the most prevalent element amongst UPR target genes. This UPRE element has been successfully used to drive expression of a number of reporter systems including LacZ, luciferase and GFP by inserting the UPRE element upstream of a crippled CYC1 promoter such that expression is only induced upon UPR activation (Cox et al., 1993; Pollard et al., 1998). As an alternative to GFP expression that requires fluorescent imaging techniques for measurement, we hypothesised that UPRE elements could be used to drive the expression of ADE2 gene which is involved in purine biosynthesis. Mutations in ADE2 cause a block in the purine nucleotide biosynthetic pathway, thus when these strains are grown on low adenine media; this results in the accumulation of precursor product, P-ribosylaminoimidazole (AIR), in the vacuole. When oxidised, AIR turns red, giving a colour change to yeast colonies that can be used as a simple and easily measured reporter readout (Ng, 2005; Ng et al., 2000). In this case, red colonies would indicate unstressed cells and low expression of ADE2 whereas induced UPR activity would drive expression of ADE2 and result in completion of the purine biosynthesis pathway seen by white colonies.

4.2 Ire1p-GFP DMA screen

4.2.1 Ire1p-GFP reporter strain construction

To generate the Ire1p-GFP reporter strain, the cytoplasmic and nuclear red fluorescent protein (RFP) marker strain yCG253 was mated with the Ire1p-GFP strain taken from the yeast GFP collection (Invitrogen, Huh et al., 2003; see Materials and Methods). The resulting strain had the correct markers but was of a *MAT**a*** mating type which is unsuitable for mating with the DMA which is also *MAT**a***. To generate a *MATα* strain, random spore selection was used (see Materials and Methods) in which the above strain was mated with the *MATα* wildtype strain yCG307 containing a selectable cen::LEU2 plasmid for use in diploid selection. Spores were generated and germinated, and the *MATα* Ire1p-GFP reporter strain yCG458 was selected in which the cen::LEU2 plasmid was lost and the desired reporters were retained.

4.2.2 Ire1p-GFP reporter validation

To test the efficacy of Ire1p-GFP as a reporter, cells had ER-stress chemically induced using DTT and were visualised with automated confocal microscopy. The optimal DTT concentration for Ire1p cluster (foci) formation was determined by performing a dose-response curve ranging from 0.125 – 8 mM DTT. Cells were grown for four hours at 30°C in a 384 well clear bottomed microtitre plate containing 50 µL of SC + DTT. Cells were imaged using the Opera confocal microscope and automated image recognition procedures were used to count the number of foci per cell (see Chapter Three). As can be seen in Figure 11a, the number of foci per cell did not have much variation over different DTT concentrations.

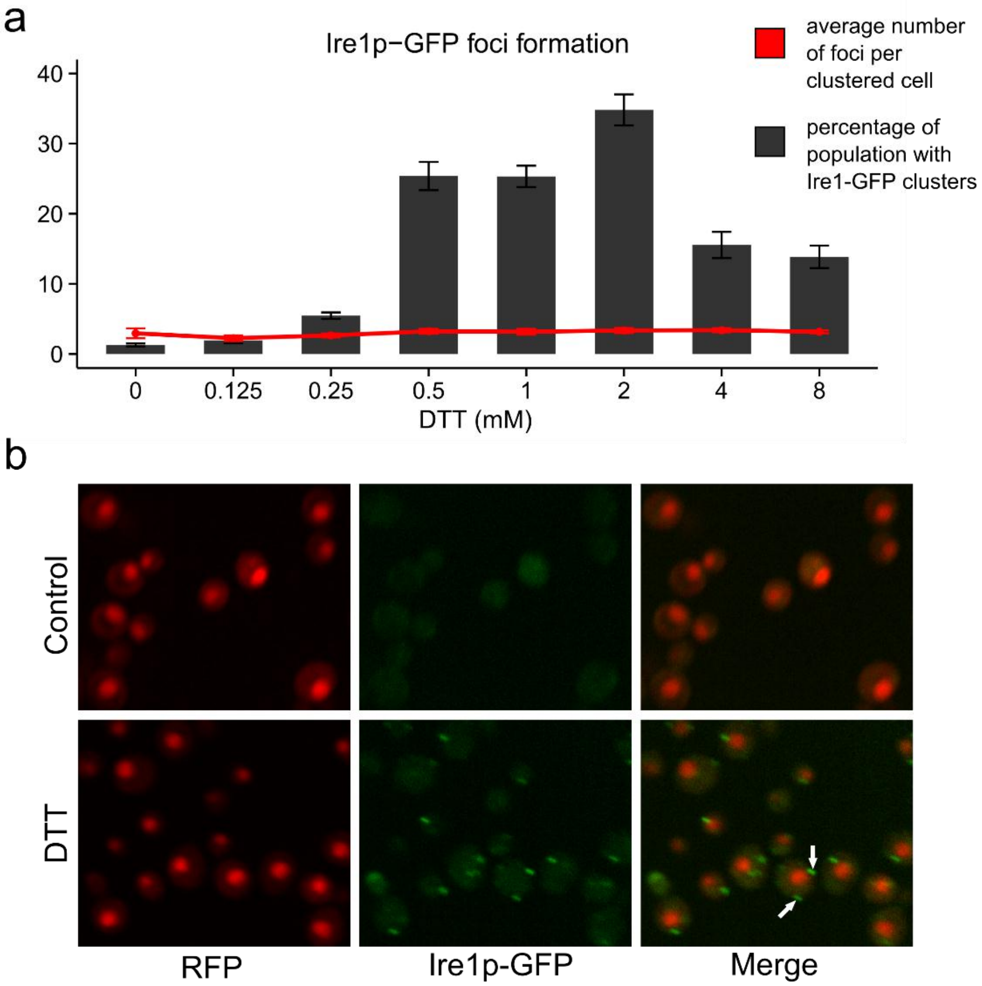


Figure 11 - Ire1p-GFP foci formation in ER-stress conditions.

a) Ire1p-GFP foci formation after 4 hours growth in different concentrations of DTT; Bars – percentage of cells of cells with foci; Line – average number of foci formed per induced cell. Error bars show standard deviation. b) Top and bottom pannels show control and 2 mM DTT treated cells respectively. Red fluorescence arises from mCherry, a cytoplasmic marker, and nuclear localisation signal RFP (NLS)-RedStar2. Green is Ire1p-GFP. White arrows point to examples of Ire1p-GFP foci.

The average number of foci per cell ranged from 2-4, and a small portion of control cells, i.e. no DTT was added, had foci. However, counting the percentage of the total cell population displaying foci appeared to be a more reliable measurement of Ire1p activation, increasing in a dose-dependent manner and peaking at 2 mM DTT. Concentration above 2 mM DTT resulted in a higher percentage of Ire1p clustered cells, however the foci were less well defined and there was an increased number of dead cells which may affect analysis. The fact that Ire1p-GFP clustered as expected under ER-stress conditions (see Figure 11b) is indicative that a C-terminal GFP tag does not impair the proteins ability to recognise misfolded proteins and consequently form high-order oligomer clusters.

4.2.3 Identification of gene deletions that *induce* Ire1p cluster formation

In order to identify gene deletions that induce Ire1p-GFP cluster formation and are therefore likely to have caused protein mis-folding, the Ire1p-GFP reporter strain yCG458 was introduced into the DMA following standard R-SGA procedure (see Materials and Methods for details). The resulting DMA/reporter library was then imaged after four hours growth in 50 µL SC media in fourteen 384 well microtitre plates using the high-throughput Opera confocal microscope and standard settings (refer to Materials and Methods). The images were assayed for foci formation using the Acapella scripts for spot detection described previously (Chapter Three) and statistical analysis was conducted using R software (R Core Team, 2013). The percentage of cells showing Ire1p foci for each deletion strain was compared to the median of its corresponding plate to account for plate to plate variation. P-values were calculated for each strain using Fisher's exact test of independence, testing for gene deletions that showed a greater than expected level of foci formation. Fisher’s method was used to combine p-values across replicates. Gene deletions were considered a hit if their p-value fell below the threshold of 0.005. Genes annotated as dubious ORFs in Saccharomyces Genome Database (SGD; Cherry et al., 2012) were removed from the analysis. Results are summarised in Table 6 and genes were grouped in broad functional categories based on those presented in Costanzo et al., 2010:

Table 6 - Gene deletions that induce Ire1p-GFP cluster formation

Gene deletions that induce Ire1p-GFP cluster formation with a p-value less than 0.005. Functional categories were adapted from those used in Costanzo et al., 2010.

Functional Category	**Genes**
Amino acid biosynthesis	ASI1, PRS5, VBA5
Cell polarity/morphogenesis	ABP140, AIM3, RGA2, SAC7, STE3
Chromatin/transcription	SGF29, YPR022C
Cell cycle progression/meiosis	SPO19
DNA replication	HSM3
Drug/ion transport	PNS1, TPO5, YBT1, ZRC1, ZRT3
ER-Golgi traffic	ERV29, RUD3, SHH3, TRS85

Cell cycle progression/meiosis	CLN3, FKH1, POG1, SPO74
Golgi/endosome/vacuole/sorting	GDA1, PRB1, SIW14, SSO2
Lipids	PLB3, SSP2, VAC14
Metabolism	ASP1, EHT1, GID8, IMP2, KGD2, MAL13, MRS4, NDE2, NPP1, OM45, PCP1, PEX1, PEX6, PXA2, RMD5, SUC2, URA8, VID30, YDL119C, YNL200C
Ribosome/translation	ATS1, RCM1, RPL37A, RPL40B, RPS18B, RSA3, SCD6, TRM82
Signalling/stress response	CNB1, NMA111, PPT1, SVF1, YAP5
Protein degradation/proteasome	DFM1, HLJ1, HRD1, STE24, UBR1, UBX2, UBX4
Protein folding/glycosylation	ACF2, ALG12, ALG5, ALG9, EMC1, EMC3, EMC5, HAC1, LAS21, NBP2, OST5, PMT1, PMT2
Unknown	BOP3, CUE3, ESL1, ESL2, FMP16, LEE1, SSP120, TOS1, YBR219C, YEL057C, YGR266W, YIL163C, YKL033W-A, YKR051W, YMR196W, YPR078C, YPR147C, YPR148C, YSY6

Non-redundant GO categories that showed enrichment with a false discovery rate (FDR) less than 0.1 were obtained from Yeastmine (Balakrishnan et al., 2012) by comparison to a background population set of all gene deletions present in the DMA. GO-Biological process terms that showed enrichment included proteolysis (17 genes), ER-associated ubiquitin-dependent protein catabolic process (7 genes), protein glycosylation (7 genes) and cellular response to topologically incorrect protein (5 genes). GO-Cellular component terms that showed enrichment included Endoplasmic reticulum (19 genes) and organelle membrane (36 genes).

Interestingly two publications, Jonikas et al., 2009; Schuldiner et al., 2005, showed enrichment for this set of genes with a FDR < 0.05, both of which were studies investigating the secretory pathway/UPR using the DMA. The Jonikas dataset is particularly interesting as they used a UPRE reporter to screen the DMA using flow cytometry for genes involved in protein folding. Comparison of their dataset of UPRE inducers to our Ire1-GFP cluster inducing genes shows an overlap of 28 genes (27 % of our hits). It should be noted that their dataset used different cut-offs to define hits and as such has a larger hits list than ours (318 compared to 104). Cells were also grown in liquid media prior to screening as opposed to the agar media in our screen which would influence differences in growth rates and protein expression.

Another two datasets of interest are presented in Travers et al., 2000 and Kimata et al., 2006, both of which used cDNA microarrays to measure mRNA expression changes in response to ER-stress to define the set of UPR target genes. Travers induced ER-stress using tunicamycin and DTT whereas Kimata used tunicamycin and a constitutively active UPR strain. Comparison of our Ire1p-GFP cluster inducing hits to their combined datasets (after removing dubious ORFs and genes not represented in our array) shows only a limited overlap of 14 genes (13% of our hits, Figure 12). This may indicate that most of the gene deletions inducing Ire1p clustering are not required for actively dealing with misfolded proteins, but rather their absence is causing protein misfolding to occur by another mechanism. This may be the reason that only around 17 % of our hits are annotated to be involved in protein folding or degradation (Table 6).

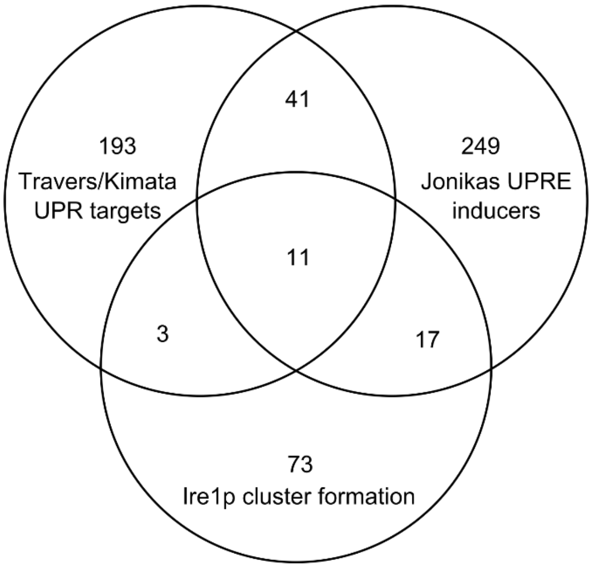


Figure 12 - Ire1p cluster formation literature comparison

Overlap between gene deletions that cause Ire1p cluster formation and UPR targets as defined by Kimata et al., 2006 and Travers et al., 2000; and gene deletions that induced expression of a UPRE reporter as defined by Jonikas et al., 2009. Dubious ORFs and genes that were not present in the DMA were removed from the datasets before comparison.

4.2.4 Gene deletions that *prevent* Ire1p cluster formation under ER-stress

To identify genes that may be required or assist in recognition of misfolded proteins and activation of Ire1p, the Ire1p-GFP reporter DMA was screened under ER-stress conditions. Cells were grown in 50 µL SC containing 2 mM DTT for four hours at 30°C to induce ER-stress. Under these conditions wild-type cells displayed an increased percentage of cells having Ire1p-GFP foci (Figure 11). Cells were screened with the Opera confocal microscope using standard conditions, and images were analysed using Acapella scripts for foci detection (see Materials & Methods). DTT treated cells were compared to untreated cells of the same strain using Fisher's exact test of independence and p-values across replicates were combined using Fisher's method. Gene deletions were considered not to have induced Ire1p clustering in strains that could not reject the null hypothesis i.e. strains with a p-value > 0.05. Additionally strains were considered to have lower than expected Ire1p clustering if the percentage of cells showing foci was less than 10 % across all replicates. Dubious ORFs were omitted resulting in a list of 13 hits, shown in Table 7:

Table 7 - Gene deletions that prevent Ire1p clustering after DTT treatment

Gene deletions that prevent proper Ire1p cluster formation. Columns showing % foci represent the percentage of cells in the population showing Ire1p clusters, averaged across replicates. P-values were calculated using Fisher's exact test and combined across replicates using Fisher's method. Strains with p-values above 0.05 (highlighted red) could not reject the null hypothesis indicating that control and treated populations are not statistically distinguishable.

ORF	Gene	Control % foci	Treated % foci	Description
YOR067C	ALG8	2.66	2.5	Glucosyl transferase
YNL169C	PSD1	0.47	2.6	Phosphatidylserine decarboxylase of the mitochondrial inner membrane
YNL284C	MRPL10	0	0.8	Mitochondrial ribosomal protein of the large subunit
YKL073W	LHS1	0	1	Molecular chaperone of the endoplasmic reticulum lumen
YJR073C	OPI3	0.92	5.3	Phospholipid methyltransferase
YOR106W	VAM3	0.13	2.1	Syntaxin-like vacuolar t-SNARE that functions in vacuolar trafficking
YKL006W	RPL14A	0	2.9	Ribosomal 60S subunit protein L14A
YJR145C	RPS4A	0.47	4.7	Protein component of the small (40S) ribosomal subunit
YNL162W	RPL42A	0.35	3.4	Ribosomal 60S subunit protein L42A
YKL081W	TEF4	0.8	5.3	Gamma subunit of translational

				elongation factor eEF1B
YMR243C	ZRC1	9.25	5.4	Vacuolar membrane zinc transporter
YML017W	PSP2	1.27	9.7	Asn rich cytoplasmic protein that contains RGG motifs
YGR092W	DBF2	0	5.7	Ser/Thr kinase involved in transcription and stress response

These genes showed no significant GO enrichment, although 10 of the 13 were shown to have physical interactions with UBI4 (ubiquitin) and 7 of the 13 with UBP3 (UBiquitin-specific Protease). It should also be noted that some of the hits which showed no Ire1p cluster formation were larger than expected for haploid cells. Of all the hits LHS1 and OPI3 are the only genes known to be a UPR target gene (Kimata et al., 2006; Travers et al., 2000), although 5 of the 13 hits were UPRE inducing gene deletions in Jonikas et al., 2009. Interestingly 10 of the 13 were UPRE inducers, identified as defined later in this chapter. Additionally it is intriguing that deletion of ZRC1 both increased the number of Ire1p foci under non-stressed conditions and decreased the number under stressed conditions.

4.3 Hac1p-GFP as a potential reporter of UPR activation

Hac1p is the transcription factor mainly responsible for the up-regulation of UPR target genes. As described previously, HAC1 transcript must be spliced by Ire1p in order to be efficiently translated, and subsequently translocated into the nucleus where it activates the transcription of target genes containing UPRE upstream activator sequences. Hac1p is able to up-regulate its own expression via a UPRE binding site within its promoter region (Ogawa and Mori, 2004). We hypothesised that measuring both nuclear translocation and expression levels of Hac1p tagged with GFP (Huh et al., 2003) would provide a useful reporter to assess UPR activation. Cells were transferred from agar plates into a 384 well clear-bottomed microtitre plate containing 50 µL SC + DTT with concentrations ranging from 0-8 mM and grown at 30°C. Cells were imaged at 30 minute intervals and assessed for nuclear translocation and nuclear GFP expression levels, selected time points are shown in Figure 13 below.

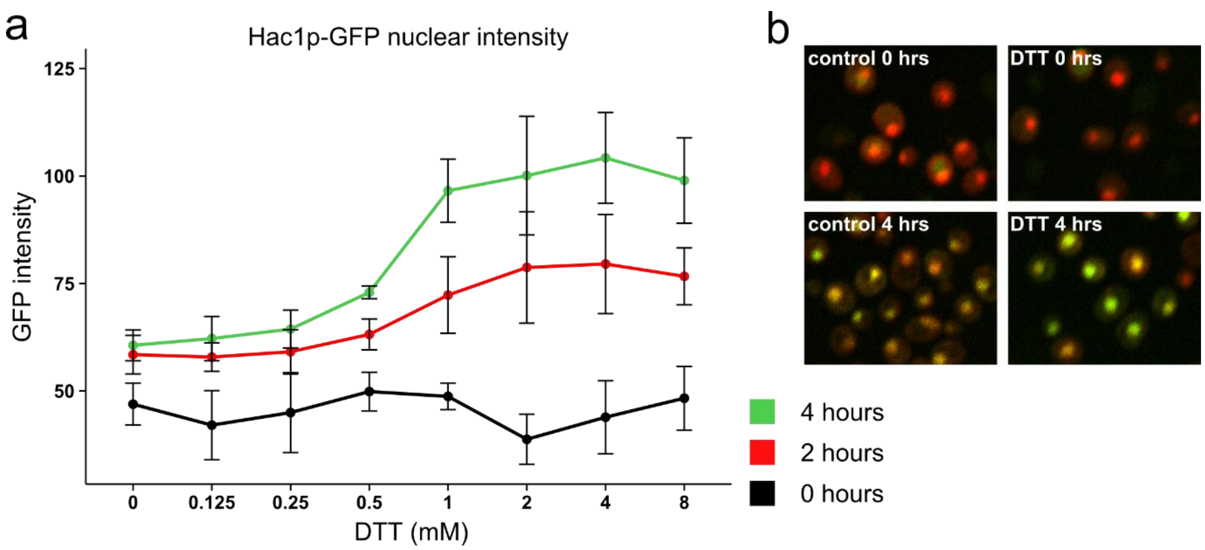


Figure 13 - Hac1p-GFP activation in ER-stress conditions.

a) DTT dose-response depicting Hac1p-GFP nuclear GFP expression. Error bars show standard deviation. b) Top and bottom panels show control and 2 mM DTT treated cells respectively. Red fluorescence arises from mCherry, a cytoplasmic marker, and nuclear localisation signal RFP (NLS)-RedStar2. Green is Hac1p-GFP. Nuclear localisation is not initially seen (red nuclei at 0 hours), however at 4 hours growth both control and treated cells show nuclear localisaed Hac1p-GFP. DTT treated cells have more nuclear Hac1p-GFP expression as seen by the distincly more green nucleus matching the data in panel a).

In our growth conditions we found rapid nuclear translocation in both control and DTT treated conditions. Presumably transferring cells into fresh media causes enough protein misfolding to induce some Hac1p activation. It has been previously shown in normal growth conditions up to 30% of nascent proteins can misfold and induce measurable but low levels of UPR activation (Jonikas et al., 2009; Schubert et al., 2000).

Although nuclear translocation occurred in both control and treated cells, nuclear GFP intensity showed a dose-dependent increase of Hac1p-GFP expression in response to DTT (Figure 13). However the dynamic range of this expression increase was not large enough to employ it as a suitable reporter. Although Hac1p-GFP expression increases with DTT concentration there is a large standard deviation for each measurement (shown as error bars in Figure 13), thereby limiting the discriminatory power of Hac1p-GFP as a reporter. This is particularly limiting in high-throughput screening given the large number of measurements and thus the need for statistically robust measurements. Because of this limitation and the lack of distinguishable nuclear translocation between control and treated cells, we deemed Hac1p-GFP to be an unsuitable reporter for measuring UPR activation in high-throughput screens.

4.4 UPRE reporter DMA screen

4.4.1 UPRE-ADE2 reporter construction

To make a simple colour based reporter assay for UPR activity that does not require further detection like fluorescent systems do, we hypothesised that Ade2p expression could be controlled by a UPRE-promoter construct as shown in previous works. By replacing the promoter of the Ade2 gene with a crippled cyc1 promoter (Guarente and Mason, 1983) containing an upstream UPRE element (Mori et al., 1992) the Ade2 enzyme should only be present when the UPR is activated. By introducing this into the YGDS, UPR activation should be apparent by the appearance of white colonies while those with compromised UPR signalling should remain red.

Construction of the UPRE-ADE2 reporter strain was carried out by a two-step sequential transformation of the SGA starting strain Y7092 (see materials and methods for PCR and transformation protocol). Firstly a crippled cyc1 promoter along with an upstream NatR antibiotic resistance cassette was introduced upstream of the ADE2 gene replacing the native promoter. This was achieved by transformation of the PCR product from plasmid pYM-N11, amplified using primers #148 and #149 into Y7092 generating the strain yCG214. This strain was further transformed with a PCR product containing a KanMX resistance cassette and 22 bp UPRE sequence. This product was amplified from plasmid pYM-N22 using primers #148 and #176, generating the UPRE-ADE2 reporter strain. This second transformation replaces NatR with the KanMX resistance cassette and introduces the UPRE activating sequence upstream of the crippled cyc1 promoter, transforming it into a UPR inducible promoter.

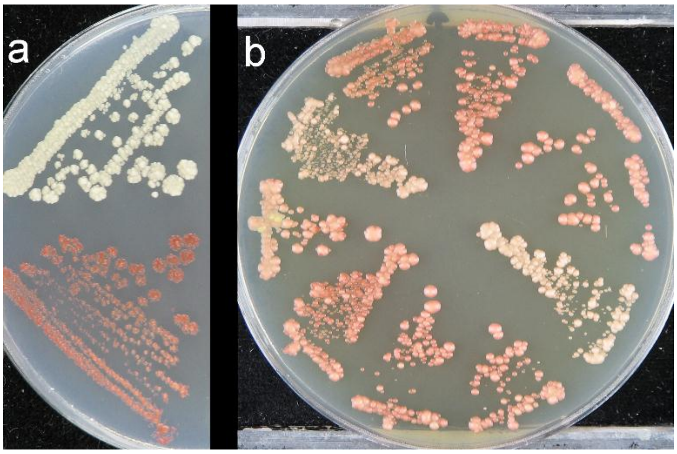


Figure 14 - UPRE-ADE2 reporter

UPRE-ADE2 reporter strain grown on low adenine media. A.) untreated conditions; top – Y7092 starting strain; bottom – UPRE-ADE2 reporter strain showing red colonies. B.) UPRE-ADE2 reporter strain replicates grown on 0.5 µg/mL tunicamycin, as can be seen there is a large amount of variation in colour between replicates and none of the strains expressed enough Ade2p to prevent pigment accumulation.

4.4.2 UPRE-ADE2 reporter testing

The UPRE promoter replacement strains resulted in generation of colonies that grow red on low adenine media as expected (Figure 14a). However colonies sometimes required a couple of days at 4°C to develop a significant amount of pigment accumulation able to be visualized. Moreover, when grown on a relatively high concentration of the UPR inducing agent tunicamycin (0.5 µg/mL), the UPRE reporter failed to consistently induce enough ADE2 expression to produce non-pigmented colonies (Figure 14b). Additionally the reporter showed a large amount of colour variation in replicates. It is possible that adding additional UPRE repeats in the promoter could enhance the response to ER-stress, however the requirement of a long pigment accumulation time could not be fixed and consequently the UPRE-ADE2 construct was not considered appropriate as reporter for acute ER-stress induction.

4.4.3 Construction of 4xUPRE-GFP strain

There have been reports in the literature describing a Hac1p responsive GFP-reporter (Jonikas et al., 2009; Pollard et al., 1998) therefore we decided to construct one for this work. Tandem repeats have been shown to increase the sensitivity of UPRE reporters

(Cox and Walter, 1996) and as such we chose to use a 4xUPRE-GFP construct as a reporter. The 4xUPRE-GFP reporter strain was created by simultaneous transformation of two PCR products with an overlapping region of homology. The first PCR product containing the URA3 (uracil biosynthesis gene) as a selectable marker and the 4xUPRE portion from plasmid pPM47 (Addgene plasmid #20132, Merksamer et al., 2008) was amplified using primers #257 and #259. The second PCR product containing the gene encoding the yeast enhanced GFP variant (yeGFP), was amplified from plasmid pYM-N21 using primers #260 and #261. A region of homology between the two PCR products to facilitate homologous recombination and also join the 4xUPRE promoter to the yeGFP gene. Flanking regions of homology to the genomic regions directly up and down-stream of the MET17 ORF, facilitating transformation into this locus. These PCR products were transformed into the SGA background strain Y7092 generating the strain yCG261. This strain was then further transformed with a PCR product consisting of the TEF2 promoter driving expression of the cytoplasmic RFP mCherry along with the selectable maker clonNAT amplified from plasmid pMJ002 with primers #224 and #252 generating yCG266.

4.4.4 4xUPRE-GFP strain validation

To test the effectiveness of the 4xUPRE reporter strain, cells were exposed to chemically induced ER-stress using DTT and visualised with confocal microscopy. A dose-response ranging from 0.125 – 8 mM DTT was run in triplicate to determine the optimal DTT concentration for UPRE expression. Cells were grown for four hours at 30°C in a 384 well clear bottomed microtitre plate containing 50 µL of SC + DTT. Cells were then imaged using the Opera confocal microscope and Acapella image recognition procedures were used to measure the level of UPRE-GFP abundance in each cell. GFP intensity was normalised against the expression of the constitutively expressed mCherry RFP protein and measured in $\log_2$ space (see Material and Methods). As can be seen in Figure 15, expression of the 4xUPRE-GFP construct increases in a dose-dependent manner, plateauing around 2 mM DTT. The response showed a good dynamic range for reporter activation and importantly there was a basal level of reporter expression enabling screening for gene deletions that decrease UPRE expression in unstressed conditions.

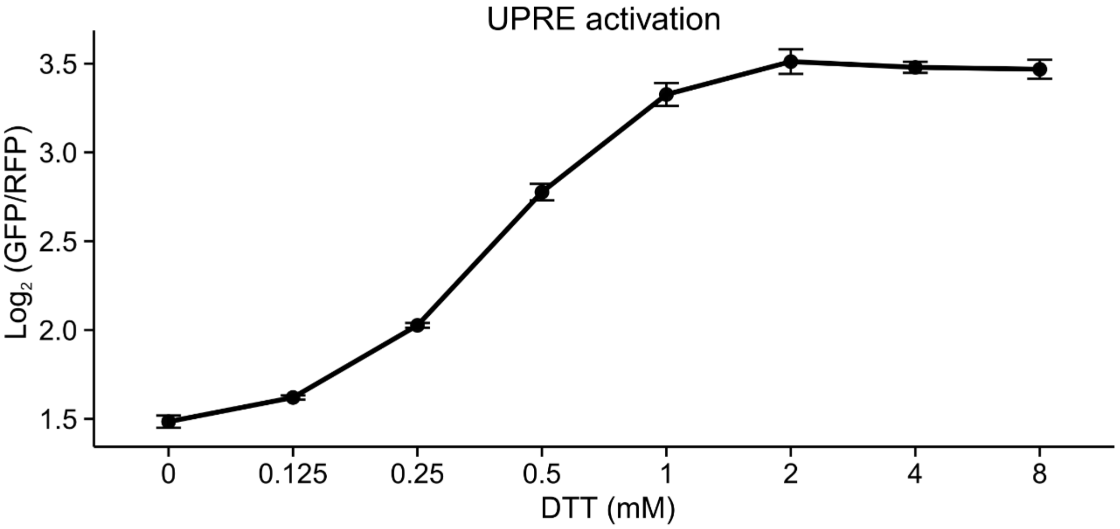


Figure 15 - 4xUPRE-GFP DTT dose response

UPR activation as a function of DTT concentration after 4 hours of treatment as measured from the expression of a 4xUPRE-GFP reporter and normalised against the expression of a constitutively expressed RFP protein TEF2pr_mCherrry. Error bars display standard deviations.

4.4.5 Gene deletions that induce UPRE-GFP expression

To screen for gene deletions that cause UPR activation and consequently UPRE-GFP reporter expression, the 4xUPRE-GFP reporter strain yCG266 was introduced into the DMA following the standard R-SGA procedure (see Materials and Methods). The resulting DMA/reporter library was then imaged in 384 well microtitre plates after four hours growth in 50 µL SC media at 30°C using the Opera confocal microscope and standard settings (see Materials and Methods). Images were then analysed to measure single cell UPRE-GFP expression using a simpler version of the Acapella scripts that do not require nuclear labelling or detection, as described in Chapter Three. GFP intensities were normalised against the intensity of the consistent mCherry cytoplasmic control RFP. UPRE activation was thus measured as Log$_2$(GFP/RFP) and statistical analysis was conducted using R software (R Core Team, 2013). The average UPRE expression of cells from each deletion strain was compared to the median of its corresponding plate to account for plate to plate variation. Z-scores were calculated by comparing the median UPRE induction of each gene deletion strain to the plate median to account for plate to plate variations and measured as:

$$Z - score = \frac{UPRE\ activation_{treated} - UPRE\ activation_{plate}}{MAD_{plate}}$$

Where $\mathrm{MAD}_{\mathrm{plate}}$ is the median absolute deviation of the plate. Z-scores were then combined across triplicates using Stouffer's Z-score method (Stouffer et al., 1949; Whitlock, 2005) and converted into p-values. A cut-off threshold of 0.01 was used for strains showing higher than expected UPRE activation and gene deletions were considered a hit if their p-value fell below this. Genes were removed if they were annotated as dubious ORFs in SGD. Results are summarised in Table 8 and grouped into broad functional categories based on those presented in Costanzo et al., 2010:

Table 8 - Gene deletions that induce UPRE-GFP expression

Gene deletions that induce UPRE-GFP expression with a p-value less than 0.01. Functional categories were adapted from those used in Costanzo et al., 2010.

Functional Category	**Genes**
Amino acid biosynthesis	BAS1, DAL80, GTR1, MEH1, MTC5, TRP3
Cell polarity/morphogenesis	ARC18, EDE1, MYO3, PIN3, PRR1, SHE4, SPS22
Chromatin/transcription	BDF2, CSE2, CTK1, FUN30, HDA3, HIR2, HIR3, HMS1, HTZ1, KNS1, LEO1, MED1, SIF2, SNT1, SPT21, SWC3, UME6
Chromosome segregation	DBF2, MAM1, RBL2, TOF2
DNA replication	ESC2, HAM1, MSC1, MSH6, RAD27, RTT109
Drug/ion transport	ICE2, PMC1, PMP3, SKY1, YHM2, ZRC1, ZRG17
ER-Golgi traffic	CPR7, ERP1, ERV25, GCS1, GET3, PHO86, SEC22, SIL1, SPC2, YER084W
Cell cycle progression/meiosis	EGT2, RME1, SAP185, SAP190, SMK1, SPR3, SUM1, YOX1
Golgi/endosome/vacuole/sorting	BCH2, COG8, DRS2, ENT3, HUT1, PKR1, RIC1, RTC2, RTT10, SNX3, SWF1, VAC17, VAC7, VAM3, VPS24, VPS30, VPS38, VPS53, VPS70, VPS9, YPT6
Lipids	DGK1, DPL1, ENV9, FAT1, FEN1, INO2, IST2, LCB4, OPI3, PAH1, PSD1, SSP2, TGL4
Metabolism	AAC3, AAH1, ALD6, BNA3, CMC1, COQ2, ECM4, HOT13, HXT3, KGD1, MET22, MET5, MIG2, NRK1, PET191, QRI7, SDT1, SOD2, SPE3, THI21, VID30, YSA1

Ribosome/translation	ATE1, EAP1, MRN1, PUS6, RPL14A, RPL20B, RPL42A, RPS10A, RPS4A, RPS4B, RPS8A, TIF1, TSR3
Signalling/stress response	CKA2, CKB1, GCY1, GPA2, GRE1, NBP2, OPY2, PHO80, PTC1, SFL1, SHR5, SNF1, SNF4, TPK1, TUS1, WHI2, YAK1
Nuclear-cytoplasmic transport	HMT1, KAP122, MOG1, SLX9, YKL069W, YRA2
Protein degradation/proteasome	ATG10, CUE1, HLJ1, HRD1, HRD3, OLA1, RPN4, SPG5, UBC7, UBP11, UBP16, UBX2, UFD2, USA1, YUH1
Protein folding/glycosylation	ALG12, ALG3, ALG5, ALG6, ALG8, ALG9, ARV1, BST1, CPR4, CSF1, EOS1, ERD1, FKS1, GAS1, GTB1, INO1, KEX2, KRE1, LAS21, LHS1, LRG1, MPD1, OST3, OST4, PER1, PMT1, PMT2, RLM1, SCJ1, SPF1, SSE1, STE24, SUN4
RNA processing	BUD31, CBC2, DBR1, DXO1, ISY1, LSM1, LSM6, LSM7, MUD2, NOT3, PAT1, SKI8, VTS1
Unknown	ILM1, IRC8, KIN3, LCL2, NIF3, OSW5, SLP1, STB6, TDA4, YBR225W, YDR186C, YFR016C, YGL230C, YGR066C, YJR124C, YJR142W, YJR154W, YKL100C, YKR011C, YKR018C, YKR070W, YLL007C, YOR062C, YPL107W, YPL150W, YSY6

GO categories that showed enrichment with a FDR < 0.05 were obtained from Yeastmine (Balakrishnan et al., 2012) by comparison to a background population set consisting of all gene deletions present in the DMA. GO-Biological process terms that showed enrichment included protein transport (39 genes), macromolecule catabolic process (33 genes), RNA processing (24 genes), cellular lipid metabolic process (24 genes), ubiquitin-dependent protein catabolic process (20 genes), phospholipid metabolic process (16 genes), mRNA processing (15 genes), ER-associated ubiquitin-dependent protein catabolic process (13 genes), protein glycosylation (13 genes), regulation of response to stress (11 genes), dolichol-linked oligosaccharide biosynthetic process (5 genes), phosphatidylinositol biosynthetic process (5 genes) and GPI anchor metabolic process (4 genes).

GO-Cellular component terms that showed enrichment included endoplasmic reticulum (55 genes), endoplasmic reticulum membrane (43 genes), endomembrane system (75 genes), endoplasmic reticulum lumen (6 genes), ER ubiquitin ligase complex (5 genes), Hrd1p ubiquitin ligase complex (3 genes)

A number of publications including several large scale UPR studies were also enriched for genes in our hit list with a FDR < 0.05. These publications included Čopič et al., 2009; Jonikas et al., 2009; Promlek et al., 2011; Schuldiner et al., 2005 and some of our previous work (Bircham et al., 2011). Of particular note is Jonikas et al., 2009 who also used the same UPRE-GFP reporter as us to screen the DMA using flow cytometry. Comparison of our results to Jonikas et al. shows a great deal of overlap, with 40 % of our hits (98 of 244, Figure 16) also present in the Jonikas dataset. There are a number of differences between the studies, discussed later, that may account for some of the differences. Additionally it is important to note that the UPRE response is being measured on a continuous scale and thus the core overlap between studies is likely to represent gene deletions that strongly induce UPR, whereas other deletions may be more variable or less severe in response.

Comparison of our hits to known UPR target genes as defined by Kimata et al., 2006; and Travers et al., 2000 shows that a more limited subset of our hits, ~14 %, are UPR targets (33 of 244, Figure 16). This is comparable to the 13 % overlap of the Ire1p-GFP foci data (Figure 12) and the 16 % overlap when comparing the Jonikas UPRE dataset to known UPR targets.

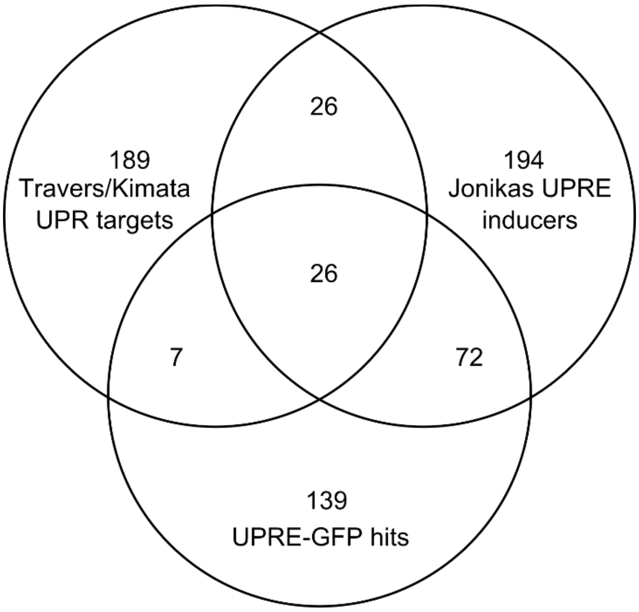


Figure 16 - UPRE-GFP inducer literature comparison

Overlap between gene deletions that induce UPRE-GFP expression and UPR targets as defined by Kimata et al., 2006 and Travers et al., 2000; and gene deletions that induced expression of a UPRE reporter as defined by Jonikas et al., 2009. Dubious ORFs and genes that were not present in our DMA collection were removed from the datasets before comparison.

4.4.6 Gene deletions that down-regulate UPRE activity

To identify gene deletions that showed a lower than expected level UPRE expression, further statistical analysis was conducted using R software (R Core Team, 2013). The average UPRE expression of cells from each deletion strain was again compared to the median of its corresponding plate as described above. A cut-off p-value threshold of 0.05 was used for strains with lower than the expected basal UPRE activation and gene deletions were considered a hit if their p-value fell below this. Genes were removed if they were annotated as dubious ORFs in SGD. Results are summarised in Table 9 and grouped in functional categories based on those presented in Costanzo et al., 2010:

Table 9 - Gene deletions that down-regulate UPRE activity

Gene deletions that down-regulate UPRE activity in unstressed conditions with a p-value less than 0.05. Functional categories were adapted from those used in Costanzo et al., 2010.

Functional Category	Genes
Cell polarity/morphogenesis	BEM2
Chromatin/transcription	BRE1, CHD1, DST1, LGE1, RRN10, RTF1, SDC1, SGF29, SIR2, SNF5, SOH1, SPT2, SPT3, SPT8, SWD1, SWD3, THP3
Chromosome segregation	LDB18
DNA replication	DPB3, TOP1
Cell cycle progression/meiosis	SWI4, SWI6
Golgi/endosome/vacuole/sorting	PIB2, RCY1, RIM8, SNC1, VAM6
Lipids	HTD2, IPK1, ISC1, MOT3, SUR2
Metabolism	CEM1, COX12, COX7, GSH2, IDH2, MDM12, MRM2, OAR1, PHO5, RIM1, RPO41, AIM22
Ribosome/translation	MAF1, MRT4, PUF6, REI1, RPL35B, RPL37A, RPL37B, RPP2B, RPS9B, SSF1, SWS2, SSD1
Nuclear-cytoplasmic transport	SXM1
Protein degradation/proteasome	UBP6
Protein folding/glycosylation	EXG1, HAC1, IRE1, MNN10, NAB6
RNA processing	MUD1, PSP2, SAC3, XRN1
Unknown	YPR117W

GO enrichments were obtained from Yeastmine (Balakrishnan et al., 2012) with FDR <0.05. GO-Biological process showed enrichment in a number of chromatin and histone processes including but not limited to: regulation of transcription, DNA-templated (26 genes), chromatin organization (15 genes), histone modification (10 genes), chromatin silencing (8 genes), histone methylation (5 genes) and histone ubiquitination (4 genes).

GO-Cellular component terms that showed enrichment included nucleus (35 genes), nucleolus (9 genes), histone acetyltransferase complex and SAGA complex (4 genes), histone methyltransferase complex and Set1C/COMPASS complex (3 genes)

4.4.7 Gene deletions that suppress UPRE-GFP induction

To identify genes that are required for UPR activation, the UPRE-GFP reporter DMA was screened under ER-stress conditions in triplicate. Cells were grown in 50 µL SC containing 2 mM DTT for four hours at 30°C to induce ER-stress, as these conditions induce the highest levels of UPRE activation and Ire1p foci formation (Figure 11 and Figure 15). Cells were screened with the Opera confocal microscope using standard conditions, and images were analysed using Acapella scripts for cell recognition and intensity measurements followed by statistical analysis. UPRE induction was measured as Log_2(GFP/RFP) as described above. Hits were defined as gene deletions strains showing a lower UPRE activation than expected from the plate median with a p-value less than 0.05. Dubious ORFs and ORFs that were missing in two of the three replicates were removed. Results are listed in Table 10 and grouped into functional categories based on those presented in Costanzo et al., 2010:

Table 10 - Gene deletions that suppress UPRE-GFP induction upon ER-stress

Gene deletions that prevent UPRE-GFP expression upon ER-stress with a p-value less than 0.05. Functional categories were adapted from those used in Costanzo et al., 2010.

Functional Category	Genes
Amino acids/Nitrogen utilisation	CPA2, MET18, SLM4, THR1, THR4, YOR302W
Cell polarity/morphogenesis	BEM1, BEM2, BNI1, HBT1, RVS161, SHE4
Chromatin/transcription	BRE1, CHD1, CUP2, DEP1, EAF1, ELF1, RTF1, SDC1, SGF29, SNF5, SPT2, SPT3, SPT8, SUB1, SWD3, SWI4
Chromosome segregation	CIN8, SRC1
DNA replication	CTF18, RTT109, SLX5, SLX8, XRS2
Drug/ion transport	FPS1, PHO88, PMA2, PMR1, ZRC1

ER-Golgi traffic	CPR7, GET2
Golgi/endosome/vacuole/sorting	CCZ1, DID2, MON1, RAV1, RCY1, RIM8, VAC8, VPS41
Lipids	CSG2
Metabolism/mitochondria	AMD1, ATP11, BCS1, BNA6, COA4, PKP1, RIM1, RPO41, TPS2, XYL2
Ribosome/translation	AEP2, ANB1, DBP7, MAF1, MRPL10, RPL37A, RPP2B, RPS10A, RPS21B, RSM25, SSD1, TMA23
Signalling/stress response	BCK2, GPB2, GPR1, PRX1, SKN7, SNF4
Nuclear-cytoplasmic transport	SXM1, YKL069W
Protein degradation/proteasome	RPN10, UBP6
Protein folding/glycosylation	HAC1, IRE1, LAS21, MNN10, OST4, PAC10
RNA processing	LRP1, NAM7, NMD2, PUF3, RRP6
Unknown	ECM34, ECM8

GO enrichments were obtained from Yeastmine (Balakrishnan et al., 2012) with FDR <0.05. GO-Biological process categories that showed enrichment encompassed a substantial number of chromatin and transcription processes including; gene expression (46 genes), regulation of DNA-templated transcription (25 genes), chromosome organization (22 genes), chromatin organization (14 genes) and histone modification (10 genes). Others included; cellular response to stress (22 genes), membrane docking (4 genes) and vesicle docking (3 genes).

GO-Cellular component categories that showed enrichment included the histone acetyltransferase complex (5 genes), SAGA complex (4 genes), Mon1-Ccz1 complex (2 genes), and SUMO-targeted ubiquitin ligase complex (2 genes).

Comparison of the UPRE suppressors to the gene deletions that down regulate unstressed UPRE expression shows an overlap of 25 genes – around 26 % of the suppressors and 36 % of down-regulators. These 25 genes show enrichment for GO terms including gene expression (21 genes), DNA-templated transcription (15 genes), chromatin organisation (10 genes), covalent chromatin modification (8 genes) and histone modification (8 genes).

Comparison to the UPRE down-regulators in the Jonikas dataset show similar overlap numbers (Figure 17) but only 8 genes overlapping between all three screens; HAC1, IRE1, MAF1, RPL37A, RPP2B, SGF29, SPT3 and SPT8. Notably these genes include the two central UPR components IRE1 and HAC1, without which UPR activation is not possible.

SGF29, SPT3 and SPT8 are all subunits of the SAGA histone acetyltransferase complex that regulates transcription of some RNA polymerase II-dependent genes. SGF29 is a component of the histone acetyltransferase core module of the SAGA complex, although the other member of this complex was not present in our DMA. MAF1 is a negative regulator of RNA polymerase III and is involved in tRNA processing and stability. RPL37A is a ribosomal 60s subunit protein and RPP2B is a ribosomal protein involved in the interaction of translational elongation factors and the ribosome.

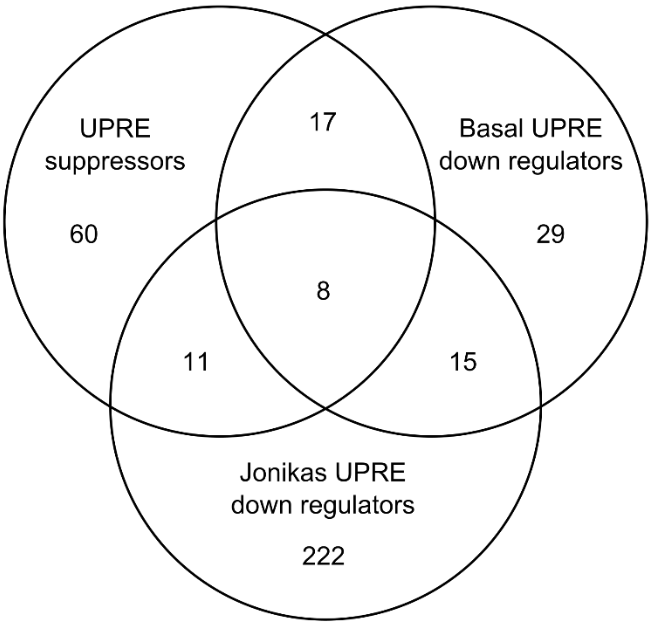


Figure 17 - Comparison of UPRE suppressors and down regulators

Overlap between gene deletions that suppress UPRE-GFP expression under ER-stress conditions and gene deletions that down regulate expression of a UPRE reporter as defined by Jonikas et al., 2009. Dubious ORFs and genes that were not present in the DMA were removed from the datasets before comparison.

4.5 Discussion

Almost 80% of yeast genes can individually be knocked out of the genome without causing an obvious growth defect, leading to the notion that measuring specific reporter phenotypes in gene deletion strains may be more revealing than simple growth measurements alone. Indeed this is the case for a number of ER-function related phenotypes shown by ourselves and others (Bircham et al., 2011; Čopič et al., 2009;

Jonikas et al., 2009). The fact that cells can adapt to the diverse impairments caused by gene knockouts that cause phenotypic consequences, is testament to the proficiency of homeostasis systems even in simple organisms such as yeast. Here we have investigated the phenotypic consequences of single gene deletions on UPR function both in standard growth and elevated ER-stress conditions. Specifically we have used reporters for function of the two UPR signalling components Ire1p and Hac1p. The use of automated microscopy to measure foci formation of Ire1p-GFP is a novel approach, allowing us to assess genes involved in early misfolded protein sensing events. This was complemented through monitoring the expression of Hac1p responsive UPRE-GFP which allowed for the elucidation of genes involved in downstream UPR activation.

There is substantial overlap between our result utilising UPRE-GFP and those of previous genome-wide studies assessing UPR function. Of particular note is the Jonikas et al., 2009 study which also used a UPRE-GFP reporter to screen the DMA. Comparing the gene deletions that increase basal UPRE expression between our datasets showed that only 40% of our hits were accounted for in their study. There are a number of differences between the two studies that could account for the discrepancies between the studies. Firstly our study used confocal microscopy to measure the reporter whereas Jonikas et al. used flow cytometry. Both techniques have their merits. Confocal microscopy is considered more sensitive and has the benefit of being able to visually inspect cells. Flow cytometry on the other hand measures thousands of cells as opposed to hundreds in microscopy, giving more statistical power to the measurements. Secondly, our study transferred yeast from overnight agar cultures into fresh SC media prior to imaging, whereas Jonikas et al. grew cells to saturation in a small amount of liquid YPD media and back-diluted the cultures with fresh media prior to screening. Both studies allowed cells to grow over a similar timeframe, however the differences in media change (liquid to liquid vs. solid to liquid), and the use of rich media (YPD) as opposed to SC likely caused some of the discrepancies. Indeed deletion strains are known to have different responses between rich and minimal growth media (Giaever et al., 2002; Hillenmeyer et al., 2008; Newman et al., 2006; Winzeler et al., 1999). Finally differences in statistics and cut-off limits for hit detection are likely to have caused differences between the screens, we used a relatively constricting cut-off identifying 244 UPRE inducing gene deletions, and by comparison

Jonikas et al. identified almost 400 hits. Given these differences we consider the overlap between the screens to be quite substantial and likely to represent the consistent set of 'core' gene deletion hits. As a case in point, all of the top 20 hits in the Jonikas et al. screen were found in our study, and 70% of their top 50. Conversely 76% of our top 50 were in the Jonikas screen, showing that as hit cut-offs are lowered more discrepancies between the studies arise. Beyond these comparisons, the two studies diverge on the use of UPRE-GFP as a reporter, Jonikas et al. created a series of double deletion strains looking for epistatic effects on UPR induction, whereas we have used DTT as an ER-stress and UPR inducing agent to identify gene deletions supressing activation of the UPR.

Another approach is that of Čopič et al., 2009, in which the DMA was screened for defects in the retention of an ER chaperone, Kar2p. This study identified a total of 87 mutants that caused secretion of Kar2p, 73 of which were non-essential genes. Of these, 14 hits showed an overlap with our UPRE-GFP inducing gene deletions, 13 of which were also identified by Jonikas et al. Notably 12 of the 14 genes are ER localised proteins and included functions in protein glycosylation, ER quality control, protein folding, GPI-protein maturation and ER-golgi transport. Inverse to the retention defect screen of Čopič et al. is previous work from out lab in Bircham et al., 2011 where we screened the DMA for defects in in the delivery of a plasma membrane protein, Mrh1p, to the cell surface. In this screen we identified 24 gene deletions that caused a partial blockage of Mrh1p delivery to the plasma membrane. Comparison of our UPR reporter results presented here identified an overlap of 11 genes between the two studies, 10 of which were UPR inducers identified by us or Jonikas et al. 2009. Comparison of these screens with different but ER-function specific reporters shows the extensive, non-growth defects caused by gene-deletions as well as highlighting the importance and power of combining multiple approaches to enrich for specific gene functions. Remarkably all of these studies identified a number of genes with unknown function, highlighting the fact that there are still many ER protein folding/quality control processes yet to be defined. Below we discuss the hits from our screens in relation to the category of reporter measurement i.e. induction or suppression caused by particular gene deletions.

4.5.1 UPR inducers

Comparing the 104 gene deletions that induced Ire1p-GFP cluster formation to the 244 gene deletions that induced UPRE-GFP expression revealed only a limited overlap of 15 genes. These overlapping genes are enriched for a sizeable number of ER-associated protein degradation (ERAD) and glycosylation related GO terms including but not limited to: protein maturation (7 genes), proteasome-mediated ubiquitin-dependent protein catabolic process (6 genes), protein glycosylation (5 genes) and endoplasmic reticulum unfolded protein response (3 genes). The Ire1p-GFP foci inducers and the UPRE-GFP inducers datasets both individually show similar ER-function/UPR GO enrichments as expected, demonstrating that both reporter methods have functioned effectively. The small overlap between the two datasets is indicative of the temporal action of the reporters (i.e. Ire1p foci formation occurs upstream of UPRE induction) rather than implying one method is superior over the other. Ire1p-GFP clustering for example is normally a transient phenotype (Ishiwata-Kimata et al., 2013a) and here we measured the proportion of cells in the population with this phenotype at a given time. UPRE-GFP expression on the other hand is a more prolonged component of the UPR, and in this case we measured the average UPR activation of all cells. With this in mind it seems possible that Ire1p-GFP clustering may have identified gene deletions that cause a transient misfolding of protein. Since there is only ever a proportion of cells with this phenotype it would be interesting in future studies to measure if any correlation exists between cell-cycle stage and the effect of the gene deletion hits identified.

It is interesting to note that of the Ire1p cluster inducers and UPRE inducers only 13% and 14% of their respective hits were UPR targets as defined in the literature or by us (see Chapter 5). By contrast almost half of the overlapping hits are UPR target genes, specifically they are HRD1, LAS21, PMT1, PMT2, STE24 and UBX2. Four of these proteins have ERAD functions, HRD1 is the ubiquitin ligase of the Hrd1p complex, UBX2 is a bridging factor associated with the Hrd1p complex, and PMT1/PMT2 forms an ER membrane Protein O-mannosyltransferase complex involved in ER quality control (see Chapter 1 for details on these genes).

Of all the UPR inducing gene deletions, whether Ire1p foci forming or UPRE activating, the two standout functional categories are glycosylation and ERAD/protein degradation with

15 and 16 genes in each category respectively. Specifically the glycosylation related genes are: ALG3, ALG5, ALG6, ALG8, ALG9, ALG12, EOS1, ERD1, GDA1, GTB1, OST3, OST4, OST5, PMT1 and PMT2, and the ERAD related genes are BST1, CUE1, DFM1, HLJ1, HRD1, HRD3, LCL2, PMT1, PMT2, SCJ1, UBC7, UBR1, UBX2, UBX4, UFD2 and USA1. It is understandable that deletions of genes with these functions leads to an activation of UPR as protein glycosylation is a critical component of protein folding and logically defects in this would lead to an increased rate of protein misfolding. Conversely ERAD is essential for the removal and degradation of terminally misfolded proteins, thus defects in ERAD would lead to the accumulation of naturally occurring misfolded proteins.

4.5.2 UPR suppressors

Given that Ire1p cluster formation is the result of high order oligomerisation through specific domain interactions rather than a mere consequence of protein aggregation (Kimata et al., 2007), its is reasonable to hypothesize that additional proteins may be involved in this conformation event. Indeed it was recently shown that actin filaments and type-II myosin are required for efficient Ire1p cluster formation (Ishiwata-Kimata et al., 2013b). It is interesting to note that here we found very few gene deletions that supressed the formation of Ire1p foci upon exposure to ER-stress (Table 7), indicating that this process may not require additional interaction partners outside of Ire1p and misfolded proteins themselves. The simplest explanation for hits identified here would be that the gene deletions themselves have caused a small induction of UPR stress and ‘primed’ the cells such that Ire1p no longer clusters in response to additional ER stress. Indeed comparing these foci inhibiting hits to the list of gene deletions that induce a chronic UPR activation (Table 8), we show that of the 13 gene deletions that prevent Ire1p clustering, 10 deletions (Δalg8, Δdbf2, Δlhs1, Δopi3, Δpsd1, Δrpl14a, Δrpl42a, Δrps4a, Δvam3 and Δzrc1) already showed UPR activity before the addition of DTT stress. Ire1p clustering has been demonstrated previously to be a short-term response to ER-stress and is not maintained in long-term stress (Ishiwata-Kimata et al., 2013a). In these cases Ire1p still exists as an activated homo-dimer and a lower level of continued UPR activity is maintained. This suggests that the cluster inhibitory effect of these 10 gene deletions is likely a secondary effect from an already active UPR. Of the three deletions that did not themselves induce UPR activity, only Δpsp2 caused a down-regulation of basal UPR expression while Δmrpl10 and Δzrc1 caused a repression of UPRE expression upon addition of external ER-stress. Both Δmrpl10 and Δpsp2 are known to be sensitive to DTT treatment (Fernandez-Ricaud et al., 2005, 2007). The Ire1p-GFP and UPRE-GFP data suggest that sensitivity to DTT stress is likely caused by the lack of a functional UPR in these strains.

The effects of ZRC1 deletion are intriguing as it both increased the number of foci beyond expected in untreated cells, and decreased the number in treated cells. By itself this could be overlooked, however the same effect was measured in the UPRE-GFP expression screens suggesting this is not an artefact. Zrc1p is a zinc transporter that transports zinc from the cytoplasm to the vacuole for detoxification and storage, and also functions to maintain ER zinc levels (Ellis et al., 2004; MacDiarmid et al., 2002). Surprisingly, the deletion of the Zrc1p paralog Cot1p showed no UPR effects in our screens. The exact mechanism behind these contrasting UPR phenotypes remains to be determined, however it may be due to unbalanced zinc homeostasis in the ER causing basal as well as ER-stress induced UPR defects. Zinc is well known to have catalytic and structural roles in almost 400 yeast proteins (Regalla and Lyons, 2006), additionally recent studies have shown that zinc deficiency can induce the UPR and is required for normal secretory pathway function (Ellis et al., 2004, 2005). Furthermore, we also found deletion of the vacuole to cytoplasm zinc transporter, ZRT3, to increase basal Ire1p-GFP clustering similar to Δzrc1. Additionally, deletions of ICE2, a gene required for ER zinc homeostasis, and ZRG17, an ER zinc transporter, also caused an increase in basal UPRE-GFP expression. Moreover, it is noteworthy that 11 of the deletions increasing basal Ire1p-GFP clustering or UPRE-GFP expression have a zinc requirement and/or zinc related functions. Taken together these data suggest that although zinc homeostasis is required for proper ER-function, these genes are unlikely to be specifically required for UPR activation as only deletion of ZRC1 had a suppressing effect on the UPR in the presence of additional ER-stress. Additionally, in the case of Δzrc1 it was not a complete blockage of UPR signalling but rather an attenuation of the response, whereas for the other zinc transporter related gene deletions only the basal UPR was affected suggesting the effect was due to an increase in protein misfolding.

It is interesting that only two gene deletions (Δmrpl10 and Δzrc1) caused both suppression of Ire1p clustering and UPRE induction under ER-stress conditions. This implies that the other 94 gene deletions that were found to inhibit ER-stress induced UPR, are not a consequence of failure to initially detect misfolded proteins. It is perhaps unsurprising then that most of the UPRE suppressor hits only caused a partial attenuation of the UPR response (see supplementary Table 1). In fact the only gene deletion that completely blocked UPR signalling to Δhac1 levels was a dubious ORF that partially overlapped with the HAC1 reading frame. Only a dozen gene deletions in the presence of DTT reduced UPRE expression to less than half what was expected, specifically: Δbem2, Δcpa2, Δhac1, Δire1, Δost4, Δpac10, Δrcy1, Δrps10a, Δthr1, Δthr4, Δubp6 and Δyor302w. Expectedly Δire1 and Δhac1 completely negated UPRE induction. Δbem2, Δrcy1 and Δubp6 also caused a decrease in basal UPRE expression. Both Δost4 and Δrps10a actually caused an increase in basal UPRE expression. Some of the other gene deletions are less interpretable, for example THR1 and THR4 for instance are genes required for threonine biosynthesis.

Functional enrichments within the hits that suppress ER-stress induced UPRE expression were concentrated around functions in gene expression, transcription and chromatin/histone modification. In fact there were 10 hits involved in chromatin modification and 5 hits involved in the histone acetyltransferase (HAT) complex. Four of these genes CHD1, SGF29, SPT3 and SPT8 are part of the SAGA complex responsible for histone acetylation during transcriptional activation, which has previously been shown to play a role in the UPR (Welihinda et al., 1997, 2000). It has been suggested previously that SAGA related effects could possibly be due to a general decrease in transcription reducing the protein load on the ER thus reducing the activation of UPR prior to IRE1 or HAC1 (Urano et al., 2000). This remains a possibility, although we believe this is unlikely as the SAGA components seen in our results also down-regulate the basal UPRE expression prior to any additional ER-stress and SGF29, a component of the HAT/Core module of the SAGA complex also increased the basal Ire1p cluster formation suggesting an increase rather than a decrease in ER protein load.

Another interesting enrichment was the two genes of the Mon1-Ccz1 complex. Mon1p and Ccz1p are required in autophagy and in the fusion of autophagosomes and vesicles with the vacuole (Meiling-Wesse et al., 2002; Wang, 2002). ER-stress is known to induce autophagy, and this has been shown to counterbalance the ER membrane expansion associated with ER-stress and UPR induction (Bernales et al., 2006; Yorimitsu et al., 2006). Why deletion of these genes, which result in an absence of autophagy (Kanki et al., 2009), results in an attenuation of UPRE expression in DTT treatment is unclear. Possibly in the short term ER-expansion alone is beneficial to dealing with ER-stress, lowering UPRE expression. Supporting this possibility, deletions of MON1 have been shown to be resistant to DTT stress (Kim et al., 2012a). Thus it may only be in the long term, or upon resolution of ER-stress that autophagy is required to compensate for the no longer necessary ER expansion.

As a final example of hit enrichment, the UPRE suppressor hits also included both members of the Slx5-Slx8 SUMO-targeted ubiquitin ligase complex (Xie et al., 2007). Small ubiquitin-like modifier (SUMO) proteins are covalently attached to a number of proteins to modify their function as a form of post-translational control. One of the functions of SUMOylation is to direct proteins for ubiquitinylation by the Slx5-Slx8 complex (Uzunova et al., 2007). Interestingly one of the gene deletions that induced basal UPRE expression was Δsnf1. Snf1p is also a UPR target gene, and is known to be regulated by SUMOylation which inhibits its function and ultimately targets Snf1p for destruction by the Slx5-Slx8 ubiquitin ligase (Simpson-Lavy and Johnston, 2013). This suggests that the deletions of SLX1 and SLX5 may actually attenuate UPR induction by supressing the degradation of UPR target genes, specifically Snf1p, although there may be other UPR targets that are modulated by SUMOylation.

4.6 Conclusion

In this chapter we describe the use of the yeast DMA – a collection of non-essential knock-out strains – to screen for gene deletions affecting UPR function. We describe the application of the image recognition scripts described in Chapter 3 to measure two reporters to specifically measure different phases of the UPR on a per cell basis. Firstly Ire1p-GFP foci formation was used as a phenotypic reporter for UPR initiation, and secondly we used the UPR/Hac1p responsive UPRE-GFP reporter to measure downstream UPR activity. Although the gene deletion strains of the DMA show no obvious growth consequences, here we show that many of the deletions affect basal UPR activity and additionally identify genes that are required for UPR activation in conditions of additional ER-stress. The Ire1p-GFP screen identified 104 gene deletions that induced Ire1p cluster formation under normal growth conditions, and 13 gene deletions that prevented the expected foci formation under conditions of induced ER-stress. The UPRE-GFP screens elicited a greater number of hits, identifying 69 gene deletions that down-regulated basal UPRE expression, 244 gene deletions that induced basal UPRE expression and 96 gene deletions that suppressed the expected UPRE induction under conditions of ER-stress. Cross examination of these results showed functional connections between UPR inducing and UPR suppressing hits that explain the nature of these responses. Comparison with the literature showed a good correspondence between our results and similar DMA studies assessing UPR function. Additionally our data provided new information not seen in the literature for a number of genes affecting the UPR. The studies presented here also show the versatility of automated microscopy for genome-wide screens, particularly in identifying subtle phenotypes that are not apparent based on cell viability alone.

5 LOCALISATION PROTEOMICS OF ER-STRESS

5.1 Background

Yeast are very malleable to a variety of stress conditions. Stress responses such as the UPR allow cells to adapt in a variety of ways, most observable of which is a change in the cellular milieu of proteins available to deal with the particular stress. Such changes may be initiated by an increase in protein expression or a change of protein localisation, both of which can be readily assayed using high-throughput fluorescence microscopy. The yeast GFP collection has accelerated the development of localisation proteomics and 'phenomics' in *Saccharomyces cerevisiae*, which has been growing in popularity as a screening technique (Breker et al., 2013; Dénervaud et al., 2013; Mazumder et al., 2013; Tkach et al., 2012). In this chapter we describe the use of high-throughput fluorescence microscopy to characterise proteome-wide changes induced by UPR activation with the ER-stress agents DTT and tunicamycin (TM).

DTT and TM are both well-established ER-stress inducers (see Chapter One). Chemical induction of the UPR is known to have several phenotypic consequences, including up-regulation of UPR target genes (Kimata et al., 2006; Travers et al., 2000), ER-expansion (Bernales et al., 2006) and Ire1p foci formation (Kimata et al., 2007). Previous genome-wide screens for UPR adaptations have used cDNA microarrays to assess regulation at an mRNA level. Travers et al., (2000) used both DTT and TM to induce ER-stress and defined UPR target genes. This has been the primary resource for UPR target genes in the literature, and has since been followed up by Kimata et al., (2006) who used TM and a constitutively active UPR strain to further define UPR target genes. Although these studies have provided a wealth of knowledge, it is well known that changes in mRNA levels do not always correspond to a change in protein level and in some cases protein level changes are not activated at the transcript level (Gygi et al., 1999). Here we sought to supplement current knowledge of UPR targets by assessing proteome-wide UPR changes at the protein level using the yeast GFP collection. By assessing the fluorescence distribution of GFP-proteins, we are also able to include UPR consequences that are actuated through a change in protein localisation rather than expression.

One of the major challenges in high-content screening is the scoring of localisation changes. Because localisation patterns show so much variation between proteins and can change unexpectedly, automated scoring of localisation phenotypes can be a difficult task. The most common approach has simply been to visually screen and annotate localisation changes by eye (Breker et al., 2013; Tkach et al., 2012). Another approach has been to measure specific changes of interest, for instance nuclear-cytoplasmic translocations (Mazumder et al., 2013). Here we present an approach that uses the speed of automated texture analysis to enrich for a set of 'likely' localisation changes followed up by visual inspection.

The aims of this chapter are as follows:

1. Introduce the dual RFP labelling system from Chapter Three into the yeast GFP collection to enable genome wide screening of protein changes by automated image analysis.

2. Screen the GFP collection using automated microscopy and image analysis to characterise the proteome-wide changes in protein expression and localisation induced by the ER stress agents DTT and tunicamycin.

3. Define UPR specific changes by comparing to DTT induced GFP changes in a Δhac1 UPR deficient GFP collection

4. Assess the efficacy of texture analysis as a means to automate screening of localisation changes

5.2 Screening for proteome changes induced by the UPR

5.2.1 Selection of optimal drug concentrations for ER-stress conditions

In order to screen the yeast GFP collection under UPR conditions we chose to use chemical induction of ER-stress with dithiothreitol (DTT) and tunicamycin (TM). Drug concentration was selected based on the optimal response of the UPRE-GFP and Ire1p-GFP reporters described in Chapter Four. Cells were transferred directly from 384 colony agar plates into a 384 well microtitre plate containing 50 µL SC + drug. Imaging was carried out as per standard screening conditions. Dose responses of DTT ranging from 0.125 - 8 mM showed optimal UPR activation and Ire1p foci formation at 2 mM DTT (Figure 18).

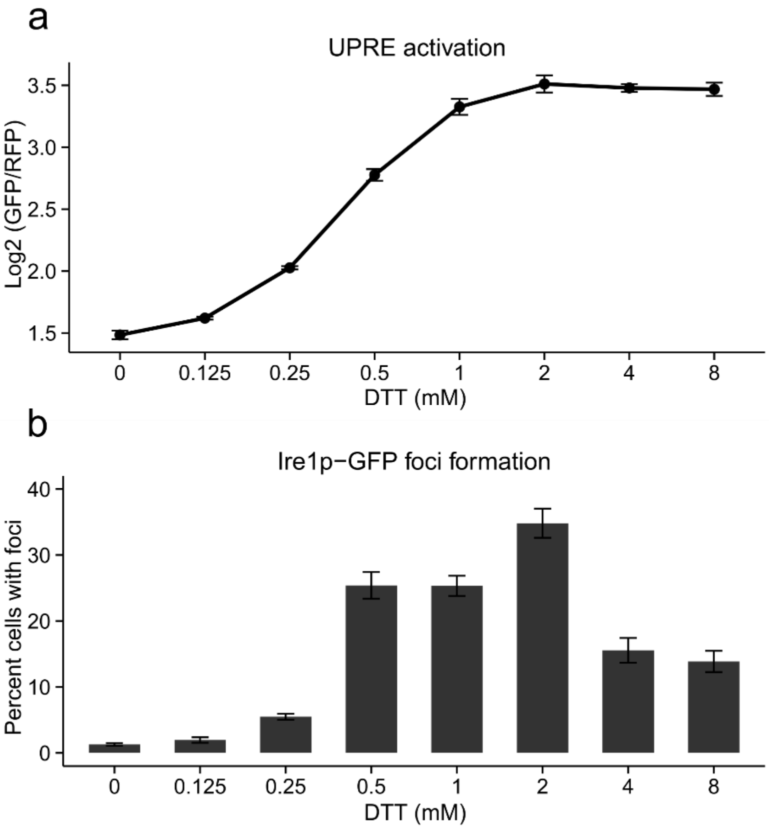


Figure 18 –UPR activation as a dose-response to Dithiothreitol treatment

UPR activation as a function of DTT concentration after 4 hours of treatment. a) UPR activation measured from the expression of a 4xUPRE-GFP reporter and normalised against the expression of a constitutively expressed RFP protein TEF2pr_mCherrry; b) Ire1p-GFP cluster formation presented as percentage of cell population displaying Ire1p foci.

Dose responses of TM ranging from 0.125 - µg/mL showed optimal UPR activation and Ire1p foci formation at 2 µg/mL TM (Figure 19). Visual inspection of the images showed healthy looking cells that were actively dividing, thus these drug concentrations were chosen for further screening. Although drug concentrations higher than this have a high UPRE-GFP activation, there was an increase in the number of dead cells. Ire1p-GFP foci counts also dropped off. Most likely as a result of less well defined foci at these concentrations which may have affected the image analysis.

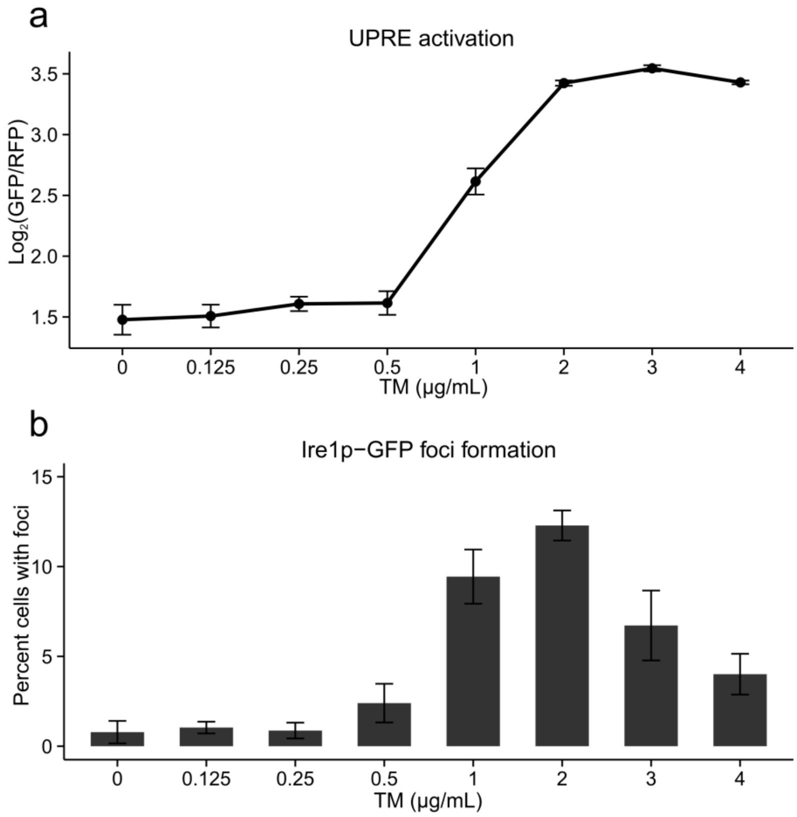


Figure 19 - UPR activation as a dose response to tunicamycin treatment

UPR activation as a function of TM concentration after 4 hours of treatment. a) UPR activation measured from the expression of a 4xUPRE-GFP reporter and normalised against the expression of a constitutively expressed RFP protein TEF2pr_mCherrry; b) Ire1p-GFP cluster formation presented as percentage of cell population displaying Ire1p foci.

5.2.2 Preparation and screening of the yeast GFP collection under UPR conditions

To enable automated image analysis of the yeast GFP collection, the dual cytoplasmic – nuclear RFP labelling system described in Chapter Three was introduced into the collection using standard SGA procedures (See Materials and Methods). Briefly, the *MATα* dual-RFP SGA starting strain yCG253 was mated to the 11 plates of the *MAT**a*** GFP collection in 384 colony format using the Singer RoToR HDA robot. Diploids were selected on SD-His + NAT media selecting for the GFP marker and mCherry RFP respectively. After sporulation *MAT**a*** haploid progeny were selected for on SD-Ura/Arg/Lys + CAN/THIA. This was followed by sequential selections for the GFP-fusion protein on media lacking histidine and the cytoplasmic and nuclear RFP on media containing nourseothricin and hygromycin B in that order.

The Opera microscope was used for high-throughput imaging of the collection using standard procedures. Cells were pinned directly from agar plates grown overnight into 384 well clear bottomed microtitre plates (cell carrier, Perkin Elmer) containing 50 µL SC ± drug. As tunicamycin is not water soluble DMSO was used as a carrier to a final concentration of 2% in both treated and control conditions. Cells were grown without shaking for 4 hours prior to imaging and transferred to the microscope 5 minutes before imaging to allow cells to settle and reduce motion blur when imaging. An exposure time of 400 ms was used for optimal GFP fluorescence across the highly variable protein levels in the collection. Treated and untreated cells were imaged in adjacent wells to minimise variation due to growth conditions and plate effects. Images were taken from two locations in each well to increase cell counts and to provide redundancy in the case of unfocused images.

5.3 Abundance changes across the proteome

5.3.1 Measurement of protein abundance

Protein abundance was assessed by automated image analysis using customised Acapella scripts described in Chapter Three. Changes in protein expression were measured by taking the total fluorescence intensity within the cell body and taking the median of all the individual cell measurements for each GFP-strain. The benefit of using total cell intensities is that changes to weakly expressed proteins that are localised to only a small portion of the cell can still be detected. An example is the Ire1-GFP strain used that is only weakly expressed but upon ER-stress forms discrete foci along the ER. As these foci are still relatively faint and only occupy a small portion of the cell, they would not be detected if the pixels were averaged instead. A possible problem when using the total cell intensity is that the measurements can become dependent on cell size. Indeed in an individual population of cells as measured per image there is a relatively strong correlation between cell size and total GFP fluorescence. A sampling of 100 cells per well for a randomly selected plate from the DTT treated screen showed an average R^2 value of 0.72. However this was not considered to be a significant problem as no noticeable change in cell size was apparent between control conditions and the treatments used, and median measurements were used to compare between populations. To test this we compared the difference in protein fluorescence against the difference in cell size between DTT treated and control conditions. No correlation was observed ($R^2 = 0.129$) indicating that changes in GFP expression were independent of cell size changes (Figure 20).

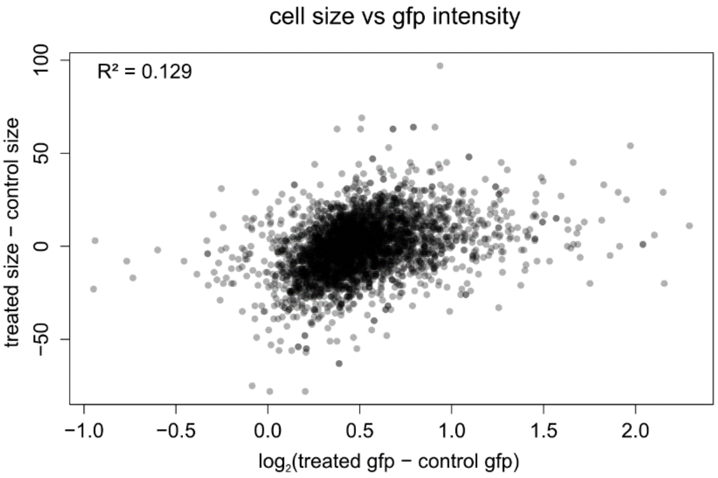


Figure 20 - Comparison between cell size and total GFP fluorescence after DTT treatment.

Comparing the difference in GFP expression levels (x axis) and difference in cell size (y axis) between treated and control conditions for each strain in the GFP collection shows no correlation between the two.

During the GFP collection screens we noticed a decrease in our 561 nm laser power used for RFP excitation. This resulted in a significant decrease in RFP fluorescence, although enough protein was visualised for the cell detection algorithms to run appropriately. Since GFP and RFP were simultaneously imaged we found that the low RFP expression led to cross-talk from high GFP fluorescence appearing in the fainter RFP channel. Because of this we chose not to use the RFP for normalisation in these screens.

5.3.2 Data normalisation and analysis:

GFP-protein intensity changes were measured as $\log_2$(treated/control). Expressing the ratio in Log_2 space better depicts the spread of data and its variation and also results in an even scale for an increase or decrease in intensity. Data analysis was conducted in R (R Core Team, 2013). Firstly each replicate screen was filtered to remove samples with a low cell count of ≤25 in either the control or treated samples. It has been reported that a large portion of proteins in the GFP collection are expressed at low levels making them effectively indistinguishable from auto-fluorescence (Breker et al., 2013; Dénervaud et al., 2013; Newman et al., 2006). Therefore, to remove strains likely to be below our detection threshold we estimated auto-fluorescence from the lowest abundance strain with greater than 150 recognised cells. The lower GFP detection limit was set as the estimated median auto-fluorescence intensity of this strain plus 2.5 times the median absolute deviation. Any strains with GFP fluorescence lower than this in both treated and control conditions were removed from the analysis. Replicate data were then filtered to remove data that were highly variable between the two replicates (Figure 21). The treated/control GFP ratio was calculated for each replicate. If the ratios are reproducible between replicates then $\log_2$(ratio replicate 1/ratio replicate 2) should be close to zero (Quackenbush, 2002). In Figure 21 we can see the plot of ratio1 against ratio2 follows closely to a slope of one, passing through zero as we would expect in reliable replicates. Outliers were defined as samples greater than 3 standard deviations away from the diagonal and removed from the analysis (highlighted red in Figure 21). Although if both replicates showed greater than 1.5 fold or less than 0.75 fold change i.e. the protein had considerably changed in expression across both replicates although to a greater degree in one replicate than the other these were included in the analysis (highlighted green in Figure 21).

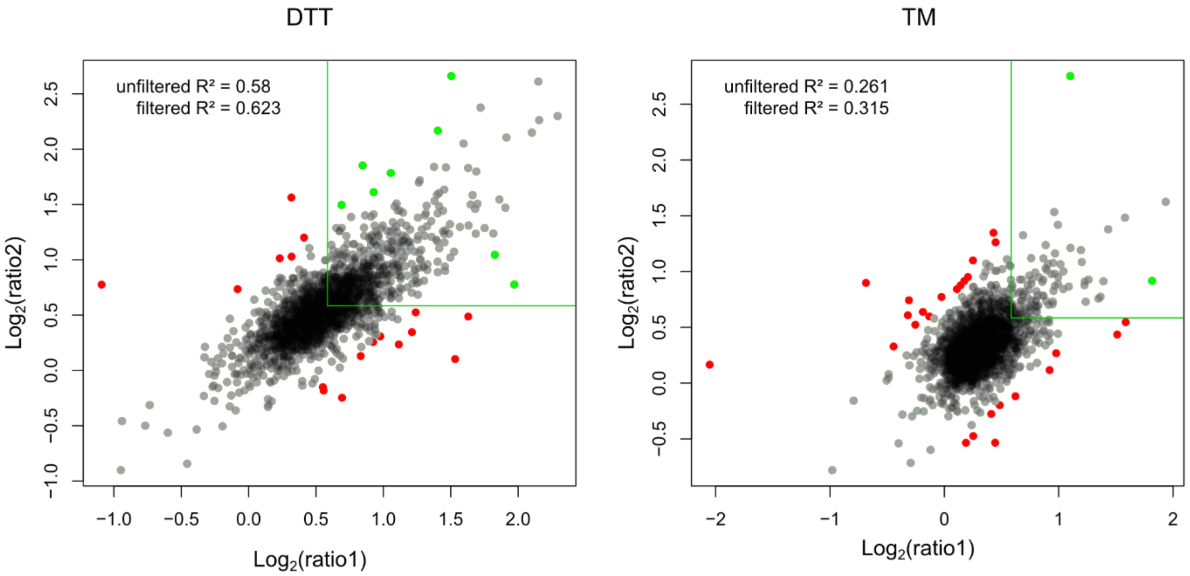


Figure 21 –Replicate filtering.

$\log_2$(treated/control) intensity ratios of replicates plotted against each other. The red dots highlight outliers that are 3 σ away from the diagonal and were not reproducibly up-regulated. These strains were considered inconsistent between replicates and were removed from the analysis. Green dots represent strains that are greater than 3 σ away from the diagonal but were kept in the analysis as both replicates increased in expression greater than 1.5 fold, as indicated by the green box.

Following data filtering the replicate screens were normalised in order to make their arbitrary fluorescent values directly comparable between screens. Since laser power decreases with use over time and between screens fluorescence intensity will decrease too. However, the units will keep their relative values within a given screen. To adjust for this quantile normalisation was used on both the control replicates and treated replicates (Bolstad et al., 2003) to adjust the replicates to have distributions with the same statistical properties (Figure 22). Briefly the two replicates are input into a matrix and their rank order noted. The replicates are then ordered by value and replaced by row means. For each replicate the new mean values are then rearranged to match the original rank order. This gives two matching distributions while maintaining the relative position of each gene within the population. To combine the replicates the mean was used to average values. Missing values were estimated from the normalised data from the other replicate.

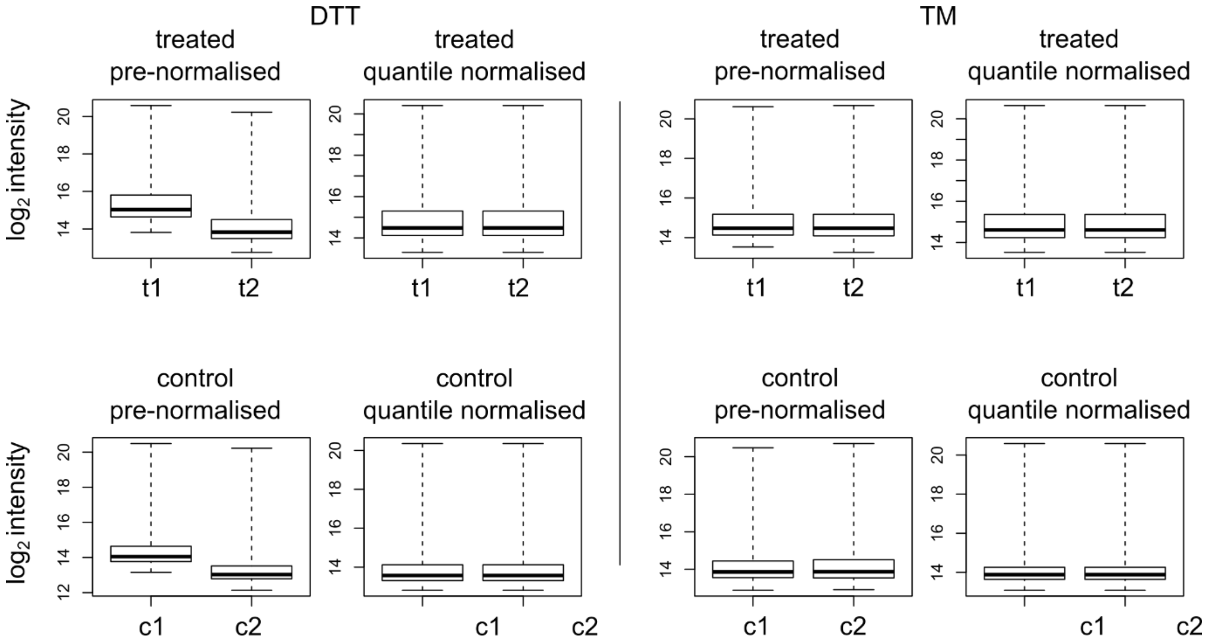


Figure 22 - boxplots showing control and treated GFP distributions across replicates before and after normalisation.

As can be seen in the DTT replicates an overall decrease in intensity due to reduced laser power in the second replicate is adjusted for by quantile normalisation.

Intensity changes are commonly measured as the ratio of treated to control in log space (Figure 23, left panels). However these plots often give an overly high sense of agreement between conditions and may not show important trends in the data. To see if there was any intensity based effects a magnitude - average (MA) plot (Dudoit et al., 2002) was used to visualise the data (Figure 23, right panels). An MA-plot is a form of ratio-intensity plot analogous to a scaled 45 ° rotation of the control-treated plot and depicts the relationship between the magnitude of intensity change (M) as the log intensity ratio, and the average log intensity (A) calculated as:

$$M = log2\left(\frac{treated}{control}\right)$$

$$A = \frac{1}{2}\, log2(treated * control)$$

Although the treated-control plot (Figure 23, left panels) appears to show good concordance between the local regression line (red) and a linear model (green), when we visualise the same data on an MA plot it becomes apparent that there is a strong non-linear effect on the intensity ratio as average intensity increases (Figure 23, right panels).

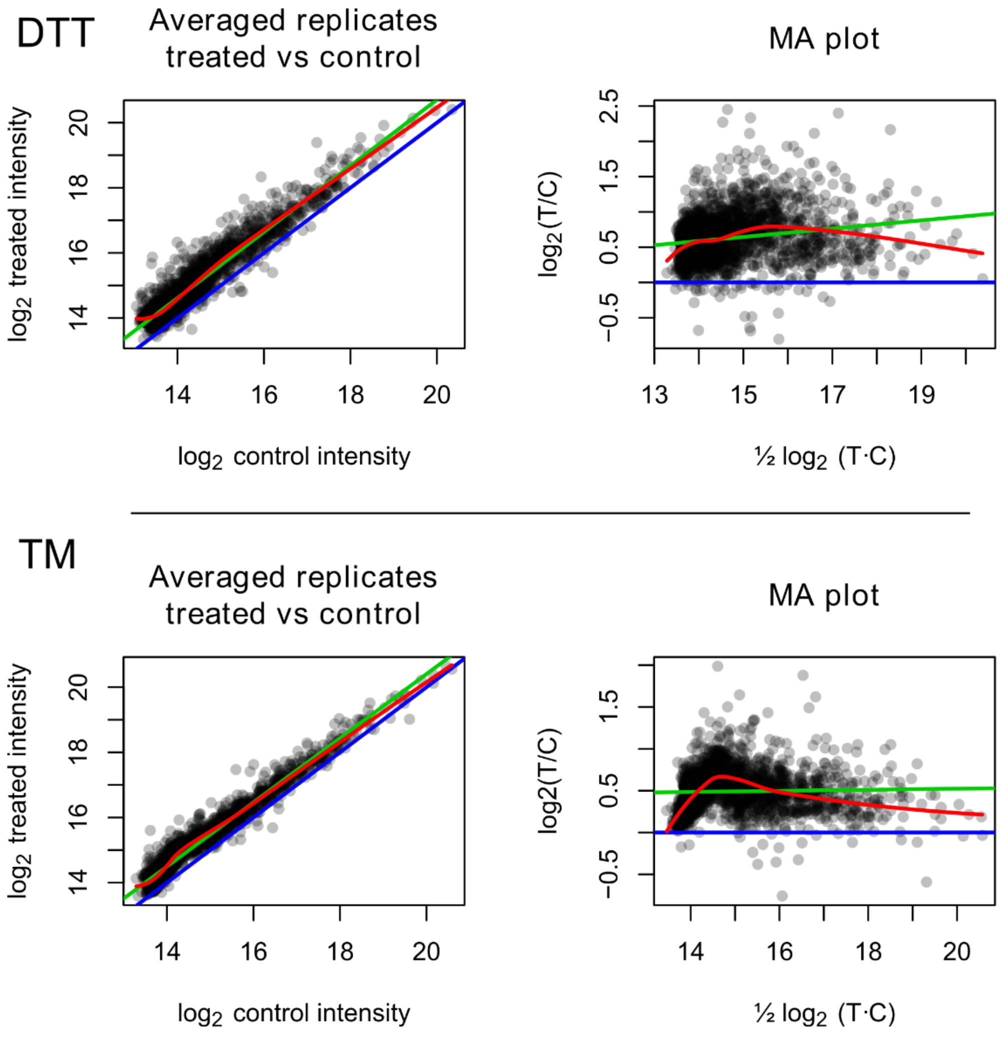


Figure 23 – Treated-Control intensity plots.

Panels on the left depict the averaged GFP intensities of treated against control for each protein. Panels on the right show MA plots – a 45° rotation of the left hand plots depicting the magnitude of expression change against the average. Green lines depict a linear model of the plot, blue lines depict a base line of no expression change and the red lines depict a local regression line. In plots on the left a general increase of protein expression in treated cells can be seen as the red and green lines have shifted upward from the zero change base line (blue). Although the linear model appears to closely match the local regression on the left, in the MA plots we can see an intensity based effect suggesting that low expression proteins behave differently from high expression proteins. As such the expression data were normalised to the red local regression lines in the MA plots for further analysis.

Because the GFP collection screen is essentially analogous to expression microarray data we employed local regression methods that are well established in the microarray literature to remove any intensity effects from the data (Colantuoni et al., 2002). To achieve this, the residuals calculated from the loess (LOcal regrESSion, Cleveland and Devlin, 1988) function in R were used to smooth the scatterplot and fit a non-linear local regression curve which the data were normalised to (Figure 24, left panels). The loess normalised data was then used to calculate local Z scores for increases and decreases in protein expression ratios (Figure 24, left panels - green lines and right panels). A similar methodology has previously been described for microarray expression data (Colantuoni et al., 2002).

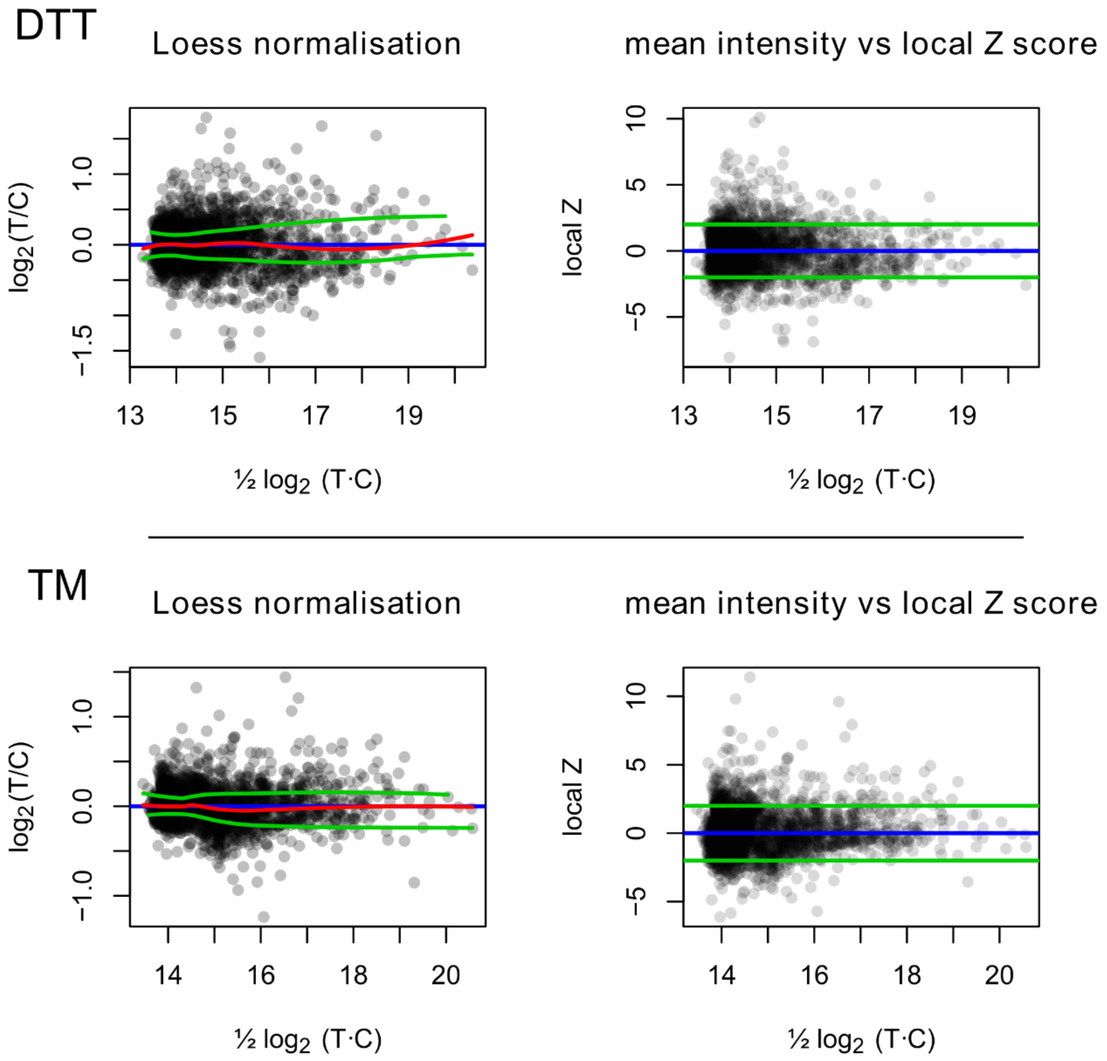


Figure 24 - Loess normalisation and local Z score calculation.

Plots on the left are the same MA plots as in Figure 23 after normalising to the Loess regression. As can be seen now the new local regression modelling (red) closely follows and is centred on the zero change base line (blue). Green lines represent local standard deviation measurement of 1σ. On the right hand panels the local Z-scores are plotted against the mean intensity. Green lines represent the hit cut-off Z-score of 2.

Proteins were considered to be significantly up-regulated in strains that showed a local Z-score ≥ 2 (Figure 25, highlighted green). Analysis revealed that very few proteins actually showed a significant decrease in protein abundance over the time-frame measured. This may be due to a number of factors. Firstly, GFP is known to be a relatively stable protein and one might not expect to see much of a protein decrease in this time frame unless proteins were specifically being targeted for degradation (Mateus and Avery, 2000). Drug treatment also caused a general increase in protein fluorescence, as can be seen in the plot of treated against control GFP expression in Figure 23. With this in mind we chose to measure repressed proteins as those that showed a lower than expected GFP fluorescence compared to proteins of similar expression levels with a local Z-score ≤ -2 (Figure 25, highlighted red).

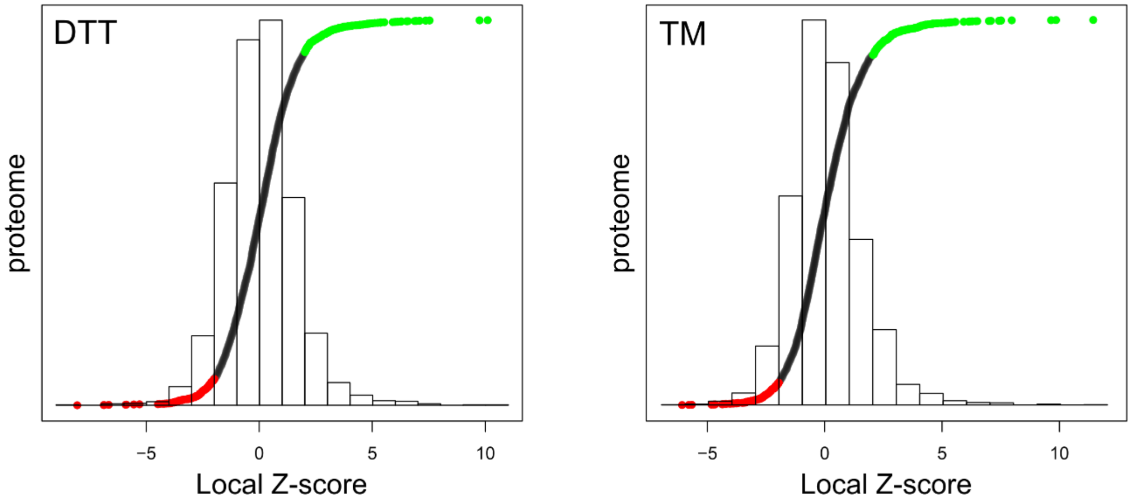


Figure 25 - Local Z-score for GFP expression changes.

Local Z-scores for proteins abundance changes across the proteome. Strains are ordered by rank. Up-regulated hits are highlighted in green, repressed proteins are highlighted in red. The histogram displays a density function of the plots.

Proteins were removed from the hit list if the GFP tag was likely to have affected localisation due to disruption of the C-terminus as defined in Breker et al., (2013). Hit lists of up-regulated proteins for DTT and TM treatments are listed in Table 11 and Table 12 respectively. Hit lists of repressed proteins for DTT and TM treatments are listed in Table

13 and Table 14. Gene ontology analysis was conducted using the online software tool YeastMine (Balakrishnan et al., 2012) by comparing the hit lists to a background set consisting of all proteins in the GFP collection. To correct for multiple testing the Benjamini Hochberg (False Discovery Rate) test correction was used and GO term enrichments were considered significant if they showed an FDR less than 0.05. Redundant GO terms were disregarded by comparison of GO term hierarchies using the web-based tool AmiGO (Carbon et al., 2009).

5.3.3 Proteins up-regulated under ER-stress conditions

Analysis identified 255 proteins that were up-regulated in expression in response to DTT (Table 11) and 217 up-regulated in response to TM (Table 12). An overlap of 80 proteins, roughly a third of each set, was observed between both treatments (Figure 26). These proteins are more likely to represent a 'core' UPR response, rather than any drug specific effect. This is likely to be a conservative estimate as statistical cut-offs are ultimately a subjective decision in data confidence and if we were to loosen the stringency we would likely find more overlap. Nonetheless it is curious that this core overlap of 80 proteins did not show enrichment for any particular GO ***biological process*** term, however GO ***cellular component*** showed enrichments for endoplasmic reticulum (23 proteins), endomembrane system (27 genes) and Sec62/Sec63 complex (3 out of 4 proteins). Additionally Yeastmine identified a publication involving a yeast GFP collection screen for DNA damaging agents (Tkach et al., 2012) as having significant enrichment for 22 of the 80 overlap proteins, this indicates that at least some of these proteins are likely to be involved in a more general stress response rather than being UPR specific.

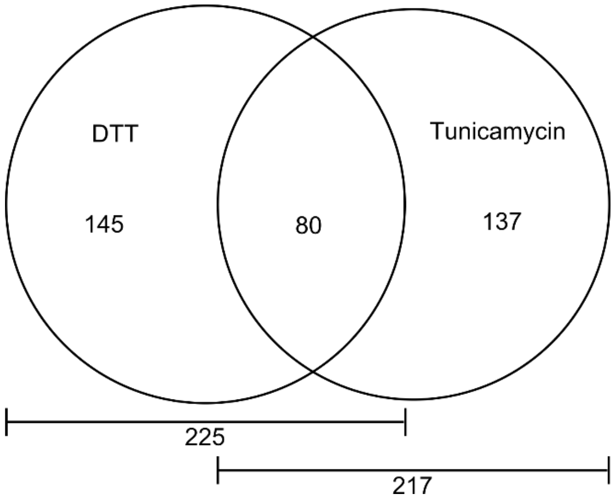


Figure 26 - Overlap between DTT and TM up-regulated proteins

Table 11 - Proteins up-regulated by DTT treatment.

Proteins that increase in abundance in response to 2 mM DTT treatment with a local Z-score ≥ 2. Functional categories were adapted from those used in Costanzo et al., 2010.

Functional Category	Genes
Amino acid biosynthesis	ARO10, HIS2, PAR32
Cell polarity/morphogenesis	ABP1, BUD5, CAP1, MYO3, PIL1, PIN3, RCR1, SAC6, STE50, SUR1, TWF1
Chromatin/transcription	SDS3, SPT2, TFC7, WTM1
DNA replication	DDR48, MHR1
Drug/ion transport	AST2, BSD2, CTR1, CUP2, ISU2, MEP1, MEP2, PDR5, PMC1, TPO4, ZRT1
ER-Golgi traffic	ERP1, ERP2, ERV29, RET2, SEC66, SEC72, SFB2, SLY1, SPC2, SRP14, TRX2, YIP3
Golgi/endosome/vacuole/sorting	APE1, BTS1, GDI1, GYP7, HSE1, IVY1, PBI2, PRC1, RUP1, SLA2, SNA3, VPS1, VPS68
Lipids	CSG2, DPP1, ERG10, ERG24, FMP45, LAP2, OPI3, OSH6, SAY1, SCS3, YDC1, YKL091C
Metabolism	ACB1, ACS1, ADH4, ALD4, ARI1, ATP2, ATP3, ATP7, AYR1, BNA1, CAR2, CBP3, COQ4, COQ6, CPR3, DAK1, DLD3, DOG2, ECM4, FBP26, GAD1, GCY1, GDB1, GLC3, GLK1, GLO1, GPD1, GPM2, GPP1, GPP2, GPX2, GRX1, GSY2, HEM1, HEM14, HEM15, HXK1, HXT2, KGD2, LSC1, LSC2, MAM3, MCT1, MDH2, MSP1, MTG1, NTH1, OCT1, OYE2, PET10, PNC1, POR1, PUT1, PYC1, QCR7, RBK1, RIB1, RIP1, RBK1,SDH4, SOL4, STF1, TAL1, TDH2, THI20, THI80, TSA2, TSL1, UBC8, UGA1, UGA2, URA10, XKS1, YAT2, YDL124W, YJL068C, YLR345W, YMR315W

Ribosome/translation	BFR1, MAF1, MRPL16, MRPL38, MRPL6, RSM23, SWS2, TMA17, YET1, YMR295C
Signalling/stress response	CMK1, CMK2, CMP2, GRE3, HSP30, MCA1, MRP8, PRM5, PRX1, PST2, RIM11, RTS3, TFS1, YBL055C
Protein degradation/proteasome	DFM1, HRD1, UBC7, UBC5, UBI4, URH1
Protein folding/glycosylation	ALG6, ERJ5, GET4, GUP1, GWT1, ORM2, PMI40, PUN1, SEC62, SHR3, SLT2, SSA4, SSE2, TDH1, YLR194C, YPS3
RNA processing	EDC2, GSP2, IGO1, KIN28, ROX3
Unknown	ADD37, COS10, COS6, COS8, CRP1, NCE102, PRM8, RAM2, RCN2, RDL1, RGI1, TDA1, UIP3, YBR085C-A, YBR287W, YET3, YHR097C, YHR138C, YIL108W, YKL151C, YML007C-A, YMR122W-A, YMR178W, YNK1, YOR289W, YOR385W, YPL107W, YPR127W, YRO2

The set of 225 proteins up-regulated by DTT treatment showed enrichment for GO ***biological process*** terms including; redox related terms oxidation-reduction process (46 genes) and oxidoreduction coenzyme metabolic process (11 genes), as well as some relatively broad carbohydrate and metabolism related terms including carbohydrate metabolic process (29 genes), carbohydrate catabolic process (16 genes) and generation of precursor metabolites and energy (23 genes). These hits were also enriched for the ***molecular Function*** GO term oxidoreductase activity (30 genes) and for the cellular component GO term endoplasmic reticulum (39 genes). A number of publications on stress responses were enriched for genes in this hit list including salinity, heat shock, DNA damage, ethanol, cell wall stress, and glycosylation deficiency (Alexandre et al., 2001; Boorsma et al., 2004; Cullen et al., 2006; Liu et al., 2007; Tkach et al., 2012; Wu and Li, 2008).

Table 12 - Proteins up-regulated by TM treatment

Proteins that increase in abundance in response to 2 µg/mL TM treatment with a local Z-score ≥ 2. Functional categories were adapted from those used in Costanzo et al., 2010.

Functional Category	Genes
Amino acid biosynthesis/transport	MEP3, HIS2, PAR32, TRP1
Cell polarity/morphogenesis	AXL2, CAP1, CHS3, PFY1, PIN3, PRR1, PXL1, RCR1, RHO2, SDS24, SKT5, SRV2, STE24, SUR1, YPT32
Cell cycle progression/meiosis	CDC14, CDC28, CDC37, FPR3, NBP1
Chromatin/transcription	CPR1, FRA1, HTZ1, IES3, IES4, MCM1, SDC1, TUP1
Chromosome segregation	MAD2, MCD1

DNA replication/repair	CDC9, MHR1, RAD52
Nuclear-cytoplasmic transport	POM33
Drug/ion transport	ARN1, BSD2, FLC1, KCH1, MEP1, MEP2, MUP1, NHX1, PMT3, SNA2, YPR003C, YVC1, ZRT1
ER-Golgi traffic	ERD2, ERP1, ERP2, ERP3, ERV25, RER1, RET2, SEC13, SEC28, SEC66, SEC72, SFB2, SFB3, SPC2, SVP26, TRS23, YIP3
Golgi/endosome/vacuole/sorting	ARF1, ARL1, BMH1, GYP7, PBI2, RUP1, SEC2, SNA3, TVP15, TVP18, VID22, VMA11, VMA21, VPS45, VPS52, VPS68
Lipids	DNF2, ERG24, FAR8, FMP45, GPT2, HMG2, ITR1, NCR1, OPI10, OPI3
Metabolism	ACB1, ADK1, ARH1, ASN2, CBP3, CDD1, COQ4, COR1, COX14, COX17, DFR1, DOG2, GPD1, GPP1, GPP2, GRX1, HEM13, HXT2, ICP55, MDM10, MIC17, MSP1, NCE103, NQM1, OYE2, PAM17, PET10, PGM2, PYK2, QCR2, RIB1, SDH4, SOL4, STF2, TAL1, TDH2, UBC8, YCF1, YDL124W, YLR345W, YMR315W
Ribosome/translation	BFR1, CCA1, MRPL36, MRPL38, NOB1, RSM25, SDO1, SGN1, SQS1, TAD2, TAN1, TIF1, TMA17, YET1, YMR295C
RNA processing	LSM2, POP5, RTC3, YSH1
Signalling/stress response	CMK2, HSP26, POG1, PRM5, PST2, RTS3, STE7, TSA1
Protein degradation/proteasome	ADD66, DFM1, NAS6, PRD1
Protein folding/glycosylation	GET4, GPI17, GWT1, HAC1, HSP104, HSP82, ORM2, PCM1, PUN1, SEC62, SHR3, SLT2, SPF1, SWP1, YLR194C, YPS3
Unknown	COS10, CSI2, GIS3, HNT2, MHO1, NCE102, NIF3, NIT3, PIN2, PMU1, RCN2, RDL1, TDA1, VBA4, YBR287W, YDR210W, YDR476C, YFR006W, YGL010W, YGR017W, YHR127W, YHR138C, YJR015W, YLR290C, YNR061C, YOL019W, YOR289W, YPL067C, YPL107W, ADD37, VID27, YBR085C-A, YCP4, YET3, YMR122W-A

For the 217 proteins up-regulated by TM treatment the only GO ***biological process*** term that showed enrichment with an FDR < 0.05 was ER to Golgi vesicle-mediated transport (15 genes). GO ***Cellular component*** enrichments included ER (48 genes), intrinsic component of membrane (83 genes), ER to Golgi transport vesicle (10 genes), Vesicle (17 genes), endomembrane system (66 genes), endoplasmic reticulum membrane (35 genes) and the Sec62/Sec63 complex (3 genes). Publications that showed enrichment included yeast genome-wide studies on DNA damage, cell wall, salt, glycosylation deficiency and heat shock (Boorsma et al., 2004; Cullen et al., 2006; Liu et al., 2007; Melamed et al., 2008; Tkach et al., 2012; Wu and Li, 2008).

5.3.4 Proteins repressed under ER-stress conditions

Analysis identified 174 proteins whose expression was repressed in DTT treatment (Table 13) and 147 in TM treatment (Table 14). An overlap of 44 proteins was observed between DTT and TM (Figure 27) which showed a GO biological process enrichment (FDR < 0.05) for ribosome biogenesis (15 proteins), ribonucleoprotein complex biogenesis (16 proteins) and ribosomal large subunit biogenesis (7 proteins).

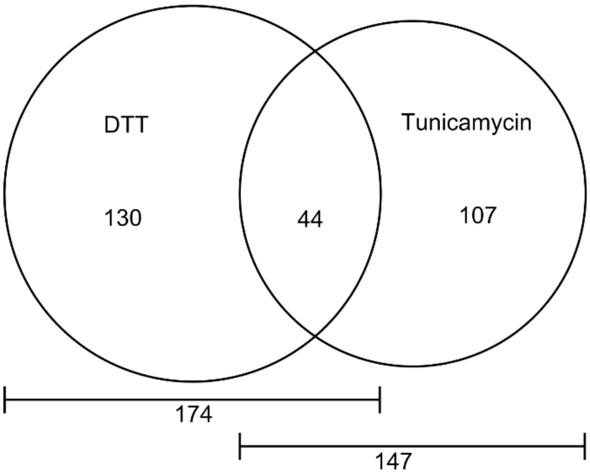


Figure 27 - Overlap between DTT and TM repressed proteins

Table 13 - Proteins repressed by DTT treatment.

Proteins that decrease in abundance or have lower than expected expression in response to 2 mM DTT treatment with a local Z-score ≤ -2. Functional categories were adapted from those used in Costanzo et al., 2010.

Functional Category	Genes
Amino acid biosynthesis	ADE6, APT1, ARG7, BAP3, CHA1, CPA1, CPA2, DIP5, HIS6, LEU9, LTV1, LYS2, TAT1, YOR302W, YPQ1
Cell polarity/morphogenesis	BZZ1, RAX2, SKM1, STE2, YPT31
Chromatin/transcription	HHF2, HTA2, RAP1, RPA34, SWI3, TAF3, ZPR1
Chromosome segregation	DAD1, SPC34
DNA replication	MKT1, POL5
Drug/ion transport	FET3, FTR1, PDR12, TNA1, TPO1, TPO3, VCX1
ER-Golgi traffic	ERP4, ERV14, ERV41, USO1
Golgi/endosome/vacuole/sorting	EMP70, VID22, VMA11, VMA16, VPH1

Lipids	CHO2, CPT1, ERG5, ERG11, FAA3, FEN1, HNM1, KEI1, SAH1, SUR2, YPR063C
Metabolism	AAH1, COX19, FSH1, GPM3, HNT1, HPT1, HXT3, HXT6, JJJ3, RPO41, SAM1, SAM3, SAM4, SPE1, TIM17, URA7, YHB1
Ribosome/translation	ALB1, BUD23, CBF5, DBP3, DPH2, DPH6, DRS1, EFG1, EFT1, ELP2, EMG1, ENP1, FCF2, IKI3, KAP123, MAK21, MDN1, MRPL24, NSA1, PRP38, PWP1, PWP2, PXR1, RBG1, RIA1, RMT2, RPL11A, RPL17A, RPL19B, RPL19A, RPL20B, RPL21A, RPL22B, RPL23A, RPL23B, RPL27A, RPL33B, RPL34B, RPL40A, RPL41A, RPL42B, RPL5, RPL7B, RPL8A, RPS10A, RPS14B, RPS17A, RPS18B, RPS19A, RPS22A, RPS21B, RPS23B, RPS24A, RPS25B, RPS6A, SRO9, SRP40, SSF1, SYO1, TIF11, TRM11, TRM12, TRM82, TSR2, UTP5
Signalling/stress response	GIS2, HMS2, SAP185
Nuclear-cytoplasmic transport	ARX1
Protein degradation/proteasome	CIC1, POC4
Protein folding/glycosylation	EMW1, NAB6, PER1, TDH3, ZIM17
RNA processing	HCA4, MTR3, MTR4, NAM8, NOP12, RPA135, RPA190, RPA43, RPA49, RRP46, RRP5, SKI6, SRB4, TAF14, TFC6
Unknown	INA1, NOG1, SKG6, YCR016W, YIL055C, YML018C, YML020W, YOR342C,

The set of 174 proteins whose expression was repressed after DTT treatment showed significant enrichment (FDR < 0.05) for a number of GO biological process terms relating to protein translation and ribosomal function including but not limited to ribosome biogenesis (45 genes), translation (36 genes), RNA processing (40 genes), rRNA processing (32 genes), ribosomal small subunit biogenesis (18 genes), ribosomal large subunit biogenesis (17 genes), cleavage involved in rRNA processing (13 genes), ribosome assembly (10 genes) and rRNA transcription (7 genes).

Table 14 - Proteins repressed by TM treatment.

Proteins that decrease in abundance or are lower in expression than expected in response to 2 µg/mL TM treatment with a local Z-score ≤ -2. Functional categories were adapted from those used in Costanzo et al., 2010.

Functional Category	Genes
Amino acid biosynthesis	ARG1, ARO8, CPA1, HIS4, LYS2, MET18, TAT1, YOR302W
DNA replication	MCM6, MCM7, SUA5
ER-Golgi traffic	SRP72

Cell polarity/morphogenesis	BZZ1, CLA4, GIN4, RAX2
Chromatin/transcription	BDF1, EAF5, IOC3, RAP1, RKM4, SWI1, UME6
Chromosome segregation	BRN1, PDS5, SLK19, SPC72
Nuclear-cytoplasmic transport	CRM1,
Drug/ion transport	FCY2, FET3, FTR1, TNA1, TPO1
Golgi/endosome/vacuole/sorting	MDR1, MRL1, NVJ1, SWA2, VAM7, VMA7
Lipids	AUR1, ERG5, FEN1, OSH3, PEX11
Metabolism	BNA6, CDC19, FBA1, GCV2, HNT1, HPT1, HXT3, HXT6, ILV6, MGR3, MRI1, OAC1, RCF2, RPO41, SPE2, SRX1, URA7, YHB1, YLH47
Ribosome/translation	ALB1, BMT6, BRX1, DBP7, DPH6, DRS1, DUS1, EMG1, FAL1, GAR1, GCD10, HGH1, KRI1, MAK11, MAK21, NAN1, NOP15, NUG1, PWP2, PXR1, RIA1, RPL1B, RPL22A, RPL23A, RPL23B, RPL8A, RPL8B, RPL9A, RPS10A, RPS10B, RPS17B, RPS21B, RPS26A, SMM1, SSF2, SQT1, SYO1, TEF4, TIF11, TOD6, TRM11, TSR2, UBA4, URB2, UTP21, UTP6, UTP8, URM1, YFL034W, YTM1
Signalling/stress response	GAL83, NAT5, SFK1, SIP1, SKM1, SKN7
Nuclear-cytoplasmic transport	ARX1
Protein folding/glycosylation	KRE6, NUS1, STT3
Protein degradation/proteasome	CIC1, ORM1, PTH2
RNA processing	DBP8, ECM16, HRB1, MLP1, NAF1, PRP45, RPA43, RPC17, RRP5, RRP8, SNU114, SYF1
Unknown	INA1, SSP120, SYH1, YCR087C-A, YGR122W, YIL055C, YKL071W, YRB30, YTA6

The set of 147 proteins whose expression was repressed after TM treatment also showed a significant enrichment (FDR <0.05) for ribosomal function including the GO biological process terms ribosome biogenesis (41 genes), RNA processing (37 genes), rRNA processing (27 genes), ribosomal large subunit biogenesis (16 genes) and ribosomal small subunit biogenesis (14 genes).

Given that translation and ribosomal function related proteins are being repressed in ER-stress conditions induced by both DTT and TM, it seems likely this represents a control mechanism of protein translation, presumably in an effort to lower the protein folding burden within the ER. Since there seems to be no global effect of lowered protein expression in our screens, this mechanism may take place over a longer timeframe than the four hour time point we measured here.

5.4 Localisation changes across the proteome

One of the major bottlenecks in proteome-wide high-throughput microscopy is the assessment of localisation changes. Typically high-throughput microscopy assesses one reporter at a time as we have done previously (Bircham et al., 2011) and with the DMA screening in chapter 4. Using single reporters allows for the development of highly accurate recognition algorithms specific to that reporter. Measuring proteome-wide localisation changes is more than a single reporter as there is such a broad range of proteins that can be visualised in the yeast GFP collection, all expressed to different levels and localised to their specific organelle. Additionally, the scope of proteins being measured means that not all possible localisation changes can be known *a priori* to define recognition procedures. Given these problems it is common for localisation shifts to be assessed by eye (Breker et al., 2013; Tkach et al., 2012). This is a time consuming and laborious process when dealing with tens of thousands of images per screen. Manual inspection can also be highly subjective and objectivity may change as more images have been observed and assessed. We sought to overcome some of the problems of manually visualising large datasets by first enriching for a subset of 'likely' localisation changes through texture analysis.

Texture analysis measures local patterns within the fluorescent distribution of an image and can provide an unbiased measurement of protein localisation shift. Texture measurements have widely been used in machine vision and facial recognition algorithms, and more recently applied to high-content microscopy. Texture analysis is not specific for any particular localisation change and therefore it is possible to broadly identify changes regardless of what they are and without prior knowledge. Many of the commonly used texture features such as Haralick textures, threshold adjacency statistics (TAS), and Gabor filters (Gabor, 1946; Hamilton et al., 2007; Haralick et al., 1973; Turner, 1986) produce a large number of features that may require complex processing such as dimensionality reduction (Dénervaud et al., 2013; Liu et al., 2004; Singan et al., 2012). Additionally these features often perform better for the analysis of more detailed images than is possible here due to the small size of yeast cell, as is the case for TAS features (N. Hamilton, personal communication Feb 2011). Here we combine the use of morphological features along with the spots edges and ridges (SER) texture analysis procedures available in

Acapella to measure localisation differences between control and treated images. We have found SER features to perform well with the small size of yeast cells and it produces a set of easily interpretable measurements.

The Acapella SER texture analysis uses a set of eight characteristic fluorescence patterns to identify textural differences between images. These are spots, holes, edges, ridges, saddles, valleys, bright and dark. The SER features are generated by applying a set of filters based on second order Gaussian derivatives to the original image. The corresponding texture feature is then characterized by mean intensity of the filtered image in that region. SER measurements are controlled by a scale factor that defines the size of the texture properties. As these features are based on fluorescence patterns, they are independent of overall intensity and should not be affected by changes due to exposure conditions.

To define a list of probable localisation changes induced by DTT and TM treatment, we measured a set of phenotypic fluorescent properties as well as texture features for control and treated images. The properties measured were; the overall GFP intensity, the ratio of peripheral GFP to internal GFP intensity, the proportion of bright pixels within the cell and on the periphery, spot formation and the eight SER texture features measured at two different scales. For each measurement a Z score was calculated using the following formula:

$$Z\ score = \frac{Median_{treated} - Median_{control}}{MAD_{control}}$$

Medians were used as they are more robust to outliers and as such the median absolute deviation was used in place of standard deviation. The Z scores from the SER features were combined for both scales using Fisher's method. Each measurement was then rank ordered based on Z score and the top 100 hits of each feature were combined to get a list of probable localisation changes. This method produces a set of ~ 400 probable protein localisation changes, reducing the number of images needed to be inspected by eye to roughly 10% of the total collection.

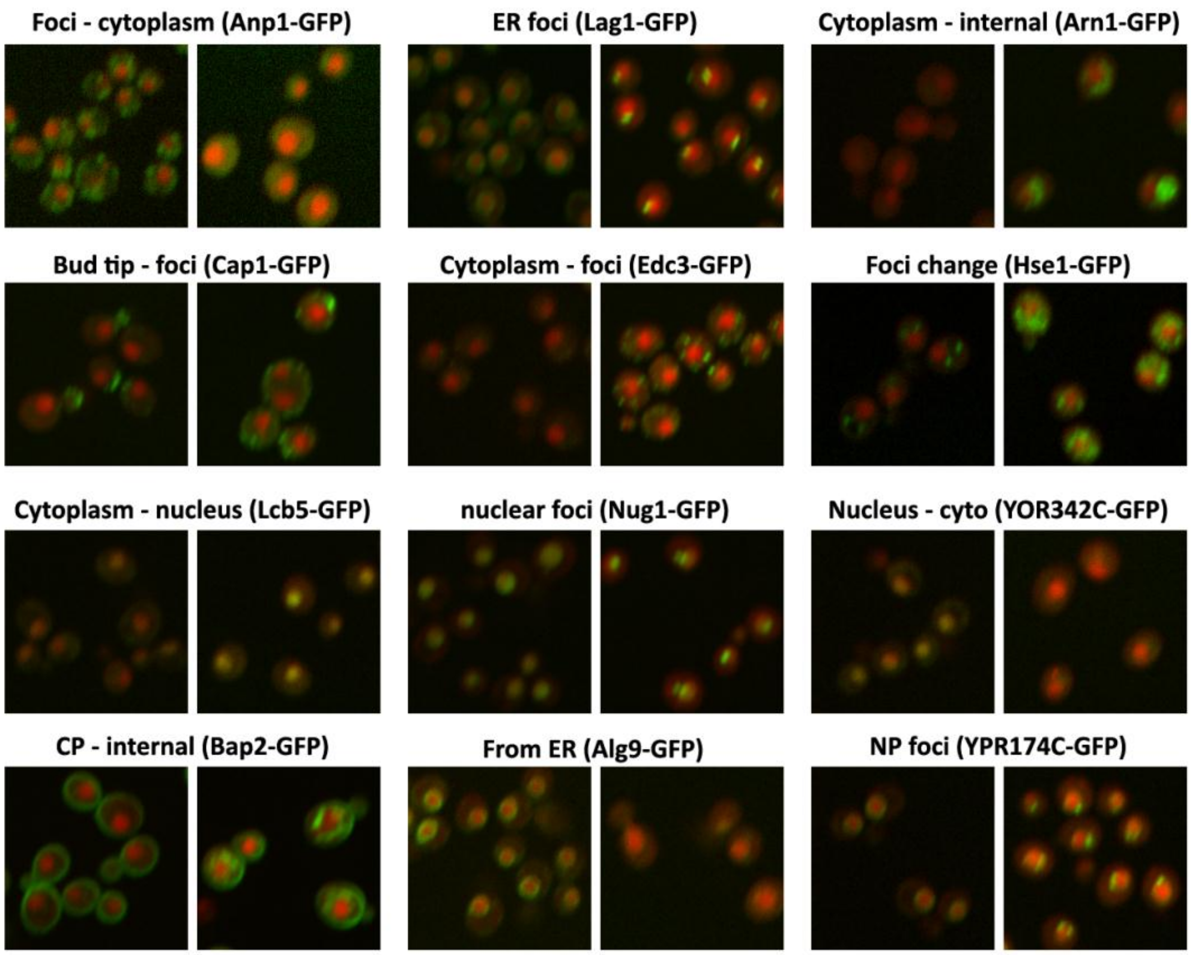

Foci - cytoplasm (Anp1-GFP)
ER foci (Lag1-GFP)
Cytoplasm - internal (Arn1-GFP)
Bud tip - foci (Cap1-GFP)
Cytoplasm - foci (Edc3-GFP)
Foci change (Hse1-GFP)
Cytoplasm - nucleus (Lcb5-GFP)
nuclear foci (Nug1-GFP)
Nucleus - cyto (YOR342C-GFP)
CP - internal (Bap2-GFP)
From ER (Alg9-GFP)
NP foci (YPR174C-GFP)

Figure 28 - Example localisation changes.

Example GFP localisation changes overlaid with nuclear/cytoplasm RFP; control is on the left and DTT treated on the right for each panel. CP -cell periphery, NP -nuclear periphery. Image brightness has been adjusted equivalently in each pair for optimal image reproduction.

To test the fidelity of this method, the entire set of TM treated images were assessed by eye and compared to the list of probable hits. All localisation changes identified by eye were encompassed in this set as well as some subtle changes that went unnoticed on first inspection. However, as texture analysis is sensitive to image quality, a significant portion of these changes were due to unfocused images or blurring caused by cell movement. As such the set of probable changes for each condition was further assessed by eye. Strains that showed genuine localisation changes in either condition were reimaged in triplicate at a higher resolution for reassessment in both conditions. Proteins that consistently showed a change in localisation across replicates were considered hits and listed in Table 15 below with examples displayed in Figure 28.

Table 15 - Protein localisations altered by ER stress induced by TM or DTT treatment

Highlighted in red are proteins that only changed in response to DTT.

Localisation Change	Genes
Bud tip -> foci	AXL2, CAP1, CAP2, EDE1, RSE1
Cytoplasm -> foci	AIP1, AKL1, AMS1, APE1, DCP1, DCP2, DYS1, EDC3, EMI2, GGA1, GLK1, GLT1, GSY1, HSP26, HSP42, KAP95, RBS1, XRN1, YAR009C, YDR170W-A, YIL108W
Cytoplasm -> internal	ARN1, ATG18, OPY2, PHM7, PRM5, RCR1, RCR2, SUR1, UIP3
Cytoplasm -> nucleus	CRZ1, LCB5, RPL18B, RRP12, STB3, TOD6, TSR3, VHR1
ER -> Cytoplasm	ALG2, ALG9
ER foci	ERG1, ERG6, FAA1, FAA4, IRE1, LAC1, LAG1, NUS1
Foci -> Cytoplasm	ANP1, CDC15, KEX2, MNN10, MNN11, MNN5, PEX21, RNR1, VAN1
Foci change (spread)	CCZ1, DID4, FCY2, HSE1, IST1, ITR1, MVP1, PEP1, PEP8, PSO2, SNF7, VPS16, VPS17, VPS24, VPS38, VPS4, VPS8, VTA1
Nuclear periphery -> foci	HMG1, HMG2, NSP1, NUP159, NUP82, YPR174C
Nucleus -> Cytoplasm	CDC20, CDC24, DUS3, GCN4, KAP123, YOR342C
Nucleus -> nucleus foci/nucleolus	CDC16, DRS1, FPR3, FPR4, GLC7, JIP5, MAK11, NOP16, NSA1, NUG1, RDH54, RIX1, SIS1, THO2, WTM1, WTM2
Cell periphery -> internal	AGP2, ARF3, AQR1, BAP2, DIP5, FLC1, FUI1, HNM1, HSP30, HXT2, HXT3, HXT6, MEP1, MRPL6, QDR3, RSN1, THI7, YDR090C

The 126 proteins that changed localisation in either TM or DTT showed enrichment in the GO biological process terms including: late endosome to vacuole transport (12 genes), endosomal transport (14 genes), vacuolar transport (16 genes), protein glycosylation (8 genes), protein targeting to vacuole (8 genes), actin filament depolymerisation (3 genes), transport (55 genes) and localisation (58 genes).

GO cellular process enrichments included: endomembrane system (47 genes), endosome (16 genes), plasma membrane (25 genes), vacuole (17 genes), alpha-1,6-mannosyltransferase complex (4 genes), ESCRT III complex (3 genes), F-actin capping protein complex (also WASH complex, 2 genes) and acyl-CoA ceramide synthase complex

(2 genes). The gene set also showed enrichment in the GO molecular function mannosyltransferase activity (7 genes). Looking at only the proteins that changed localisation in both DTT and TM did not show any changes in GO enrichements.

Publications that showed enrichment for the hit list included a DNA damage screen of the yeast GFP collection (Tkach et al., 2012) and various studies on protein sorting and vacuole function: (Banta et al., 1988; Nickerson et al., 2010; Robinson et al., 1988; Rothman et al., 1989).

5.5 Changes compared to UPR deficient GFP set

To determine the contribution of the UPR to the proteomic changes induced by ER-stress, we created a UPR deficient yeast GFP collection for comparison. A modified SGA procedure was used to introduce a *Δhac1* knock-out mutation into the yeast GFP collection to prevent the normal transcriptional up-regulation of UPR targets. This collection was imaged under DTT induced ER-stress conditions as previously described for comparison to wild-type cells. Abundance changes are listed in Table 16 and Table 17 below. Localisation changes are listed in Table 18 .

Table 16 - Proteins up-regulated by DTT in the absence of a functional UPR (Δhac1).

Proteins that increase in abundance in response to 2 mM DTT treatment with a local Z-score ≥ 2. Functional categories were adapted from those used in Costanzo et al., 2010. Proteins highlighted in red were not significantly up-regulated by DTT in the presence of a functional UPR (Table 11)

Functional Category	proteins
Amino acid biosynthesis/transport	ARO10, GAP1
Cell polarity/morphogenesis	ABP1, AIP1, BUD5, LSB3, MYO3, PIL1, PIN3, RCR1, SAC6, SDS24
Cell cycle progression/meiosis	OSW5, SDS22
Chromatin/transcription	BDF1, IES4, WTM1
ER-Golgi traffic	BET3
DNA replication/repair	DDR48, MSC1
Drug/ion transport	FLC1, INH1, MEP1, MEP2, PDR5

Golgi/endosome/vacuole/sorting	APE1, GGA1, GYP7, PBI2, RUP1, SAC1, SNA3, VPS68, VPS1
Lipids	ERG24, FMP45, GPT2
Metabolism	ACH1, ADH4, ALD3, ALD4, ARA1, ARI1, BNA1, DAK1, DUG1, GAD1, GDB1, GLK1, GLO1, GOR1, GPD1, GPP1, GPP2, GRX1, GSY2, GTT1, IDH2, IGD1, ISU1, KGD2, LSC1, MBF1, NCE103, NRG1, PAM17, PGM2, PNC1, PUT2, RIB4, SER3, SOD2, SRX1, STF1, STF2, TAL1, TPS1, TPS2, TSA2, TSL1, URA10, YDL124W, YMR315W
Ribosome/translation	MNP1, YMR295C
Signalling/stress response	AHP1, CMK2, GRE3, HSP12, HSP26, HYR1, MCA1, MRP8, RTS3, TFS1, TSA1
Protein degradation/proteasome	AFG3, UBC5
Protein folding/glycosylation	HSP104, HSP42, HSP78, OCH1, PUN1, SLT2, TDH1, YLR194C, YPS3
RNA processing	DCS1, EDC2, IGO1, RTC3, SRB6, SSA4, SSE2
Unknown	AIM17, AIM41, COS6, FMP21, PHM7, RCN2, RDL1, RGI1, TDA1, VID27, YBR085C-A, YBR287W, YDR391C, YHR097C, YHR138C, YJL016W, YJR085C, YKL151C, YLR257W, YNK1, YNL134C, YNR014W, YNR034W-A YOR385W, YPL067C, YPL260W, YPR127W, YPR172W

76 of 225 proteins up-regulated by DTT in wild-type cells were also up-regulated in the UPR deficient collection (142 proteins up-regulated in total). Interestingly there were an additional set of 67 genes only that were up-regulated in the UPR deficient strains. This may be due to an altered homeostasis and buffering effects due to the UPR deficiency, but may also include some genes that were just below the significance threshold in the functional UPR DTT screen.

This set of 142 proteins showed enrichment for a number of GO terms including the expected, redox related terms. GO biological process enrichments included oxidation-reduction process (28 genes), response to oxidative stress (14 genes), trehalose metabolic process (5 genes) and cellular carbohydrate biosynthetic process (9 genes). Interestingly this set was no longer enriched for endoplasmic reticulum genes as the WT DTT screen hits list was.

Table 17 - Proteins repressed by DTT treatment in the absence of a functional UPR (Δhac1).

Proteins that decrease in abundance or are lower in expression than expected in response to 2 mM DTT treatment with a local Z-score ≤ -2. Functional categories were adapted from those used in Costanzo et al., 2010. Proteins highlighted in red were not significantly repressed by DTT in the presence of a functional UPR (Table 13).

Functional Category	Genes
Amino acid biosynthesis	ADE6, ARG7, ARO2, ARO4, BAP3, CHA1, CPA1, CPA2, GNP1, HIS5, HOM3, LTV1, LYS2, PRO1, PRS1, TRP4, URE2
Cell polarity/morphogenesis	BEM2
Chromatin/transcription	HHO1, MOT1, SCP160, SFH1
Chromosome segregation	SMC1
DNA replication	RFA2
Golgi/endosome/vacuole/sorting	EMP70, RTT10
Drug/ion transport	CAR1
Cell cycle progression/meiosis	SAP185
Lipids	HNM1
Metabolism	AAH1, GCV2, PDC1, SAM1, SAM4, SHB17
Ribosome/translation	ALB1, BUD20, DBP3, DPH6, EFG1, EFT1, EFT2, FYV7, IPI1, LIA1, MAK21, NAN1, NEW1, NIP1, NIP7, NOC2, NSA1, NSA2, PXR1, RIX1, RPG1, RPL11A, RPL13B, RPL17A, RPL18B, RPL21A, RPL22B, RPL23A, RPL27B, RPL2A, RPL31B, RPL33B, RPL34B, RPL38, RPL40A, RPL42A, RPL42B, RPL43B, RPL4B, RPL4A, RPL5, RPL6B, RPL7B, RPL8A, RPL8B, RPP1A, RPS14B, RPS16B, RPS17A, RPS18B, RPS19A, RPS21A, RPS22B, RPS23B, RPS24A, RPS24B, RPS25B, RPS30B, RPS7B, RPS8A, RPS8B, RRP12, RRP15, TRM82, RRP7, SGD1, SRO9, SUI2, TSR2, URB1, UTP14, UTP21, UTP6
Signalling/stress response	BCY1, GIS2
Nuclear-cytoplasmic transport	ARX1, SXM1
Protein degradation/proteasome	CIC1, UBP1, UBP10
Protein folding/glycosylation	CWP1, PLP2, TDH3, WBP1
RNA processing	BUD31, DBP8, ECM16, HCA4, MSL5, NOP12, PRP43, RPA12, RPA135, RPA190, RPA43, RPO31, RRP5, TAF14
Unknown	INA1, YIL055C

56 of the 174 proteins repressed in the wild-type DTT screen were also repressed in UPR deficient conditions. An additional 79 proteins were considered to be repressed in the UPR deficient *Δhac1* screen, which as mentioned in the up-regulated protein results

above, may be due to homeostasis changes and/or proteins that fell just outside of significance in the WT screen. Of note, the set of 135 proteins repressed in the UPR deficient screen were still enriched for translational and ribosomal related GO terms (FDR < 0.05) including the GO biological process terms; gene expression (90 genes), ribosome biogenesis (54 genes), translation (47 genes), rRNA processing (38 genes), ribosomal small subunit biogenesis (28 genes), ribosomal large subunit biogenesis (21 genes), ribosome assembly (12 genes) and translational elongation (11 genes). This indicates that any translational repression due to this is likely to be UPR independent and may be part of a more general stress-response.

Table 18 - Localisation changes induced by ER-stress in UPR deficient cells.

Proteins highlighted in red did not change localisation in cells with a functional UPR (Table 15)

Localisation Change	Proteins
Cytoplasm -> foci	AIP1, AKL1, APE1, DCP1, DCP2, EDC3, EMI2, GGA1, GLK1, GLT1, GSY1, HSP26, HSP42, KAP95, PPZ1, RBS1, SRP1, XRN1, YAR009C, YCR043C, YDR170W-A, YIL108W, YJL017W
Cytoplasm -> internal	OPY2, PHM7, PRM5, RCR1, RCR2, SUR1, UIP3
Cytoplasm -> nucleus	CRZ1, TSR1, LCB5, RPL18B, RRP12, STB3, TSR3, VHR1
ER foci	ERG1, ERG6, FAA1, FAA4, IRE1, LAC1, LAG1, NUS1
Foci -> Cytoplasm	ANP1, CDC15, GEA2, KEI1, KEX2, MNN5, PEX21, RNR1, VAN1
Foci change (spread)	CCZ1, FCY2, HSE1, IST1, ITR1, MVP1, PEP1, PSO2, VPS16, VPS17, VPS24, VPS38, VPS4, VPS8, VTA1
Nuclear periphery -> foci	HMG1, HMG2, NSP1, NUP159, NUP82, YPR174C
Nucleus -> Cytoplasm	CDC24, DUS3, GCN4, KAP123, YOR342C
Nucleus -> nucleus foci/nucleolus	CDC16, CGR1, DRS1, FPR3, FPR4, GLC7, JIP5, MAK11, MRT4, NOP16, NSA1, NUG1, RDH54, SIS1, THO2, WTM1, WTM2
Cell periphery -> internal	ARF3, AQR1, BAP2, DIP5, FLC1, FUI1, HNM1, HSP30, HXT2, HXT3, HXT6, INA1, MEP1, MRPL6, PDR5, QDR3, RSN1, TAT1, THI7, VHT1, YDR090C

Interestingly almost all of the localisation changes in the *Δhac1* UPR deficient DTT screen were also observed in the functional UPR screen, indicating that UPR regulation of the proteome is mainly actuated through increased protein expression rather than control through localisation. Only 13 new localisation changes were observed in the UPR deficient set. These are likely to be effects either through increased levels of ER-stress due to a lack of UPR response and/or homeostatic changes that may include buffering systems to deal with the lack of a functional UPR. It is noteworthy that a number of the changes seen in the functional UPR DTT screen were not seen as changes in the *Δhac1* screen as the proteins were already in the secondary localisation prior to DTT treatment. Specifically these proteins were: AGP2, ALG2, ALG9, AMS1, CAP1, CAP2, CDC20, DID4, EDE1, MNN10, MNN11 and PEP8. This is probably due to the increased level of basal ER-stress in UPR deficient cells that cannot effectively deal with the normal levels of protein misfolding. Therefore, these localisation changes must be UPR independent responses. Only seven proteins; ARN1, ATG18, AXL2, DYS1, RIX1, RSE1 and TOD6 were found to change localisation only in the functional UPR screen, as such these localisation changes are likely due to a direct effect of UPR activation. One protein SNF7 could not be assessed in the Δhac1 screen as there were too few cells for analysis.

5.6 Transcription factor analysis

Transcription factor (TF) analysis was used to identify any potential regulators, other than Hac1p, the well-established TF responsible for the expression changes induced by ER-stress. YEASTRACT, a comprehensive database of all known transcription factor interactions in yeast (Abdulrehman et al., 2011; Monteiro et al., 2008; Teixeira et al., 2006, 2013) was used to identify all TFs present in either the expression or localisation hit sets. These TFs were then assessed for activating regulatory interactions amongst the up-regulated protein expression data.

5.6.1.1 Transcription factors activated in response to DTT

Taking all TFs in the DTT abundance and localisation data, 11 TFs were found to change. Nine of these had potentially regulating interactions amongst the DTT hits and had regulatory interactions with 50% (112/225) of the proteins in the hit list (57% or 129/225 including Hac1p interactions). Three of these TFs were considered to be significantly enriched for interactions amongst up-regulated hits (pval<0.001 from YEASTRACT), Crz1p (20% of hits), Wtm2p (9.33%) and Cup2p (15.11%) together accounting for 67 possible regulatory interactions with protein hits. By comparison Hac1p, the TF component of the UPR showed potential regulatory functions with 15.56% of the hits list. Other TFs that were identified as interacting with the hit list but fell below statistical significance were Gcn4p (34.67%), Spt2p (5.78%), Sds3p (1.78%), Wtm1p (1.33%), Tod6p (0.89%) andTho2p (0.44%).

Searching the full set of known TFs in the yeast genome we found 23 TFs potentially regulating the hit set with significant target enrichment, together accounting for possible interactions with 196 of the 225 proteins. We searched the Prophecy database (Fernandez-Ricaud et al., 2005, 2007) for DTT sensitivity to deletions of these TFs and found that deletion mutants of 8 TFs (HAC1, MGA2, SWI4, RLM1, CRZ1, FLO8, MSN4, YAP6) are designated as sensitive to 1.6mM DTT. Notably Crz1p was the only DTT sensitive TF to show a phenotypic change upon DTT treatment in our data.

5.6.1.2 Transcription factors activated in response to TM

Assessment of the TM abundance and localisation data showed that Hac1p was the most enriched TF with reported interactions covering 17.51% of the TM up-regulated hit set (pval< 0.001), followed by Crz1p interacting with 12.44% of the hit set and Wtm2p with 5.53%. TFs that fell below statistical significance but had regulatory interactions with the hit set included Mcm1p (18.89%), Gcn4p (27.19%), Tup1p (15.21%), Wtm1p (2.30%) and Pog1p (1.38%).

Searching the full set of known TFs we found a total of 15 TFs with potential activating regulatory interactions (p-value<0.01) accounting for possible interactions with 154 of the 217 proteins in the TM hit list. Searching SGD we found that deletion mutants in 3 of these TFs (CRZ1, HAC1 and RLM1) result in a sensitivity to TM.

5.6.1.3 Transcription factors that respond to DTT in UPR deficient cells.

Searching the *Δhac1* UPR deficient screen for TFs that change in expression or localisation we identified seven TFs with potential activator interactions. Four of these showed significant enrichment of target genes in the *Δhac1* up-regulated proteins hit set; Crz1p (30.28% of hits), Gcn4 (73.94% of hits), Wtm2 (16.9%) and Nrg1 (22.5%). Others that fell below statistical significance were Mbf1p (6.34% of hits) and Wtm1p (3.52%).

Searching the full set of yeast TFs revealed 34 potential regulators, nine of which cause DTT sensitivity in null mutations (MGA2, SWI4, RLM1, CRZ1, CIN5, FLO8, ADR1, MSN4, YAP6). Altogether the 34 possible TFs account for activator interactions with 96% of

Δhac1 up-regulated proteins hits. Limiting to the nine DTT sensitive TFs still accounts for interactions with 91% of hits. As expected Hac1p was no longer enriched as a TF for interactions with the hit set.

5.7 Discussion

5.7.1 Comparison of DTT and TM as ER-stress inducers

Comparing the expression data from the DTT and TM screens, we see a significant portion of each hit set is unique to each stress condition (Figure 26). These discrepancies can be rationalised as a drug specific component of the cells stress response. The protein set up-regulated by DTT for example, is uniquely enriched for the redox related gene ontology terms oxidation-reduction process (46 genes) and oxidoreduction coenzyme metabolic process (11 genes). These processes can evidently be attributed to DTT's mode of action as a reducing agent. These condition-specific hits are not in themselves particularly pertinent in the context of studying the UPR and so are not discussed here in detail. More interesting is the overlap of 80 up-regulated proteins between DTT and TM treatments. These proteins are likely to represent the set of proteins required to deal with ER-stress and the resulting physiological consequences. Although this set of 80 proteins is up-regulated in both conditions, they may not necessarily be UPR specific hits. In fact 20 of the 80 proteins are also significantly up-regulated by DNA damaging agents in a similar GFP-based screen (Tkach et al., 2012), see Figure 29 and Supplementary appendix 2. Additionally, searching Yeastmine for publication enrichments showed that a number of the overlapping hits were also enriched in a number of stress condition studies including cell wall stress and salinity (Boorsma et al., 2004; García et al., 2009; Liu et al., 2007; Melamed et al., 2008). Transcription factor analysis using YEASTRACT identified 9 possible transcription factors with statistically significant activating interactions amongst the 80 overlap hits. As expected, Hac1p is amongst these with reported regulatory interactions with 26.25% of the hits. Other possible TFs were Rsc1, Mga2, Cbf1, Spt23, Rgm1, Crz1, Wtm2 and Rlm1. Distinct among these are Crz1p and Wtm2p which both show a phenotypic change in DTT and TM treatment. Crz1p is a calcineurin-dependent TF known to activate stress response genes (Yoshimoto et al., 2002) and has potential activating interactions with 26.25% of the 80 overlap hits. Wtm2p showed potential activating regulatory interactions with 10% of the overlap hits. Wtm2p is a transcriptional regulator that has been implicated in replication stress (Pemberton and Blobel, 1997; Tringe et al., 2006). Wtm2p has a paralog Wtm1p which also showed a localisation shift in DTT and TM treatments. DTT treatment also caused a foci to cytoplasm localisation shift of the Wtm2p target gene Rnr1p. Rnr1p is a known DNA damage response gene and a similar localisation shift has previously been described for autophagy mutants (Dyavaiah et al., 2011).

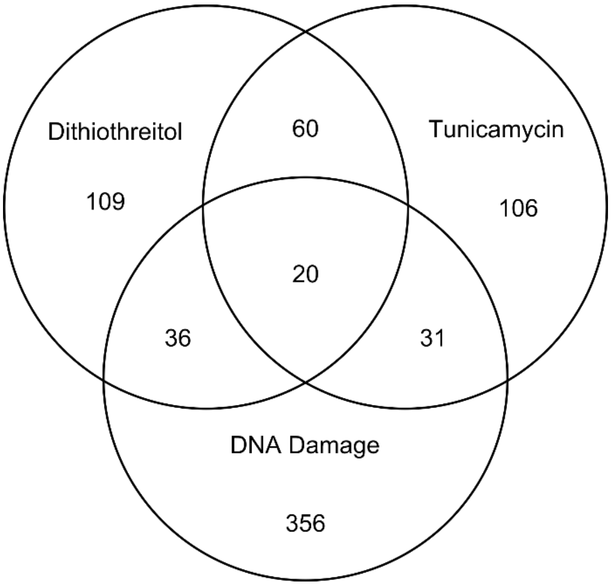


Figure 29 - Overlap between ER-stress inducers and DNA damaging agents.

Overlap between proteins up-regulated by DNA damage (Tkach et al., 2012) and the ER-stress agents DTT and TM.

5.7.2 GFP proteomic screening reveals novel UPR target genes

We initially defined possible UPR targets as proteins whose expression was significantly up-regulated in DTT treatment but not in UPR deficient cells i.e. the *Δhac1* screen (Figure 30 yellow and green). From this we identified 149 UPR target proteins (see supplementary appendix 2 for a list of genes), of which 38 overlapped with previously identified UPR targets (Kimata et al., 2006; Travers et al., 2000) and 28 of which overlapped with DNA damage hits (Tkach et al., 2012). Among these 149 possible UPR targets are 15 proteins of unknown function. A second group of possible UPR targets are the set of 172 proteins whose expression was significantly up-regulated in TM treatment but not in UPR deficient cells (Figure 30 blue and green). However as we did not screen the UPR deficient cells with TM induced ER-stress, we cannot be sure how many of these may be TM-specific effects that are not genuine UPR targets and as such less emphasis was put on these hits.

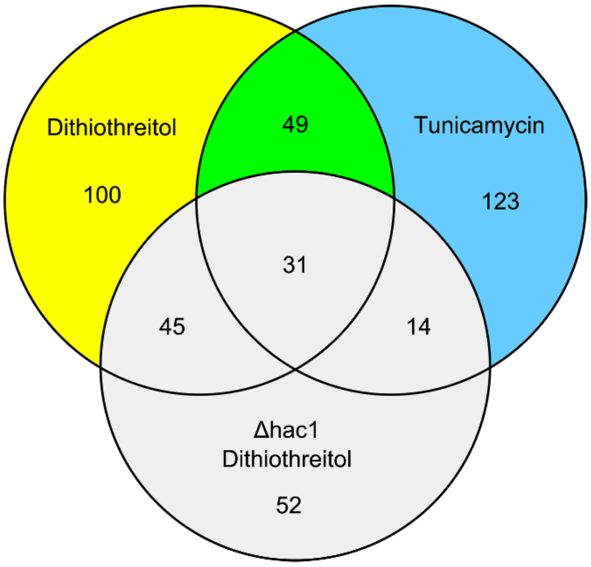


Figure 30 - overlap of up-regulated proteins between DTT, TM and UPR deficient cells

By further refining a set of definite UPR targets as those proteins whose expression was significantly unregulated in both DTT and TM treatment but not in UPR deficient cells, we identify a total of 49 UPR targets (Table 19 and Figure 30 green). By limiting the set of definitive UPR targets to those that respond to both DTT and TM, secondary drug-specific hits from DTT or TM should be eliminated. Of these more stringent hits 27 are novel UPR targets when compared to UPR targets as defined in the literature (Kimata et al., 2006; Travers et al., 2000). A search on YEASTRACT revealed that for these 27 novel UPR targets there is no known TFs having significant activator interactions with this group. This lends weight to the likelihood that these are unique UPR hits rather than from additional stress responses. It is also noteworthy that seven of these novel UPR targets are proteins of unknown function.

Table 19 – UPR target Genes.

UPR-specific target proteins. The 27 novel targets are highlighted in red and proteins of unknown function are underlined.

UPR specific up-regulated proteins
ACB1, ADD37, BFR1, BSD2, CAP1, CBP3, COQ4, COS10, DFM1, DOG2, ERP1, ERP2, GET4, GWT1, HIS2, HXT2, MHR1, MRPL38, MSP1, NCE102, OPI3, ORM2, OYE2, PAR32, PET10, PRM5, PST2, RET2, RIB1, SDH4, SEC62, SEC66, SEC72, SFB2, SHR3, SOL4, SPC2, SUR1, TDH2, TMA17, UBC8, YET1, YET3, YIP3, YLR345W, YMR122W-A, YOR289W, YPL107W, ZRT1

In addition to up-regulated proteins, 20 localisation changes were found only in UPR functional cells (Table 20). However, when inspecting the untreated *Δhac1* images we found that 12 of these proteins already displayed the stress induced phenotype. This suggests that the deletion of hac1 itself causes enough ER-stress to induce these phenotypes, and that these phenotypes are not UPR regulated.

Table 20 - UPR localisation changes.

Proteins that showed a localisation shift to DTT or TM in wild-type but not Δhac1 cells. Underlined proteins are those that already showed the ER-stress phenotype in untreated Δhac1 cells.

Proteins showing a localisation change in TM or DTT but not UPR deficient cells
AGP2, ALG2, ALG9, AMS1, ARN1, ATG18, AXL2, CAP1, CAP2, CDC20, DID4, DYS1, EDE1, MNN10, MNN11, PEP8, RIX1, RSE1, SNF7, TOD6

Among the novel UPR targets were two proteins Yet1p and Yet3p, homologues to the mammalian B-cell receptor-associated protein BAP31 implicated in ER quality control and secretory cargo export (Toikkanen et al., 2006). These have been recently shown to form a heteromeric complex that interacts with the ER translocon Sec complex, potentially regulating the biogenesis of specific transmembrane proteins (Wilson and Barlowe, 2010). Previous studies have also shown YET1, YET2 and YET3 mRNA levels increase in response to DTT (Gasch et al., 2000), however they had not previously been shown to be UPR specific targets. As there is no GFP-tagged variant of Yet2p in the GFP collection we were unable to assess whether Yet2p is also a UPR target. However Yet2p seems to function separately from the Yet complex (Wilson and Barlowe, 2010). Additionally it was recently shown that the YET complex has regulatory interactions with Opi1p a repressor of Ino2p/Ino4p mediated control of phospholipid biosynthesis (Wilson et al., 2011). Indeed Opi3p a known Ino2/Ino4 target is also an up-regulated UPR target here, suggesting a possible role of the YET complex in UPR mediated up-regulation of phospholipid biosynthesis.

Another novel UPR target is Shr3p an ER packaging chaperone. Shr3p is required for specific packaging of amino acid permeases into COPII coated vesicles for delivery to the cell surface (Gilstring et al., 1999; Ljungdahl et al., 1992). Interestingly we also show that two of the permease substrates of Shr3p, Dip5p and Bap2p, (Herzig et al., 2012; Wright et al., 1997) show localisation shifts from the plasma membrane to an ER-retention-like phenotype similar to shr3 null mutants. In this case the mislocalisation of these permeases is either due to TM/DTT induced misfolding or general secretory pathway shutdown, Shr3p is presumably up-regulated to cope with the processing of such proteins.

The ubiquitin ligase adaptor Bsd2p was also identified as a novel UPR target protein. Bsd2p acts as an adapter protein for the ubiquitin ligase Rsp5p, an essential protein for ubiquitin-dependent trafficking events in yeast including trans-membrane proteins (Belgareh-Touzé et al., 2008). Bsd2p recognises exposed hydrophilic residues in the trans-membrane domain of target proteins that lack the Rsp5p recognition ‘PY’ motif. Bsd2p itself contains PY motifs and thus is able to recruit Rsp5p to target proteins via this interaction (Hettema et al., 2004). Ubiquitination then acts as a signalling mechanism targeting these proteins to the vacuole for degradation. Bsd2p has been implicated in the homeostatic regulation of a number of trans-membrane proteins and permeases including heavy metal ion homeostasis. Bds2p facilitates the trafficking of the heavy metal ion transporters Smf1p and Smf2p to the vacuole for degradation thus preventing heavy metal accumulation within the cell (Liu et al., 1997; Sullivan et al., 2007).

It has been suggested that Bsd2p may in fact act as a general mechanism to target misfolded membrane proteins for degradation in the vacuole (Hettema et al., 2004). However, given that Bsd2p is involved in the homeostatic trafficking of proteins for physiological function it has remained unclear whether Bsd2p plays a general role in protein quality control or not (Fredrickson and Gardner, 2012). Our results presented here suggest that Bsd2p is likely to play a role in general protein quality control given that it is up-regulated in a UPR specific manner.

Additionally we identified Tma17p (also known as Acd17p) as a novel UPR target. Tma17p has recently been shown to act as an ATPase dedicated chaperone crucial for homeostatic maintenance of the proteasome under stress conditions (Hanssum et al., 2014). As discussed previously in chapter 1, the proteasome is tightly linked to the UPR and is ultimately responsible for the degradation of terminally misfolded proteins. A number of proteasome subunits are known to be upregulated under conditions of protein folding stress, which in yeast are under the control of the transcription factor Rpn4p (Xie and Varshavsky, 2001). Notably, deletion of RPN4 showed up in the DMA screens presented here as an inducer of UPRE expression. However, expression of subunits alone is not enough to increase proteasome levels. Tma17 has recently been shown to increase under ER-stress conditions and acts to assist in the assembly of complete proteasomes (Hanssum et al., 2014). Interestingly TMA17 was not previously identified as a UPR target at the mRNA level (Kimata et al., 2006). Closer inspection of their data shows that TM treatment did induce TMA17 expression at the transcript level but it failed to meet their requirements to define UPR targets as a constitutively active $HAC1^i$ mutant did not induce expression. This is interesting as Kimata et al. state that $HAC1^i$ is overall a stronger inducer of UPR targets than TM in their system, yet both DTT and TM induced Tma17p expression in a UPR specific manner in the results presented here. This allows the novel suggestion that simple IRE-HAC1 regulation is not enough to induce expression of all UPR targets. There likely exists cross-talk between other transcription factors that modulate the level and selection of UPR target genes required for a response. This concept is further discussed later on in this chapter.

As expected the UPR-specific proteins we identified here include previously identified UPR targets, such as members of the sec63 complex (Sec62, Sec66 and Sec72), Erp1p and Erp2, members of the p24 complex involved in ER-Golgi traffic and Opi3p required for phospholipid biosynthesis. The major notable difference between our screen for UPR targets and previous genome-wide screens for UPR targets (i.e. Kimata et al., 2006; Travers et al., 2000) is that the previous screens used cDNA microarrays. The prior screens therefore measured regulation at the transcript level, whereas our approach measured regulation at the protein level. Both approaches inherently have their merits. Microarrays are able to measure early transcript expression changes whereas GFP fusion proteins require time to fold and mature properly before they can be measured. Additionally C-terminal GFP tagging may cause protein defects and/or interfere with localisation or regulation. An example of this is the canonical UPR target protein Kar2p, an essential ER chaperone whose localisation is maintained by a C-terminal HDEL tag. C-terminal GFP tagging disrupts the HDEL tag causing the GFP variant of Kar2p to mislocalise (Huh et al., 2003). Because of this there are some proteins that cannot be accurately assessed by GFP tagging. In our analysis we assessed 4024 proteins that we considered unlikely to be affected by the GFP tag, which still represents 69% of the currently known verified ORFs. GFP screens do however offer some major advantages, primarily as the fact that measurements are taken at the protein level and therefore directly measure physical changes in protein regulation. It is well known that changes in transcript level do not always account for a change of protein level (Gygi et al., 1999). Additionally, GFP provides the unique ability to assess localisation changes in protein localisation on a proteome scale. These unique factors of each approach complement each other, as highlighted by the fact that each methodology produced a set of unique hits as well as an overlap of consistent UPR targets between them.

5.7.3 Additional stress responses act alongside the UPR

Looking at the amount of up-regulated proteins in ER-stress conditions that are still activated in UPR deficient cells (Figure 30 grey overlaps), we can see there are a significant number of proteins induced that are likely regulated through pathways acting in parallel to the UPR. We assessed the TFs that may regulate the 31 UPR-independent up-regulated proteins overlapping between the DTT, TM and *Δhac1* screens (Figure 31). Statistically Crz1p and Wtm2p are the stand out candidates with potential interactions amongst 42% and 23% of the hits respectively (target enrichment p-value < 0.001). Additionally Gcn4p, a TF which was found to localise more to the nucleus, interacts with 58% of hits although this is less statistically significant given the large number of documented interactions of Gcn4p in YEASTRACT. Gcn4p is responsible for the general amino acid control response (GAAC) and is a basic-leucine zipper transcription factor, like

Hac1p. Gcn4p is well known to interact with Hac1p and is required for induction of many UPR target proteins (Patil et al., 2004). ER-stress induced by TM has also been shown to modulate GCN4 mRNA expression independently of Hac1p to similar levels as amino acid starvation (Herzog et al., 2013). Cross talk between the UPR and GAAC pathways has been well documented (Herzog et al., 2013; Patil et al., 2004), therefore it is not surprising that Gcn4p may be activated alongside the UPR in response to ER-stress, even in UPR deficient cells. Interestingly the TF Mbf1p was up-regulated in *Δhac1* cells treated with DTT but not in wild-type cells and shows potential interactions with 10% of the UPR independent hits. Mbf1p is a transcriptional co-activator of Gcn4p that functions by bridging the TATA binding protein Spt1p to the DNA binding region of Gcn4p (Takemaru et al., 1998). This activation of Mbf1p suggests an increased dependence on the GAAC pathway in UPR deficient cells exposed to ER-stress.

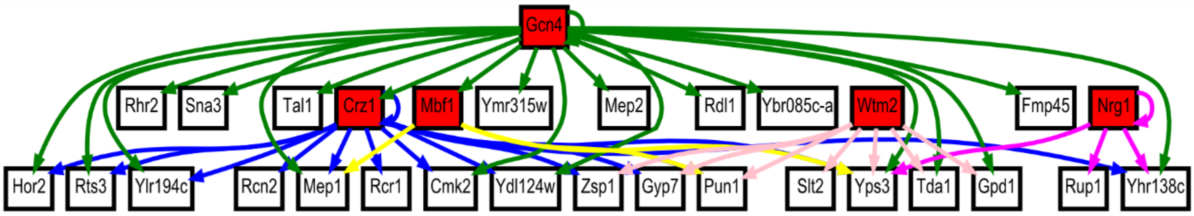


Figure 31 - UPR independent TF regulation.

Transcription factor interactions amongst UPR independent hits generated using YEASTRACT. TFs are highlighted in red. Arrows denote activator interaction direction, colours distinguish TFs.

Comparing the UPR independent up-regulations to the Tkach et al., (2012) dataset we find that 10 of the 31 proteins are up-regulated by DNA damage, and further to this SGD annotations describe a total 16 of the 31 hits to be involved in a non-UPR stress response including oxidative, osmotic and cell wall stress.

Interestingly Cap1p was shown to be specifically UPR up-regulated, while both Cap1p and Cap2p showed a UPR independent localisation change in response to ER-stress, shifting from the bud neck to foci along the cell periphery. Cap1p and Cap2p are respectively the alpha and beta subunits of the actin capping protein that binds to the barbed ends of actin filaments and prevents further polymerization (Kim et al., 2004). Recently DTT treatment was shown to cause actin cytoskeleton depolarisation, and a role for actin filaments was demonstrated in the efficient formation of Ire1p clusters, which may explain the CAP protein localisation shift (Ishiwata-Kimata et al., 2013b). Additionally in the same study Sac6p (fimbrin) an actin-bundling protein was shown to be required for Ire1p clustering. Here we found Sac6p was also up-regulated in a UPR independent manner in response to DTT *but not* TM.

5.7.4 ER-stressed repressed proteins are enriched for ribosomal proteins

There is a general increase in GFP protein fluorescence seen in cells treated with DTT or TM compared to control (Figure 23) probably because of major ER expansion that occurs under these conditions (Bernales et al., 2006). Because of this very few proteins showed GFP down-regulation to levels below unstressed conditions. Therefore we measured proteins that showed significantly less expression than we expected after quantile normalisation and local Z-score calculation and classified these as ER-stress repressed proteins. Looking at the effect of ER-stress on protein repression in either DTT or TM the standout feature is the multitude of ribosomal/translation related proteins that are repressed. 4% and 7% of up-regulated proteins in DTT and TM treatments comprised ribosomal/translation related hits respectively. The repressed proteins by comparison showed a substantial portion of hits to be ribosomal/translation related constituting 37% and 34% of DTT and TM repressed proteins respectively. This repression appears to be UPR independent as *Δhac1* cells treated with DTT showed 54% of repressed hits to be ribosomal/translational related proteins, a significantly larger portion of hits than the UPR functional cells. Additionally, the specifically UPR repressed proteins- that is those repressed by DTT and TM in wild-type cells but not in *Δhac1* cells - showed no enrichment for ribosomal or translational proteins. Defects in the secretory pathway have been well documented in the repression of ribosomal genes in a UPR independent manner (Mizuta and Warner, 1994).

5.8 Conclusion

Here we have described the use of high-content automated microscopy to evaluate the yeast GFP collection, in this instance to characterise the proteomic changes induced by the UPR in response to DTT and TM. Automated microscopy is increasingly being used as a screening technique, particularly with the yeast GFP collection. One of the major bottlenecks in these studies is the manual assessment of localisation changes by inspection. This process is incredibly time consuming considering the tens of thousands of images that need to be inspected. Here we successfully describe the use of automated analysis of textural image features to identify a subset of likely localisation changes. This drastically reduces the number of images that need to be inspected by eye to around 10% of the total collection, increasing throughput and efficiency. Other automation attempts have either been restricted to specific localisation changes, for example nuclear-cytoplasm translocation or localisation to mating projections (Mazumder et al., 2013; Narayanaswamy et al., 2009a). Here we show that texture features can be successfully used to predict numerous localisation changes without *a priori* knowledge. Using this system we identified the localisation changes and protein expression changes induced by the ER-stress agents DTT and TM. We identified a total of 126 localisation changes in either DTT or TM. One of the most interesting findings from these data is that most of these localisation changes occur independently of UPR regulation as they are also evident in *Δhac1* cells. This is in contrast to the expression data which clearly defines a large set of UPR-specific protein up-regulations and thus demonstrates that the UPR is primarily a transcriptional/protein expression response relying only on a limited selection of localisation changes for the response. We show that the use of the yeast GFP collection for protein-level expression measurements compliments more traditional mRNA expression data, and identified an additional 27 novel UPR targets. The studies here also provide an insight into the current limitations of live cell GFP screening techniques which are discussed in detail later.

6 DISCUSSION

The work described throughout this thesis details the development of an automated image recognition procedure for analysing high-throughput image data generated from yeast genomic screens. Specifically these methodologies were applied to measure phenotypes associated with the UPR in yeast using a two-pronged approach. Firstly, the effects of single-gene deletions were assessed for their impact on UPR function using two UPR-specific reporters, namely Ire1p-GFP and UPRE-GFP. This approach allowed us to identify gene deletions that caused an increase in basal UPR expression, as well as gene deletions that caused an attenuation of the UPR under conditions of exogenous ER-stress. Secondly, we used automated image recognition and texture analysis as a means of identifying UPR-induced changes throughout the yeast proteome, including both protein abundance and localization effects. From these data we were able to identify additional novel UPR targets, as well as UPR specific localization changes which had not previously been assessed. Although here we have applied our image analysis methodologies specifically to the study of UPR, these systems have been designed with a broader range of applications in mind allowing flexibility and customization to easily suit future experiments. The implications of our findings and an outlook on the procedures used are discussed in detail below.

6.1 Automated recognition procedures

One of the early frustrations encountered in this work was the challenge of extracting precisely the information desired from high-content imaging data. As humans we heavily rely on visual perception. Thus manually inspecting a set of images for localization patterns provides an immediate sense of any changes. Of course when one increases the scale of an experiment enough, it simply becomes too much to keep up with by manual inspection. The difficulty then becomes translating how we perceive these localisation patterns into a set of computational instructions to develop quantitative image recognition algorithms. Suddenly seemingly obvious phenotypic differences become incredibly difficult to define in absolute terms. Most commercial high-throughput imaging platforms come with some sort of image analysis software, such as the Acapella software used here. These programs tend to provide drag and drop scripting functionality for designing basic procedures, and usually comprise a set of algorithms designed for broad functionality. Although these procedures offer quick script development, they are often not as specific as one would like – a jack of all trades, master of none – so to speak. True customisation requires some level of programming knowledge, something that many end users of high-content microscopy coming from a background in biological sciences do not have. To this end, although the yeast-specific recognition procedures developed here were applied specifically to UPR study, they are also designed to offer future flexibility and customisable options for future users without programming knowledge.

Firstly, a set of marker RFP proteins were chosen to label cells, namely; mCherry (Shaner et al., 2004) as a cytoplasm marker and NLS-RedStar2 (Hodel et al., 2006, 2001; Janke et al., 2004) as a nucleus marker. Both of these proteins are expressed under the control of a constitutive TEF2 promoter to provide uniformity throughout the cell population. These markers were genomically integrated into an SGA starting strain to provide an easy means of introducing the markers into a range of yeast clone collections – namely the DMA and GFP collections – using modified SGA mating procedures. Two versions were created each with different advantages. A dual marker system using both mCherry and NLS-RedStar2 is used for a highly accurate cell recognition procedure that holds up well even in images of high cell density. This is particularly important when measuring localisation based effects, particularly in the case of plasma membrane proteins which are sensitive to inaccuracies in cell border recognition. Secondly, mCherry cytoplasmic labelling can be used alone in cell recognition. This is appropriate in cases such as when measuring cytoplasmic reporter proteins, which are not as sensitive to inaccuracies at the cell border. Additionally the mCherry system only uses up a single antibiotic selection, which can be an important consideration if additional markers or reporters are required that also need antibiotic markers.

In order to make the algorithms designed here malleable to different experimental designs a number of customisable input parameters were included for easy assay optimisation. By default the cell analysis module already measures most common features including; spot detection, fluorescence intensity in various regions (whole cells, nucleus, cytoplasm and membrane), cell size, roundness, and three different measures of texture features. Although these scripts are already optimised for yeast, adjustable parameters include optimisations for variables such as the expected size of cell objects recognised, which may change for example between haploid and diploid cells or if higher resolution images were used. Other useful variables include initial masking threshold adjustments (can account for changes in laser power), the amount of smoothing applied to cell borders, and the minimum distance between two nuclei. There is also an option to display ‘testing images’ depicting each stage of the algorithms, making it easy to understand what different parameters do and what stages may need optimisation. Finally there are also a number of filtering parameters that can be adjusted, including the distance from the image border that cells should be excluded from, the minimum ratio of nucleus/cell size, the number of standard deviations away to define bright or faint outlier cells, maximum and minimum cell size, and the minimum circularity allowed (useful for instance if one were assaying cell shape).

The procedures presented here performed very well when tested when compared to a set of manually drawn cell borders. Over 95% of the automatically identified cells matched with a centre no further than two pixels away from the manually drawn cells, with the average distance less than one pixel. Comparison of cell size showed a very high correlation of 0.912 between manual and automated cells. These statistics compare favourably against other methodologies presented in the literature for which data were available, performing just as well as the recent methodologies presented by Handfield et al. and out performing established procedures available in the well-known open source software CellProfiler (Carpenter et al., 2006; Handfield et al., 2013). In terms of identifying localisation changes, as noted by others manual inspection by the human eye produces miniscule error rates in comparison to previous fully automated techniques (Breker and Schuldiner, 2009; Rimon and Schuldiner, 2011). However, there is simply too much data produced when performing multiple screens to rely on manual inspection alone. There are a number of factors that can affect accuracy in automated procedures including; noise, dead cells, contamination and out of focus objects (Handfield et al., 2013). Because of this there should always be a human component in verifying localisation changes. Here we take a hybrid approach to localisation analysis where we first use texture analysis to identify a set of likely localisation changes, reducing the amount of images to be manually inspected by 90%. Using manual inspection as a final quality control measure allows for easy removal of false positives and a higher reliably. We found our automated procedures to perform well, and correctly encompassed all localisation changes identified by complete manual inspection of the dataset.

In terms of future directions for the automated recognition analysis presented here, there are a number of performance enhancements that could be included in future versions. Firstly moving to an open source platform would make these procedures more widely distributable and applicable to other systems. Additionally, producing a compiled version of the software rather than running in an interpreted scripting language would provide significant speed improvements to the analysis. There are a number of additional features that could prove useful in terms of cell recognition performance. The total number of cells identified could be increased in a number of ways. Firstly, the optimal mid-section selection could be applied per cell as opposed to selecting an entire image. This would be particularly useful in the dual RFP system as the nucleus provides a very clear optimal midsection. Additionally the cytoplasm only procedures could potentially be optimised for more accurate cell separation by using data from non-midsection images. As the focal plane moves away from the mid-section cells appear smaller and more separated, albeit less well defined as you approach the top or bottom of the cell. Although these peripheral images cannot provide additional cell boundary information, they could be used to define object centres as surrogates for nuclei staining – from which the better performing dual marker algorithms could be used.

There are also a number of analysis techniques used in other studies that could improve recognition of protein localisation and/or abundance changes. Firstly, one of the things we noticed in some cases was distinct populations of responders within the total cell population, particularly in cytoplasm-nucleus translocations. By including testing for bi-modal populations we may be better able to identify cases such as these, particularly if they are less distinct (Breker et al., 2013). Similarly including cell-cycle stage, or mother-bud analysis may identify a relationship to bi-modal expression patterns (Handfield et al., 2013). Finally we could further optimise the texture recognition procedures to be more yeast specific and include a number of more interpretable measurements, an approach that has proved useful for others and could further reduce the number of images to be manually inspected by eye (Dénervaud et al., 2013; Handfield et al., 2013).

6.2 Characterising UPR phenotypes

6.2.1 Deletion array screens and GFP screens reveal different aspects of the UPR

We applied our automated cell recognition procedures in a dual strategy to identify proteins with UPR function. Firstly we measured UPR-specific reporters to screen for effects on UPR activation as consequences of non-essential gene deletions. This was followed up with a proteome-wide GFP collection screen to identify UPR specific protein abundance and localisation changes. These approaches complement each other well. The DMA represents a collection of loss-of-function mutants and as such was used to identify genes required for normal UPR function in two ways. Firstly gene deletions that induce UPR activation are likely to be involved in the normal functioning of protein folding and/or ER homeostasis. Secondly, gene deletions that prevent the normal activation of UPR in response to ER-stress are likely to be involved in the UPR or maintenance of the response itself. To investigate these scenarios we used two GFP reporter assays; the first was Ire1p-GFP foci formation which measures early misfolded protein recognition events of the UPR, the second reporter was UPRE-GFP expression which measures downstream UPR activation. The GFP collection on the other hand represents a set of functional proteins for which we can measure abundance and localisation changes. Thus this collection was used to identify UPR specific targets and additional down-stream effects.

These differences between the DMA and GFP approaches are also reflected in our results. The UPR inducers, whether Ire1-GFP cluster or UPRE expression inducing, were enriched for genes involved in a number of ER and protein folding related processes. In particular protein glycosylation and ERAD functions. It is interesting to note the lack of oxidative protein folding genes here but this may be due to the essentiality of this process, for which the key members are all essential genes whose deletions are not found in the DMA (e.g. PDI1 and ERO1). What is important to note however is that of the gene deletions inducing UPR activation, only a small portion, ~13%, are UPR targets as defined by Kimata et al., 2006; Travers et al., 2000 or our results described here. This is in stark contrast to the results from our GFP collection screen for which 61% of all the ER-stress specific GFP changes (i.e. occurring in both TM and DTT treatment) were also UPR specific effects (i.e. they did not occur in Δhac1 cells).

6.2.2 GFP screening reveals protein-level UPR effects

Less than half of the UPR-targets we identified in our GFP collection screens were previously known targets as measured at the mRNA level (Kimata et al., 2006; Travers et al., 2000). It is interesting to note this distinction between the studies as Travers et al. and Kimata et al. measured UPR targets at the transcript level, whereas here we have measured effects at the protein level. These results indicate two things; that regulation of at least some of these UPR targets does not occur at the mRNA level, and that not all UPR induced transcript changes result in a significant change at the protein level. This may be due to differences in post-UPR regulation or differential protein functional requirements, i.e. minor increases in some proteins can have a dramatic effect, whereas other functions may require a significantly larger increase in protein levels. Furthermore we identified a number of possible transcription factors that may function alongside Hac1p upon UPR activation. It would be interesting to further explore the possible connection of these in the response, and likewise to see if deletion mutants lacking these TFs had a significant impact on UPR target regulation particularly of UPR targets lacking any known UPRE elements.

6.2.3 Ire1p activation appears to be a committal event

Interestingly early UPR activation in response to ER-stress appears to be a committal event in terms of Ire1p foci formation. We could not measure any significant increase in the number of foci per cell in response to DTT induced ER-stress. However, the total number of cells within the population displaying foci increased in a dose dependant manner. Presumably all cells are undergoing protein misfolding in the relatively high 2 mM DTT concentration used, yet not all cells commit to UPR activation. It would be interesting to explore any relationship between this UPR committal event and cell cycle stage or any other commonality between the sub-population of non-responders. Alternatively a vital stain would show if these cells had undergone cell death instead. It is interesting that 10 of the 13 gene deletion strains we found to prevent normal Ire1p cluster formation also induced UPRE expression. Our interpretation of this is that these gene deletions are probably priming the UPR in these cells with some level of chronic misfolding. Further ER-stress does not then result in Ire1p foci formation which is usually only a transient event. What remains unclear is why not all UPRE inducers cause a repression of Ire1p foci formation. Possibly the type of ER-stress these deletions cause, or the strength of the chronic UPR are factors in this. On the other side of the UPR time scale it may be interesting to investigate events at the end of UPR activation. Ire1 foci rapidly disassociate within 15 minutes after removal of ER-stress, the mechanism of which is currently unclear (Kimata et al., 2007). Thus it would be interesting to assay for any gene deletions preventing the normal dissociation of Ire1p foci.

6.2.4 Downstream UPR activity is blocked by SAGA complex components

The set of gene deletion strains that suppressed the normal UPRE expression in response to ER-stress was enriched for a number of chromatin and transcription related genes. Specifically four of these hits are components of the SAGA histone acetyltransferase complex (CHD1, SGF29, SPT3 and SPT8) which has previously been shown to play a role in the UPR (Welihinda et al., 1997, 2000). As expected these genes were not enriched in the list of Ire1p foci preventers, as these genes are acting at a translation level downstream of Ire1p. However, what is interesting is that deletion of SGF29, a component of the SAGA core histone acetyltransferase module, instead resulted in an induction of Ire1p foci. One possibility here is that blocking downstream Hac1p transcription activity is preventing the UPR from dealing with basal levels of protein misfolding and hence resulting in enough build-up to induce Ire1p foci formation. It would also be interesting to see how much of the UPR is blocked by these mutations, for instance are UPR target genes that contain the UPRE2 or UPRE3 upstream activator sequences, as opposed to the UPRE1 used here, also inhibited in the same manner?

6.3 Limitations and future directions

6.3.1 DMA screening

Essential Genes

One of the major short comings of the R-SGA approach is that essential genes are not present in the collection, leaving around 20% of the genome unevaluated. Potentially there are genes of major implication here – although given their essential status these are likely to be genes required for fundamental processes rather than anything specifically of the UPR. However, this is not to suggest essential genes are unlikely to contribute to UPR function, in fact a good example of an essential process impacting on UPR function is the requirement for F-actin and type-II myosin in efficient Ire1p foci formation. Recently Ishiwata-Kimata et al. showed that the actin disrupting agent latrunculin-A prevents Ire1p foci formation upon DTT stress. Additionally the type-II myosin gene MYO1 – which is not in our DMA array – is required for Ire1p foci (Ishiwata-Kimata et al., 2013b). The DMA collection we used here has also had slow growing strains removed that can be problematic for traditional SGA analysis. However, these strains could potentially reveal interesting results, clearly these deletions have some consequence based on the growth defect. Unlike essential genes we can probe the effect of a complete knock-out mutation in these strains. Again an example can be seen in the work by Ishiwata-Kimata and colleagues who also showed that deletion of the actin bundling gene SAC6 – a strain removed from our DMA because of slow growth – additionally prevented Ire1p foci formation. These findings suggest that it would be worthwhile to further screen the slow growing strains, as well as the essential genes using either the temperature-sensitive or DAmP hypomorphic strain collections available (Ben-Aroya et al., 2008; Breslow et al., 2008; Li et al., 2011). Fluorescent protein reporters have previously been successful for identifying phenotypes associated with these conditional essential gene alleles validating the approach (Herzig et al., 2012; Li et al., 2011).

Optimal growth conditions for early UPR events

Although here we used a four hour incubation and drug treatment time to be consistent across all screens, it would be interesting to also investigate the UPR activation events in a shorter time frame. Particularly for the early UPR activation events measured by Ire1p foci activity, which unlike downstream UPRE activity is not limited by protein synthesis rates. Ire1p clustering for instance has been shown to occur as rapidly as 6 minutes (Aragón et al., 2009). This may present different early UPR dynamics between deletion mutants, for instance some strains may take longer to form foci that others. Additionally it would be interesting to find conditions where Hac1p translocation to the nucleus was a measurable phenotype. The problem we encountered was that control cells also showed

Hac1p translocation to the nucleus, presumably because of protein synthesis rates in actively growing cells produces high enough levels of misfolded proteins to induce low levels of UPR. It may be possible that using stationary phase cells or reducing temperature to slow down cell growth rates could be a way around this problem. This would provide potentially interesting information on the dynamics between early Ire1p activation and Hac1p splicing events.

6.3.2 GFP set screening

C-terminal tagging and possible protein disruptions

One of the limitations of using C-terminal GFP tagging is that it intrinsically removes any proteins dependant on the C-terminus for proper function. This includes proteins that contain HDEL-ER retention signals, which naturally are of interest when studying ER function. In fact of the 93 strains systematically removed from our analysis, 13 had previously been defined as UPR targets including key players such as the chaperones Kar2p and Lhs1p as well as the protein disulphide isomerases Pdi1p, Mpd1p and Mpd2p required for oxidative protein folding. As an alternative, using N-terminal GFP fusions in these cases is an appealing option (Prein et al., 2000). Additionally this strategy could be applied to the proteins completely missing from the commercial yeast GFP-collection that did not produce notable GFP fluorescence with C-terminal tagging (Huh et al., 2003).

Proteomic context of biological changes

One of the interesting points to note is the systematic increase in protein abundance upon ER-stress. This is unlikely to be a consequence of drug induced changes to GFPs fluorescence properties as both DTT and TM caused an increase and other data from our lab showed that overexpression of a misfolded protein causes the same general increase in abundance (Low, 2013). This raises a few points of intriguing significance, particularly as to the cause of the overall increased protein abundance. This seems counterintuitive given that we saw a specific repression of ribosomal and translation related proteins. However one possible explanation is that rather than being an overall increase in protein expression we are instead seeing the consequences of decreased protein turnover, possibly due to the proteasome being overwhelmed with ERAD substrates as a consequence of ER-stress and misfolding. This also brings up a general point worth discussing given the number of proteins with increased abundance, particularly after DTT treatment. If we had used a standard Z-score calculation rather than the local Z-score that takes into account intensity effects, almost 25% of proteins in the GFP collection would have been considered upregulated at the same significance level (data not shown). This highlights an important feature of proteome-wide screens in which the degree of abundance change can be taken in context of the whole proteome. Using a local Z-score we found closer to 5% of proteins to actually be significantly upregulated from DTT treatment in context of the whole proteome. This highlights an aspect of improved statistical strength in whole proteome screening, where the entire proteome is effectively acting as a control. In contrast if we had cherry picked only a small collection of proteins for measurement many of them would have shown individually significant increases with strong p-values, underscoring the importance of control selection in such cases. Along this same line of thought, it is important to consider the biological context of localisation changes in proteome screens. For instance a sizeable portion of the ER-stress induced localisation changes were cell periphery proteins that also localised internally within the cell. It is unclear whether such changes have specific functional significance or are instead a consequence of reduced secretory pathway function. Further to this, it can be incredibly difficult interpreting which localisation hits are most relevant. Take the example of ire1 which has a major role in activating the UPR, yet due to low protein abundance the localisation effect is far more subtle than the majority of other changes identified. This is in contrast to measuring protein abundance or a single specific reporter for which there is a comparable quantitative component to the measurements. One final note of interest is the dominance of UPR induced abundance changes in comparison to the number of UPR specific localisation changes. This might be explained by the fact that the UPR is a transcriptional response and hence localisation changes are likely to be down-stream effects from up-regulation of primary UPR targets.

Separating the UPR specific response from general ER-stress consequences.

It is clear that there are a number of physiological changes that are a consequence of ER-stress rather than just UPR induced changes. Although here we attempt to separate the two by comparison to a Δhac1 strain lacking a functional UPR, it is appears that many

‘UPR specific’ changes are likely to involve complex downstream regulation and/or interactions amongst UPR targets. This is particularly apparent when considering that most known UPR targets do not contain a recognisable UPRE sequence (Patil et al., 2004), and that many of the UPR targets we identified are not known to be regulated at the transcript level (Kimata et al., 2006; Travers et al., 2000). It would thus be interesting to attempt to totally separate UPR activity from other ER-stress consequences, possibly using a similar system to Kimata et al. where they used a constitutively activated $HAC1^i$ mutant, rather than just treating with ER stress agents (Kimata et al., 2006). Although constitutively active UPR strains can be problematic to maintain due to slow growth and spontaneous loss of function mutations at the $HAC1^i$ locus (Y. Kimata, personal communication September 2012) an appealing option would instead be to use a $HAC1^i$ allele under control of an inducible promoter to activate the UPR, without the detrimental consequences of inducing ER-stress.

CONCLUSION

In conclusion we have described an efficient framework for automated image analysis of large high-throughput microscopy datasets. In the context of this thesis we have applied these methods to the characterisation of phenotypic changes induced by the UPR, and in the process identified a number of new UPR target proteins. We have also separated the genes required for the early events in the UPR from the later ones involving UPR targets. In the broader scheme this study highlights the usefulness of high-content imaging and GFP based proteomic screening as a systems biology tool, particularly when used in combination with traditional deletion mutant array screening. With the advances currently occurring in automating both image acquisition and data analysis this field will continue to grow and undoubtedly provide a wealth of information. It is foreseeable that as the number of such studies increases more and more image data will be compiled into online databases as traditional growth data has, and potentially lead to image-based searching or clustering methods for drug mode of action screening and/or characterisation of novel gene functions.

REFERENCES

Abdulrehman, D., Monteiro, P.T., Teixeira, M.C., Mira, N.P., Lourenço, A.B., Santos, S.C. dos, Cabrito, T.R., Francisco, A.P., Madeira, S.C., Aires, R.S., et al. (2011). YEASTRACT: providing a programmatic access to curated transcriptional regulatory associations in Saccharomyces cerevisiae through a web services interface. Nucleic Acids Res. *39*, D136–D140.

Abramoff, M.D., Magalhães, P.J., and Ram, S.J. (2004). Image processing with ImageJ. Biophotonics Int. *11*, 36–42.

Aebi, M., Bernasconi, R., Clerc, S., and Molinari, M. (2010). N-glycan structures: recognition and processing in the ER. Trends Biochem. Sci. *35*, 74–82.

Alexandre, H., Ansanay-Galeote, V., Dequin, S., and Blondin, B. (2001). Global gene expression during short-term ethanol stress in Saccharomyces cerevisiae. FEBS Lett. *498*, 98–103.

Amberg, D., Burke, D., and Strathern, J. (2005). Methods in Yeast Genetics: A Cold Spring Harbor Laboratory Course Manual, 2005 Edition (Cold Spring) ({Cold Spring Harbor Laboratory Press}).

Appenzeller-Herzog, C. (2011). Glutathione- and non-glutathione-based oxidant control in the endoplasmic reticulum. J. Cell Sci. *124*, 847–855.

Aragón, T., van Anken, E., Pincus, D., Serafimova, I.M., Korennykh, A.V., Rubio, C.A., and Walter, P. (2009). Messenger RNA targeting to endoplasmic reticulum stress signalling sites. Nature *457*, 736–740.

Ben-Aroya, S., Coombes, C., Kwok, T., O’Donnell, K.A., Boeke, J.D., and Hieter, P. (2008). Toward a Comprehensive Temperature-Sensitive Mutant Repository of the Essential Genes of Saccharomyces cerevisiae. Mol. Cell *30*, 248–258.

Bader, M., Muse, W., Ballou, D.P., Gassner, C., and Bardwell, J.C.A. (1999). Oxidative Protein Folding Is Driven by the Electron Transport System. Cell *98*, 217–227.

Balakrishnan, R., Park, J., Karra, K., Hitz, B.C., Binkley, G., Hong, E.L., Sullivan, J., Micklem, G., and Michael Cherry, J. (2012). YeastMine--an integrated data warehouse for Saccharomyces cerevisiae data as a multipurpose tool-kit. Database *2012*, bar062–bar062.

Banta, L.M., Robinson, J.S., Klionsky, D.J., and Emr, S.D. (1988). Organelle assembly in yeast: characterization of yeast mutants defective in vacuolar biogenesis and protein sorting. J. Cell Biol. *107*, 1369–1383.

Barlowe, C.K., and Miller, E.A. (2013). Secretory Protein Biogenesis and Traffic in the Early Secretory Pathway. Genetics *193*, 383–410.

Battle, A., Jonikas, M.C., Walter, P., Weissman, J.S., and Koller, D. (2010). Automated identification of pathways from quantitative genetic interaction data. Mol. Syst. Biol. *6*.

Becker, J., Walter, W., Yan, W., and Craig, E.A. (1996). Functional interaction of cytosolic hsp70 and a DnaJ-related protein, Ydj1p, in protein translocation in vivo. Mol. Cell. Biol. *16*, 4378–4386.

Belgareh-Touzé, N., Léon, S., Erpapazoglou, Z., Stawiecka-Mirota, M., Urban-Grimal, D., and Haguenauer-Tsapis, R. (2008). Versatile role of the yeast ubiquitin ligase Rsp5p in intracellular trafficking. Biochem. Soc. Trans. *36*, 791.

Bernales, S., McDonald, K.L., and Walter, P. (2006). Autophagy Counterbalances Endoplasmic Reticulum Expansion during the Unfolded Protein Response. PLoS Biol. *4*, e423.

Bertolotti, A., Zhang, Y., Hendershot, L.M., Harding, H.P., and Ron, D. (2000). Dynamic interaction of BiP and ER stress transducers in the unfolded-protein response. Nat. Cell Biol. *2*, 326–332.

Bevis, B.J., and Glick, B.S. (2002). Rapidly maturing variants of the Discosoma red fluorescent protein (DsRed). Nat. Biotechnol. *20*, 83–87.

Bircham, P.W., Maass, D.R., Roberts, C.A., Kiew, P.Y., Low, Y.S., Yegambaram, M., Matthews, J., Jack, C.A., and Atkinson, P.H. (2011). Secretory pathway genes assessed by high-throughput microscopy and synthetic genetic array analysis. Mol BioSyst *7*, 2589–2598.

Böhni, P.C., Deshaies, R.J., and Schekman, R.W. (1988). SEC11 is required for signal peptide processing and yeast cell growth. J. Cell Biol. *106*, 1035–1042.

Bolstad, B.M., Irizarry, R.A., Åstrand, M., and Speed, T.P. (2003). A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. Bioinformatics *19*, 185–193.

Boone, C., Bussey, H., and Andrews, B.J. (2007). Exploring genetic interactions and networks with yeast. Nat. Rev. Genet. *8*, 437–449.

Boorsma, A., Nobel, H. de, Riet, B. ter, Bargmann, B., Brul, S., Hellingwerf, K.J., and Klis, F.M. (2004). Characterization of the transcriptional response to cell wall stress in Saccharomyces cerevisiae. Yeast *21*, 413–427.

Botstein, D., and Fink, G.R. (2011). Yeast: An Experimental Organism for 21st Century Biology. Genetics *189*, 695–704.

Botstein, D., Chervitz, S.A., and Cherry, J.M. (1997). Yeast as a Model Organism. Science *277*, 1259.

Breker, M., and Schuldiner, M. (2009). Explorations in topology–delving underneath the surface of genetic interaction maps. Mol. Biosyst. *5*, 1473.

Breker, M., Gymrek, M., and Schuldiner, M. (2013). A novel single-cell screening platform reveals proteome plasticity during yeast stress responses. J. Cell Biol. *200*, 839–850.

Breslow, D.K., Cameron, D.M., Collins, S.R., Schuldiner, M., Stewart-Ornstein, J., Newman, H.W., Braun, S., Madhani, H.D., Krogan, N.J., and Weissman, J.S. (2008). A comprehensive strategy enabling high-resolution functional analysis of the yeast genome. Nat. Methods *5*, 711–718.

Brodsky, J.L., and Schekman, R. (1993). A Sec63p-BiP complex from yeast is required for protein translocation in a reconstituted proteoliposome. J. Cell Biol. *123*, 1355–1363.

Brown, C.R., Hong-Brown, L.Q., Biwersi, J., Verkman, A.S., and Welch, W.J. (1996). Chemical chaperones correct the mutant phenotype of the ΔF508 cystic fibrosis transmembrane conductance regulator protein. Cell Stress Chaperones *1*, 117.

Burston, H.E., Maldonado-Báez, L., Davey, M., Montpetit, B., Schluter, C., Wendland, B., and Conibear, E. (2009). Regulators of yeast endocytosis identified by systematic quantitative analysis. J. Cell Biol. *185*, 1097–1110.

Butler, G.S., and Overall, C.M. (2009). Proteomic identification of multitasking proteins in unexpected locations complicates drug targeting. Nat. Rev. Drug Discov. *8*, 935–948.

Campbell, R.E., Tour, O., Palmer, A.E., Steinbach, P.A., Baird, G.S., Zacharias, D.A., and Tsien, R.Y. (2002). A monomeric red fluorescent protein. Proc. Natl. Acad. Sci. *99*, 7877–7882.

Caramelo, J.J., and Parodi, A.J. (2008). Getting In and Out from Calnexin/Calreticulin Cycles. J. Biol. Chem. *283*, 10221–10225.

Carbon, S., Ireland, A., Mungall, C.J., Shu, S., Marshall, B., and Lewis, S. (2009). AmiGO: online access to ontology and annotation data. Bioinformatics *25*, 288–289.

Carpenter, A.E., Jones, T.R., Lamprecht, M.R., Clarke, C., Kang, I.H., Friman, O., Guertin, D.A., Chang, J.H., Lindquist, R.A., Moffat, J., et al. (2006). CellProfiler: image analysis software for identifying and quantifying cell phenotypes. Genome Biol. *7*, R100.

Carvalho, P., Goder, V., and Rapoport, T.A. (2006). Distinct Ubiquitin-Ligase Complexes Define Convergent Pathways for the Degradation of ER Proteins. Cell *126*, 361–373.

De Carvalho, M.A.G., Lotufo, R. de A., and Couprie, M. (2007). Morphological segmentation of yeast by image analysis. Image Vis. Comput. *25*, 34–39.

Chakravarthi, S., and Bulleid, N.J. (2004). Glutathione Is Required to Regulate the Formation of Native Disulfide Bonds within Proteins Entering the Secretory Pathway. J. Biol. Chem. *279*, 39872–39879.

Chakravarthi, S., Jessop, C.E., and Bulleid, N.J. (2006). The role of glutathione in disulphide bond formation and endoplasmic-reticulum-generated oxidative stress. EMBO Rep. *7*, 271–275.

Chalfie, M., Tu, Y., Euskirchen, G., Ward, W.W., and Prasher, D.C. (1994). Green fluorescent protein as a marker for gene expression. Science *263*, 802–805.

Chen, S.-C., Zhao, T., Gordon, G.J., and Murphy, R.F. (2007). Automated image analysis of protein localization in budding yeast. Bioinformatics *23*, i66–i71.

Chen, X., VanValkenburgh, C., Liang, H., Fang, H., and Green, N. (2001). Signal Peptidase and Oligosaccharyltransferase Interact in a Sequential and Dependent Manner within the Endoplasmic Reticulum. J. Biol. Chem. *276*, 2411–2416.

Chen, X., Velliste, M., Weinstein, S., Jarvik, J.W., and Murphy, R.F. (2003). Location proteomics: building subcellular location trees from high-resolution 3D fluorescence microscope images of randomly tagged proteins. In Biomedical Optics 2003, pp. 298–306.

Cherry, J.M., Hong, E.L., Amundsen, C., Balakrishnan, R., Binkley, G., Chan, E.T., Christie, K.R., Costanzo, M.C., Dwight, S.S., Engel, S.R., et al. (2012). Saccharomyces Genome Database: the genomics resource of budding yeast. Nucleic Acids Res. *40*, D700–D705.

Chong, Y.T., Cox, M.J., and Andrews, B. (2012). Proteome-Wide Screens in Saccharomyces cerevisiae Using the Yeast GFP Collection. In Advances in Systems Biology, I.I. Goryanin, and A.B. Goryachev, eds. (New York, NY: Springer New York), pp. 169–178.

Chudakov, D.M., Matz, M.V., Lukyanov, S., and Lukyanov, K.A. (2010). Fluorescent Proteins and Their Applications in Imaging Living Cells and Tissues. Physiol. Rev. *90*, 1103–1163.

Clerc, S., Hirsch, C., Oggier, D.M., Deprez, P., Jakob, C., Sommer, T., and Aebi, M. (2009). Htm1 protein generates the N-glycan signal for glycoprotein degradation in the endoplasmic reticulum. J. Cell Biol. *184*, 159–172.

Cleveland, W.S., and Devlin, S.J. (1988). Locally Weighted Regression: An Approach to Regression Analysis by Local Fitting. J. Am. Stat. Assoc. *83*, 596–610.

Cochella, L., and Green, R. (2005). Fidelity in protein synthesis. Curr. Biol. *15*, R536–R540.

Colantuoni, C., Henry, G., Zeger, S., and Pevsner, J. (2002). SNOMAD (Standardization and NOrmalization of MicroArray Data): web-accessible gene expression data analysis. Bioinformatics *18*, 1540–1541.

Cooper, A.A., Gitler, A.D., Cashikar, A., Haynes, C.M., Hill, K.J., Bhullar, B., Liu, K., Xu, K., Strathearn, K.E., Liu, F., et al. (2006). α-Synuclein Blocks ER-Golgi Traffic and Rab1 Rescues Neuron Loss in Parkinson’s Models. Science *313*, 324–328.

Čopič, A., Dorrington, M., Pagant, S., Barry, J., Lee, M.C.S., Singh, I., Hartman, J.L., and Miller, E.A. (2009). Genomewide Analysis Reveals Novel Pathways Affecting Endoplasmic Reticulum Homeostasis, Protein Modification and Quality Control. Genetics *182*, 757–769.

Copley, S.D. (2012). Moonlighting is mainstream: Paradigm adjustment required. BioEssays *34*, 578–588.

Cormack, B.P., Valdivia, R.H., and Falkow, S. (1996). FACS-optimized mutants of the green fluorescent protein (GFP). Gene *173*, 33–38.

Costanzo, M., Baryshnikova, A., Bellay, J., Kim, Y., Spear, E.D., Sevier, C.S., Ding, H., Koh, J.L.Y., Toufighi, K., Mostafavi, S., et al. (2010). The Genetic Landscape of a Cell. Science *327*, 425–431.

Cox, J.S., and Walter, P. (1996). A Novel Mechanism for Regulating Activity of a Transcription Factor That Controls the Unfolded Protein Response. Cell *87*, 391–404.

Cox, J.S., Shamu, C.E., Walter, P., and others (1993). Transcriptional induction of genes encoding endoplasmic reticulum resident proteins requires a transmembrane protein kinase. Cell *73*, 1197–1206.

Credle, J.J., Finer-Moore, J.S., Papa, F.R., Stroud, R.M., and Walter, P. (2005). On the mechanism of sensing unfolded protein in the endoplasmic reticulum. Proc. Natl. Acad. Sci. U. S. A. *102*, 18773–18784.

Cullen, P.J., Xu-Friedman, R., Delrow, J., and Sprague, G.F. (2006). Genome-wide analysis of the response to protein glycosylation deficiency in yeast. FEMS Yeast Res. *6*, 1264–1273.

Cuozzo, J.W., and Kaiser, C.A. (1999). Competition between glutathione and protein thiols for disulphide-bond formation. Nat. Cell Biol. *1*, 130–135.

Delic, M., Rebnegger, C., Wanka, F., Puxbaum, V., Haberhauer-Troyer, C., Hann, S., Köllensperger, G., Mattanovich, D., and Gasser, B. (2012). Oxidative protein folding and unfolded protein response elicit differing redox regulation in endoplasmic reticulum and cytosol of yeast. Free Radic. Biol. Med. *52*, 2000–2012.

Delic, M., Valli, M., Graf, A.B., Pfeffer, M., Mattanovich, D., and Gasser, B. (2013). The secretory pathway: exploring yeast diversity. FEMS Microbiol. Rev. n/a–n/a.

Dénervaud, N., Becker, J., Delgado-Gonzalo, R., Damay, P., Rajkumar, A.S., Unser, M., Shore, D., Naef, F., and Maerkl, S.J. (2013). A chemostat array enables the spatio-temporal analysis of the yeast proteome. Proc. Natl. Acad. Sci. *110*, 15842–15847.

Dolinski, K., and Botstein, D. (2007). Orthology and functional conservation in eukaryotes. Annu Rev Genet *41*, 465–507.

Dudoit, S., Yang, Y.H., Callow, M.J., and Speed, T.P. (2002). Statistical methods for identifying differentially expressed genes in replicated cDNA microarray experiments. Stat. Sin. *12*, 111–140.

Dyavaiah, M., Rooney, J.P., Chittur, S.V., Lin, Q., and Begley, T.J. (2011). Autophagy-Dependent Regulation of the DNA Damage Response Protein Ribonucleotide Reductase 1. Mol. Cancer Res. *9*, 462–475.

Ellis, C.D., Wang, F., MacDiarmid, C.W., Clark, S., Lyons, T., and Eide, D.J. (2004). Zinc and the Msc2 zinc transporter protein are required for endoplasmic reticulum function. J. Cell Biol. *166*, 325–335.

Ellis, C.D., MacDiarmid, C.W., and Eide, D.J. (2005). Heteromeric Protein Complexes Mediate Zinc Transport into the Secretory Pathway of Eukaryotic Cells. J. Biol. Chem. *280*, 28811–28818.

Farquhar, R., Honey, N., Murant, S.J., Bossier, P., Schultz, L., Montgomery, D., Ellis, R.W., Freedman, R.B., and Tuite, M.F. (1991). Protein disulfide isomerase is essential for viability in Saccharomyces cerevisiae. Gene *108*, 81–89.

Fassio, A., and Sitia, R. (2002). Formation, isomerisation and reduction of disulphide bonds during protein quality control in the endoplasmic reticulum. Histochem. Cell Biol. *117*, 151–157.

Fernández, F.S., Trombetta, S.E., Hellman, U., and Parodi, A.J. (1994). Purification to homogeneity of UDP-glucose:glycoprotein glucosyltransferase from Schizosaccharomyces pombe and apparent absence of the enzyme from Saccharomyces cerevisiae. J. Biol. Chem. *269*, 30701–30706.

Fernandez-Ricaud, L., Warringer, J., Ericson, E., Pylvänäinen, I., Kemp, G.J.L., Nerman, O., and Blomberg, A. (2005). PROPHECY—a database for high-resolution phenomics. Nucleic Acids Res. *33*, D369–D373.

Fernandez-Ricaud, L., Warringer, J., Ericson, E., Glaab, K., Davidsson, P., Nilsson, F., Kemp, G.J.L., Nerman, O., and Blomberg, A. (2007). PROPHECY—a yeast phenome database, update 2006. Nucleic Acids Res. *35*, D463–D467.

Fillingham, J., Kainth, P., Lambert, J.-P., van Bakel, H., Tsui, K., Peña-Castillo, L., Nislow, C., Figeys, D., Hughes, T.R., Greenblatt, J., et al. (2009). Two-Color Cell Array Screen Reveals Interdependent Roles for Histone Chaperones and a Chromatin Boundary Regulator in Histone Gene Repression. Mol. Cell *35*, 340–351.

Finke, K., Plath, K., Panzner, S., Prehn, S., Rapoport, T.A., Hartmann, E., and Sommer, T. (1996). A second trimeric complex containing homologs of the Sec61p complex functions in protein transport across the ER membrane of S. cerevisiae. EMBO J. *15*, 1482.

Frand, A.R., and Kaiser, C.A. (1998). The ERO1 Gene of Yeast Is Required for Oxidation of Protein Dithiols in the Endoplasmic Reticulum. Mol. Cell *1*, 161–170.

Frand, A.R., and Kaiser, C.A. (1999). Ero1p Oxidizes Protein Disulfide Isomerase in a Pathway for Disulfide Bond Formation in the Endoplasmic Reticulum. Mol. Cell *4*, 469–477.

Fredrickson, E.K., and Gardner, R.G. (2012). Selective destruction of abnormal proteins by ubiquitin-mediated protein quality control degradation. Semin. Cell Dev. Biol. *23*, 530–537.

Gabor, D. (1946). Theory of communication. Part 1: The analysis of information. J. Inst. Electr. Eng. - Part III Radio Commun. Eng. *93*, 429–441.

Gancedo, C., and Flores, C.-L. (2008). Moonlighting Proteins in Yeasts. Microbiol. Mol. Biol. Rev. *72*, 197–210.

García, R., Rodríguez-Peña, J.M., Bermejo, C., Nombela, C., and Arroyo, J. (2009). The High Osmotic Response and Cell Wall Integrity Pathways Cooperate to Regulate Transcriptional Responses to Zymolyase-induced Cell Wall Stress in Saccharomyces cerevisiae. J. Biol. Chem. *284*, 10901–10911.

Gardner, B.M., and Walter, P. (2011). Unfolded Proteins Are Ire1-Activating Ligands That Directly Induce the Unfolded Protein Response. Science *333*, 1891–1894.

Gasch, A.P., Spellman, P.T., Kao, C.M., Carmel-Harel, O., Eisen, M.B., Storz, G., Botstein, D., and Brown, P.O. (2000). Genomic Expression Programs in the Response of Yeast Cells to Environmental Changes. Mol. Biol. Cell *11*, 4241–4257.

Gauss, R., Jarosch, E., Sommer, T., and Hirsch, C. (2006). A complex of Yos9p and the HRD ligase integrates endoplasmic reticulum quality control into the degradation machinery. Nat. Cell Biol. *8*, 849–854.

Geda, P., Patury, S., Ma, J., Bharucha, N., Dobry, C.J., Lawson, S.K., Gestwicki, J.E., and Kumar, A. (2008). A small molecule-directed approach to control protein localization and function. Yeast *25*, 577–594.

Ghaemmaghami, S., Huh, W.-K., Bower, K., Howson, R.W., Belle, A., Dephoure, N., O’Shea, E.K., and Weissman, J.S. (2003). Global analysis of protein expression in yeast. Nature *425*, 737–741.

Giaever, G., Chu, A.M., Ni, L., Connelly, C., Riles, L., Véronneau, S., Dow, S., Lucau-Danila, A., Anderson, K., André, B., et al. (2002). Functional profiling of the Saccharomyces cerevisiae genome. Nature *418*, 387–391.

Gietz, R.D., and Schiestl, R.H. (2007). High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. Nat. Protoc. *2*, 31–34.

Gilbert, D.G. (2002). euGenes: a eukaryote genome information system. Nucleic Acids Res. *30*, 145–148.

Gilstring, C.F., Melin-Larsson, M., and Ljungdahl, P.O. (1999). Shr3p Mediates Specific COPII Coatomer–Cargo Interactions Required for the Packaging of Amino Acid Permeases Into ER-derived Transport Vesicles. Mol. Biol. Cell *10*, 3549–3565.

Goffeau, A., Barrell, B.G., Bussey, H., Davis, R.W., Dujon, B., Feldmann, H., Galibert, F., Hoheisel, J.D., Jacq, C., Johnston, M., et al. (1996). Life with 6000 Genes. Science *274*, 546–567.

Gonzalez, T.N., Sidrauski, C., Dörfler, S., and Walter, P. (1999). Mechanism of non-spliceosomal mRNA splicing in the unfolded protein response pathway. EMBO J. *18*, 3119–3132.

Gordon, A., Colman-Lerner, A., Chin, T.E., Benjamin, K.R., Yu, R.C., and Brent, R. (2007). Single-cell quantification of molecules and rates using open-source microscope-based cytometry. Nat. Methods *4*, 175–181.

Guarente, L., and Mason, T. (1983). Heme regulates transcription of the CYC1 gene of S. cerevisiae via an upstream activation site. Cell *32*, 1279–1286.

Gygi, S.P., Rochon, Y., Franza, B.R., and Aebersold, R. (1999). Correlation between Protein and mRNA Abundance in Yeast. Mol. Cell. Biol. *19*, 1720–1730.

Hamilton, N.A., Pantelic, R.S., Hanson, K., and Teasdale, R.D. (2007). Fast automated cell phenotype image classification. BMC Bioinformatics *8*, 110.

Handfield, L.-F., Chong, Y.T., Simmons, J., Andrews, B.J., and Moses, A.M. (2013). Unsupervised Clustering of Subcellular Protein Expression Patterns in High-Throughput Microscopy Images Reveals Protein Complexes and Functional Relationships between Proteins. PLoS Comput Biol *9*, e1003085.

Hanssum, A., Zhong, Z., Rousseau, A., Krzyzosiak, A., Sigurdardottir, A., and Bertolotti, A. (2014). An Inducible Chaperone Adapts Proteasome Assembly to Stress. Mol. Cell *DOI: 10.1016/j.molcel.2014.06.017*.

Haralick, R.M., Shanmugam, K., and Dinstein, I. (1973). Textural Features for Image Classification. IEEE Trans. Syst. Man Cybern. *SMC-3*, 610–621.

Harty, C., Strahl, S., and Römisch, K. (2001). O-Mannosylation Protects Mutant Alpha-Factor Precursor from Endoplasmic Reticulum-associated Degradation. Mol. Biol. Cell *12*, 1093–1101.

Hebert, D.N., and Molinari, M. (2007). In and Out of the ER: Protein Folding, Quality Control, Degradation, and Related Human Diseases. Physiol. Rev. *87*, 1377–1408.

Heim, R., Cubitt, A.B., and Tsien, R.Y. (1995). Improved green fluorescence. Nature *373*, 663–664.

Heinicke, S., Livstone, M.S., Lu, C., Oughtred, R., Kang, F., Angiuoli, S.V., White, O., Botstein, D., and Dolinski, K. (2007). The Princeton Protein Orthology Database (P-POD): A Comparative Genomics Analysis Tool for Biologists. PLoS ONE *2*, e766.

Herzig, Y., Sharpe, H.J., Elbaz, Y., Munro, S., and Schuldiner, M. (2012). A Systematic Approach to Pair Secretory Cargo Receptors with Their Cargo Suggests a Mechanism for Cargo Selection by Erv14. PLoS Biol *10*, e1001329.

Herzog, B., Popova, B., Jakobshagen, A., Shahpasandzadeh, H., and Braus, G.H. (2013). Mutual Cross Talk between the Regulators Hac1 of the Unfolded Protein Response and

Gcn4 of the General Amino Acid Control of Saccharomyces cerevisiae. Eukaryot. Cell *12*, 1142–1154.

Hettema, E.H., Valdez-Taubas, J., and Pelham, H.R.B. (2004). Bsd2 binds the ubiquitin ligase Rsp5 and mediates the ubiquitination of transmembrane proteins. EMBO J. *23*, 1279–1288.

Hillenmeyer, M.E., Fung, E., Wildenhain, J., Pierce, S.E., Hoon, S., Lee, W., Proctor, M., St.Onge, R.P., Tyers, M., Koller, D., et al. (2008). The Chemical Genomic Portrait of Yeast: Uncovering a Phenotype for All Genes. Science *320*, 362–365.

Hirayama, H., Fujita, M., Yoko-o, T., and Jigami, Y. (2008). O-Mannosylation is Required for Degradation of the Endoplasmic Reticulum-associated Degradation Substrate Gas1*p via the Ubiquitin/Proteasome Pathway in Saccharomyces cerevisiae. J. Biochem. (Tokyo) *143*, 555–567.

Hodel, A.E., Harreman, M.T., Pulliam, K.F., Harben, M.E., Holmes, J.S., Hodel, M.R., Berland, K.M., and Corbett, A.H. (2006). Nuclear Localization Signal Receptor Affinity Correlates with in Vivo Localization in Saccharomyces cerevisiae. J. Biol. Chem. *281*, 23545–23556.

Hodel, M.R., Corbett, A.H., and Hodel, A.E. (2001). Dissection of a Nuclear Localization Signal. J. Biol. Chem. *276*, 1317–1325.

Huh, S., Lee, D., and Murphy, R.F. (2009). Efficient framework for automated classification of subcellular patterns in budding yeast. Cytometry A *75A*, 934–940.

Huh, W.-K., Falvo, J.V., Gerke, L.C., Carroll, A.S., Howson, R.W., Weissman, J.S., and O'Shea, E.K. (2003). Global analysis of protein localization in budding yeast. Nature *425*, 686–691.

Hung, M.-C., and Link, W. (2011). Protein localization in disease and therapy. J. Cell Sci. *124*, 3381–3392.

Huxley, C., Green, E.D., and Dunham, I. (1990). Rapid assessment of S. cerevisiae mating type by PCR. Trends Genet. TIG *6*, 236.

Huyer, G., Piluek, W.F., Fansler, Z., Kreft, S.G., Hochstrasser, M., Brodsky, J.L., and Michaelis, S. (2004). Distinct Machinery Is Required in Saccharomyces cerevisiae for the Endoplasmic Reticulum-associated Degradation of a Multispanning Membrane Protein and a Soluble Luminal Protein. J. Biol. Chem. *279*, 38369–38378.

Hwang, C., Sinskey, A.J., and Lodish, H.F. (1992). Oxidized redox state of glutathione in the endoplasmic reticulum. Science *257*, 1496–1502.

Imai, Y., Soda, M., and Takahashi, R. (2000). Parkin Suppresses Unfolded Protein Stress-induced Cell Death through Its E3 Ubiquitin-protein Ligase Activity. J. Biol. Chem. *275*, 35661–35664.

Ishiwata-Kimata, Y., Promlek, T., Kohno, K., and Kimata, Y. (2013a). BiP-bound and nonclustered mode of Ire1 evokes a weak but sustained unfolded protein response. Genes Cells *18*, 288–301.

Ishiwata-Kimata, Y., Yamamoto, Y., Takizawa, K., Kohno, K., and Kimata, Y. (2013b). F-actin and a Type-II Myosin Are Required for Efficient Clustering of the ER Stress Sensor Ire1. Cell Struct. Funct. *38*, 135–143.

Janke, C., Magiera, M.M., Rathfelder, N., Taxis, C., Reber, S., Maekawa, H., Moreno-Borchart, A., Doenges, G., Schwob, E., Schiebel, E., et al. (2004). A versatile toolbox for PCR-based tagging of yeast genes: new fluorescent proteins, more markers and promoter substitution cassettes. Yeast *21*, 947–962.

Jeffery, C.J. (1999). Moonlighting proteins. Trends Biochem. Sci. *24*, 8–11.

Jermy, A.J., Willer, M., Davis, E., Wilkinson, B.M., and Stirling, C.J. (2006). The Brl Domain in Sec63p Is Required for Assembly of Functional Endoplasmic Reticulum Translocons. J. Biol. Chem. *281*, 7899–7906.

Jessop, C.E., and Bulleid, N.J. (2004). Glutathione Directly Reduces an Oxidoreductase in the Endoplasmic Reticulum of Mammalian Cells. J. Biol. Chem. *279*, 55341–55347.

Jonikas, M.C., Collins, S.R., Denic, V., Oh, E., Quan, E.M., Schmid, V., Weibezahn, J., Schwappach, B., Walter, P., Weissman, J.S., et al. (2009). Comprehensive characterization of genes required for protein folding in the endoplasmic reticulum. Sci. STKE *323*, 1693.

Kainth, P., Sassi, H.E., Peña-Castillo, L., Chua, G., Hughes, T.R., and Andrews, B. (2009). Comprehensive genetic analysis of transcription factor pathways using a dual reporter gene system in budding yeast. Methods *48*, 258–264.

Kanki, T., Wang, K., Baba, M., Bartholomew, C.R., Lynch-Day, M.A., Du, Z., Geng, J., Mao, K., Yang, Z., Yen, W.-L., et al. (2009). A Genomic Screen for Yeast Mutants Defective in Selective Mitochondria Autophagy. Mol. Biol. Cell *20*, 4730–4738.

Katayama, T., Imaizumi, K., Sato, N., Miyoshi, K., Kudo, T., Hitomi, J., Morihara, T., Yoneda, T., Gomi, F., Mori, Y., et al. (1999). Presenilin-1 mutations downregulate the signalling pathway of the unfolded-protein response. Nat. Cell Biol. *1*, 479–485.

Kaufmann, A. (2009). A plasmid collection for PCR-based gene targeting in the filamentous ascomycete Ashbya gossypii. Fungal Genet. Biol. *46*, 595–603.

Kawahara, T., Yanagi, H., Yura, T., and Mori, K. (1997). Endoplasmic Reticulum Stress-induced mRNA Splicing Permits Synthesis of Transcription Factor Hac1p/Ern4p That Activates the Unfolded Protein Response. Mol. Biol. Cell *8*, 1845–1862.

Keenan, R.J., Freymann, D.M., Stroud, R.M., and Walter, P. (2001). The Signal Recognition Particle. Annu. Rev. Biochem. *70*, 755–775.

Kim, H., Kim, A., and Cunningham, K.W. (2012a). Vacuolar H+-ATPase (V-ATPase) Promotes Vacuolar Membrane Permeabilization and Nonapoptotic Death in Stressed Yeast. J. Biol. Chem. *287*, 19029–19039.

Kim, K., Yamashita, A., Wear, M.A., Maéda, Y., and Cooper, J.A. (2004). Capping protein binding to actin in yeast biochemical mechanism and physiological relevance. J. Cell Biol. *164*, 567–580.

Kim, S., Sideris, D.P., Sevier, C.S., and Kaiser, C.A. (2012b). Balanced Ero1 activation and inactivation establishes ER redox homeostasis. J. Cell Biol. *196*, 713–725.

Kimata, Y., Kimata, Y.I., Shimizu, Y., Abe, H., Farcasanu, I.C., Takeuchi, M., Rose, M.D., and Kohno, K. (2003). Genetic Evidence for a Role of BiP/Kar2 That Regulates Ire1 in Response to Accumulation of Unfolded Proteins. Mol. Biol. Cell *14*, 2559–2569.

Kimata, Y., Oikawa, D., Shimizu, Y., Ishiwata-Kimata, Y., and Kohno, K. (2004). A role for BiP as an adjustor for the endoplasmic reticulum stress-sensing protein Ire1. J. Cell Biol. *167*, 445–456.

Kimata, Y., Ishiwata-Kimata, Y., Yamada, S., and Kohno, K. (2006). Yeast unfolded protein response pathway regulates expression of genes for anti-oxidative stress and for cell surface proteins. Genes Cells *11*, 59–69.

Kimata, Y., Ishiwata-Kimata, Y., Ito, T., Hirata, A., Suzuki, T., Oikawa, D., Takeuchi, M., and Kohno, K. (2007). Two regulatory steps of ER-stress sensor Ire1 involving its cluster formation and interaction with unfolded proteins. J. Cell Biol. *179*, 75–86.

Kohno, K., Normington, K., Sambrook, J., Gething, M.J., and Mori, K. (1993). The promoter region of the yeast KAR2 (BiP) gene contains a regulatory domain that responds to the presence of unfolded proteins in the endoplasmic reticulum. Mol. Cell. Biol. *13*, 877–890.

Kopito, R.R. (1999). Biosynthesis and Degradation of CFTR. Physiol. Rev. *79*, S167–S173.

Korennykh, A.V., Egea, P.F., Korostelev, A.A., Finer-Moore, J., Zhang, C., Shokat, K.M., Stroud, R.M., and Walter, P. (2009). The unfolded protein response signals through high-order assembly of Ire1. Nature *457*, 687–693.

Kozutsumi, Y., Segal, M., Normington, K., Gething, M.-J., and Sambrook, J. (1988). The presence of malfolded proteins in the endoplasmic reticulum signals the induction of glucose-regulated proteins. Nature *332*, 462–464.

Laboissière, M.C.A., Sturley, S.L., and Raines, R.T. (1995). The Essential Function of Protein-disulfide Isomerase Is to Unscramble Non-native Disulfide Bonds. J. Biol. Chem. *270*, 28006–28009.

Lakkaraju, A.K.K., Thankappan, R., Mary, C., Garrison, J.L., Taunton, J., and Strub, K. (2012). Efficient secretion of small proteins in mammalian cells relies on Sec62-dependent posttranslational translocation. Mol. Biol. Cell *23*, 2712–2722.

Laws, K.I. (1980a). Rapid Texture Identification. pp. 376–381.

Laws, K.I. (1980b). Textured Image Segmentation. Ph.D. University of Southern California.

Leidich, S.D., Drapp, D.A., and Orlean, P. (1994). A conditionally lethal yeast mutant blocked at the first step in glycosyl phosphatidylinositol anchor synthesis. J. Biol. Chem. *269*, 10193–10196.

Li, Z., Vizeacoumar, F.J., Bahr, S., Li, J., Warringer, J., Vizeacoumar, F.S., Min, R., VanderSluis, B., Bellay, J., DeVit, M., et al. (2011). Systematic exploration of essential yeast gene function with temperature-sensitive mutants. Nat. Biotechnol. *29*, 361–367.

Liko, D., Conway, M.K., Grunwald, D.S., and Heideman, W. (2010). Stb3 Plays a Role in the Glucose-Induced Transition from Quiescence to Growth in Saccharomyces cerevisiae. Genetics *185*, 797–810.

Liu, D.-H., Lam, K.-M., and Shen, L.-S. (2004). Optimal sampling of Gabor features for face recognition. Pattern Recognit. Lett. *25*, 267–276.

Liu, X., Zhang, X., Wang, C., Liu, L., Lei, M., and Bao, X. (2007). Genetic and Comparative Transcriptome Analysis of Bromodomain Factor 1 in the Salt Stress Response of Saccharomyces cerevisiae. Curr. Microbiol. *54*, 325–330.

Liu, X.F., Supek, F., Nelson, N., and Culotta, V.C. (1997). Negative Control of Heavy Metal Uptake by the Saccharomyces cerevisiae BSD2 Gene. J. Biol. Chem. *272*, 11763–11769.

Ljungdahl, P.O., Gimeno, C.J., Styles, C.A., and Fink, G.R. (1992). SHR3: A novel component of the secretory pathway specifically required for localization of amino acid permeases in yeast. Cell *71*, 463–478.

Low, Y.S. (2013). Kinetochores are required to fully activate secretory pathway in elevated ER stress. Doctoral dissertation. Victoria University of Wellington.

Lyles, M.M., and Gilbert, H.F. (1991). Catalysis of the oxidative folding of ribonuclease A by protein disulfide isomerase: dependence of the rate on the composition of the redox buffer. Biochemistry (Mosc.) *30*, 613–619.

MacDiarmid, C.W., Milanick, M.A., and Eide, D.J. (2002). Biochemical Properties of Vacuolar Zinc Transport Systems ofSaccharomyces cerevisiae. J. Biol. Chem. *277*, 39187–39194.

Machi, K., Azuma, M., Igarashi, K., Matsumoto, T., Fukuda, H., Kondo, A., and Ooshima, H. (2004). Rot1p of Saccharomyces cerevisiae is a putative membrane protein required for normal levels of the cell wall 1,6-β-glucan. Microbiology *150*, 3163–3173.

Mateus, C., and Avery, S.V. (2000). Destabilized green fluorescent protein for monitoring dynamic changes in yeast gene expression with flow cytometry. Yeast *16*, 1313–1323.

Matlack, K.E.S., Misselwitz, B., Plath, K., and Rapoport, T.A. (1999). BiP Acts as a Molecular Ratchet during Posttranslational Transport of Prepro-α Factor across the ER Membrane. Cell *97*, 553–564.

Matz, M.V., Fradkov, A.F., Labas, Y.A., Savitsky, A.P., Zaraisky, A.G., Markelov, M.L., and Lukyanov, S.A. (1999). Fluorescent proteins from nonbioluminescent Anthozoa species. Nat. Biotechnol. *17*, 969–973.

Mazumder, A., Pesudo, L.Q., McRee, S., Bathe, M., and Samson, L.D. (2013). Genome-wide single-cell-level screen for protein abundance and localization changes in response to DNA damage in S. cerevisiae. Nucleic Acids Res. gkt715.

Meiling-Wesse, K., Barth, H., Voss, C., Barmark, G., Murén, E., Ronne, H., and Thumm, M. (2002). Yeast Mon1p/Aut12p functions in vacuolar fusion of autophagosomes and cvt-vesicles. FEBS Lett. *530*, 174–180.

Melamed, D., Pnueli, L., and Arava, Y. (2008). Yeast translational response to high salinity: Global analysis reveals regulation at multiple levels. RNA *14*, 1337–1351.

Merksamer, P.I., Trusina, A., and Papa, F.R. (2008). Real-time redox measurements during endoplasmic reticulum stress reveal interlinked protein folding functions. Cell *135*, 933–947.

Mizuta, K., and Warner, J.R. (1994). Continued functioning of the secretory pathway is essential for ribosome synthesis. Mol. Cell. Biol. *14*, 2493–2502.

Molteni, S.N., Fassio, A., Ciriolo, M.R., Filomeni, G., Pasqualetto, E., Fagioli, C., and Sitia, R. (2004). Glutathione Limits Ero1-dependent Oxidation in the Endoplasmic Reticulum. J. Biol. Chem. *279*, 32667–32673.

Monteiro, P.T., Mendes, N.D., Teixeira, M.C., d' Orey, S., Tenreiro, S., Mira, N.P., Pais, H., Francisco, A.P., Carvalho, A.M., Lourenço, A.B., et al. (2008). YEASTRACT-DISCOVERER: new tools to improve the analysis of transcriptional regulatory associations in Saccharomyces cerevisiae. Nucleic Acids Res. *36*, D132–D136.

Moreno, F., and Herrero, P. (2002). The hexokinase 2-dependent glucose signal transduction pathway of Saccharomyces cerevisiae. FEMS Microbiol. Rev. *26*, 83–90.

Mori, K., Sant, A., Kohno, K., Normington, K., Gething, M., and Sambrook, J. (1992). A 22 bp cis-acting element is necessary and sufficient for the induction of the yeast KAR2 (BiP) gene by unfolded proteins. EMBO J. *11*, 2583.

Mori, K., Ma, W., Gething, M.-J., and Sambrook, J. (1993). A transmembrane protein with a cdc2+CDC28-related kinase activity is required for signaling from the ER to the nucleus. Cell `, 743–756.

Mullins, C., Lu, Y., Campbell, A., Fang, H., and Green, N. (1995). A Mutation Affecting Signal Peptidase Inhibits Degradation of an Abnormal Membrane Protein in Saccharomyces cerevisiae. J. Biol. Chem. *270*, 17139–17147.

Murphy, R.F. (2005). Location proteomics: a systems approach to subcellular location. Biochem. Soc. Trans. *33*, 535–538.

Murphy, R.F., Velliste, M., and Porreca, G. (2003). Robust numerical features for description and classification of subcellular location patterns in fluorescence microscope images. J. VLSI Signal Process. Syst. Signal Image Video Technol. *35*, 311–321.

Mursalin, T.-E. (2013). PARALLEL IMAGE PROCESSING FOR HIGH CONTENT SCREENING DATA. M.Sc. McMaster University.

Nadler-Holly, M., Breker, M., Gruber, R., Azia, A., Gymrek, M., Eisenstein, M., Willison, K.R., Schuldiner, M., and Horovitz, A. (2012). Interactions of subunit CCT3 in the yeast chaperonin CCT/TRiC with Q/N-rich proteins revealed by high-throughput microscopy analysis. Proc. Natl. Acad. Sci. *109*, 18833–18838.

Narayanaswamy, R., Moradi, E.K., Niu, W., Hart, G.T., Davis, M., McGary, K.L., Ellington, A.D., and Marcotte, E.M. (2009a). Systematic Definition of Protein Constituents along the Major Polarization Axis Reveals an Adaptive Reuse of the Polarization Machinery in Pheromone-Treated Budding Yeast. J. Proteome Res. *8*, 6–19.

Narayanaswamy, R., Levy, M., Tsechansky, M., Stovall, G.M., O'Connell, J.D., Mirrielees, J., Ellington, A.D., and Marcotte, E.M. (2009b). Widespread reorganization of metabolic enzymes into reversible assemblies upon nutrient starvation. Proc. Natl. Acad. Sci. *106*, 10147–10152.

Needham, P.G., and Brodsky, J.L. (2013). How early studies on secreted and membrane protein quality control gave rise to the ER associated degradation (ERAD) pathway: The early history of ERAD. Biochim. Biophys. Acta BBA - Mol. Cell Res. *1833*, 2447–2457.

Newberg, J., Hua, J., and Murphy, R.F. (2009). Location Proteomics: Systematic Determination of Protein Subcellular Location. In Systems Biology, I.V. Maly, ed. (Totowa, NJ: Humana Press), pp. 313–332.

Newman, J.R.S., Ghaemmaghami, S., Ihmels, J., Breslow, D.K., Noble, M., DeRisi, J.L., and Weissman, J.S. (2006). Single-cell proteomic analysis of S. cerevisiae reveals the architecture of biological noise. Nature *441*, 840–846.

Ng, D.T.W. (2005). Screening for mutants defective in secretory protein maturation and ER quality control. Methods *35*, 366–372.

Ng, D.T., Brown, J.D., and Walter, P. (1996). Signal sequences specify the targeting route to the endoplasmic reticulum membrane. J. Cell Biol. *134*, 269–278.

Ng, D.T.W., Spear, E.D., and Walter, P. (2000). The Unfolded Protein Response Regulates Multiple Aspects of Secretory and Membrane Protein Biogenesis and Endoplasmic Reticulum Quality Control. J. Cell Biol. *150*, 77–88.

Nickerson, D.P., West, M., Henry, R., and Odorizzi, G. (2010). Regulators of Vps4 ATPase Activity at Endosomes Differentially Influence the Size and Rate of Formation of Intralumenal Vesicles. Mol. Biol. Cell *21*, 1023–1032.

Niwa, M., Patil, C.K., DeRisi, J., and Walter, P. (2005). Genome-scale approaches for discovering novel nonconventional splicing substrates of the Ire1 nuclease. Genome Biol. *6*, 1–10.

Nojima, H., Leem, S.-H., Araki, H., Sakai, A., Nakashima, N., Kanaoka, Y., and Ono, Y. (1994). Hac1: A novel yeast bZIP protein binding to the CRE motif is a multicopy suppressor for cdcW mutant of Schizosaccharomyces pombe. Nucleic Acids Res. *22*, 5279–5288.

Nombela, C., Gil, C., and Chaffin, W.L. (2006). Non-conventional protein secretionin yeast. Trends Microbiol. *14*, 15–21.

Noree, C., Sato, B.K., Broyer, R.M., and Wilhelm, J.E. (2010). Identification of novel filament-forming proteins in Saccharomyces cerevisiae and Drosophila melanogaster. J. Cell Biol. *190*, 541–551.

Novick, P., Field, C., and Schekman, R. (1980). Identification of 23 complementation groups required for post-translational events in the yeast secretory pathway. Cell *21*, 205–215.

Novick, P., Ferro, S., and Schekman, R. (1981). Order of events in the yeast secretory pathway. Cell *25*, 461–469.

Ogawa, N., and Mori, K. (2004). Autoregulation of the HAC1 gene is required for sustained activation of the yeast unfolded protein response. Genes Cells *9*, 95–104.

Ogle, J.M., and Ramakrishnan, V. (2005). Structural Insights into Translational Fidelity. Annu. Rev. Biochem. *74*, 129–177.

Oikawa, D., Kimata, Y., and Kohno, K. (2007). Self-association and BiP dissociation are not sufficient for activation of the ER stress sensor Ire1. J. Cell Sci. *120*, 1681–1688.

Okamura, K., Kimata, Y., Higashio, H., Tsuru, A., and Kohno, K. (2000). Dissociation of Kar2p/BiP from an ER Sensory Molecule, Ire1p, Triggers the Unfolded Protein Response in Yeast. Biochem. Biophys. Res. Commun. *279*, 445–450.

Orlean, P., and Menon, A.K. (2007). Thematic review series: Lipid Posttranslational Modifications. GPI anchoring of protein in yeast and mammalian cells, or: how we learned to stop worrying and love glycophospholipids. J. Lipid Res. *48*, 993–1011.

Osborne, A.R., Rapoport, T.A., and van den Berg, B. (2005). Protein Translocation by the Sec61/Secy Channel. Annu. Rev. Cell Dev. Biol. *21*, 529–550.

Ozcan, L., and Tabas, I. (2012). Role of Endoplasmic Reticulum Stress in Metabolic Disease and Other Disorders. Annu. Rev. Med. *63*, 317–328.

Panzner, S., Dreier, L., Hartmann, E., Kostka, S., and Rapoport, T.A. (1995). Posttranslational protein transport in yeast reconstituted with a purified complex of Sec proteins and Kar2p. Cell *81*, 561–570.

Patil, C.K., Li, H., and Walter, P. (2004). Gcn4p and Novel Upstream Activating Sequences Regulate Targets of the Unfolded Protein Response. PLoS Biol *2*, e246.

Pemberton, L.F., and Blobel, G. (1997). Characterization of the Wtm proteins, a novel family of Saccharomyces cerevisiae transcriptional modulators with roles in meiotic regulation and silencing. Mol. Cell. Biol. *17*, 4830.

Pincus, D., Chevalier, M.W., Aragón, T., van Anken, E., Vidal, S.E., El-Samad, H., and Walter, P. (2010). BiP Binding to the ER-Stress Sensor Ire1 Tunes the Homeostatic Behavior of the Unfolded Protein Response. PLoS Biol. *8*, e1000415.

Pittet, M., and Conzelmann, A. (2007). Biosynthesis and function of GPI proteins in the yeast Saccharomyces cerevisiae. Biochim. Biophys. Acta BBA - Mol. Cell Biol. Lipids *1771*, 405–420.

Pollard, M.G., Travers, K.J., and Weissman, J.S. (1998). Ero1p: A Novel and Ubiquitous Protein with an Essential Role in Oxidative Protein Folding in the Endoplasmic Reticulum. Mol. Cell *1*, 171–182.

Prasher, D.C., Eckenrode, V.K., Ward, W.W., Prendergast, F.G., and Cormier, M.J. (1992). Primary structure of the Aequorea victoria green-fluorescent protein. Gene *111*, 229–233.

Prein, B., Natter, K., and Kohlwein, S.D. (2000). A novel strategy for constructing N-terminal chromosomal fusions to green fluorescent protein in the yeast Saccharomyces cerevisiae. FEBS Lett. *485*, 29–34.

Promlek, T., Ishiwata-Kimata, Y., Shido, M., Sakuramoto, M., Kohno, K., and Kimata, Y. (2011). Membrane aberrancy and unfolded proteins activate the endoplasmic reticulum stress sensor Ire1 in different ways. Mol. Biol. Cell *22*, 3520–3532.

Proszynski, T.J., Klemm, R.W., Gravert, M., Hsu, P.P., Gloor, Y., Wagner, J., Kozak, K., Grabner, H., Walzer, K., Bagnat, M., et al. (2005). A genome-wide visual screen reveals a role for sphingolipids and ergosterol in cell surface delivery in yeast. Proc. Natl. Acad. Sci. U. S. A. *102*, 17981–17986.

Quackenbush, J. (2002). Microarray data normalization and transformation. Nat. Genet. *32*, 496–501.

R Core Team (2013). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria. URL http://www.R-project.org/. Version 3.0.1.

Randez-Gil, F., Herrero, P., Sanz, P., Prieto, J.A., and Moreno, F. (1998). Hexokinase PII has a double cytosolic-nuclear localisation in Saccharomyces cerevisiae. FEBS Lett. *425*, 475–478.

Rapoport, T.A. (2007). Protein translocation across the eukaryotic endoplasmic reticulum and bacterial plasma membranes. Nature *450*, 663–669.

Regalla, L.M., and Lyons, T.J. (2006). Zinc in yeast: mechanisms involved in homeostasis. In Molecular Biology of Metal Homeostasis and Detoxification, M.J. Tamas, and E. Martinoia, eds. (Springer Berlin Heidelberg), pp. 37–58.

Rimon, N., and Schuldiner, M. (2011). Getting the whole picture: combining throughput with content in microscopy. J. Cell Sci. *124*, 3743–3751.

Ritter, N., and Cooper, J. (2009). New Resolution Independent Measures of Circularity. J. Math. Imaging Vis. *35*, 117–127.

Robinson, J.S., Klionsky, D.J., Banta, L.M., and Emr, S.D. (1988). Protein sorting in Saccharomyces cerevisiae: isolation of mutants defective in the delivery and processing of multiple vacuolar hydrolases. Mol. Cell. Biol. *8*, 4936–4948.

Roomans, G.M. (2003). Pharmacological approaches to correcting the ion transport defect in cystic fibrosis. Am. J. Respir. Med. *2*, 413–431.

Rosenberger, R.F., and Hilton, J. (1983). The frequency of transcriptional and translational errors at nonsense codons in the lacZ gene of Escherichia coli. Mol. Gen. Genet. MGG *191*, 207–212.

Rothman, J.H., Howald, I., and Stevens, T.H. (1989). Characterization of genes required for protein sorting and vacuolar function in the yeast Saccharomyces cerevisiae. EMBO J. *8*, 2057–2065.

Rüegsegger, U., Leber, J.H., and Walter, P. (2001). Block of HAC1 mRNA Translation by Long-Range Base Pairing Is Released by Cytoplasmic Splicing upon Induction of the Unfolded Protein Response. Cell *107*, 103–114.

Sato, B.K., Schulz, D., Do, P.H., and Hampton, R.Y. (2009). Misfolded Membrane Proteins Are Specifically Recognized by the Transmembrane Domain of the Hrd1p Ubiquitin Ligase. Mol. Cell *34*, 212–222.

Schubert, U., Antón, L.C., Gibbs, J., Norbury, C.C., Yewdell, J.W., and Bennink, J.R. (2000). Rapid degradation of a large fraction of newly synthesized proteins by proteasomes. Nature *404*, 770–774.

Schuldiner, M., Collins, S.R., Thompson, N.J., Denic, V., Bhamidipati, A., Punna, T., Ihmels, J., Andrews, B., Boone, C., Greenblatt, J.F., et al. (2005). Exploration of the Function and Organization of the Yeast Early Secretory Pathway through an Epistatic Miniarray Profile. Cell *123*, 507–519.

Schuldiner, M., Metz, J., Schmid, V., Denic, V., Rakwalska, M., Schmitt, H.D., Schwappach, B., and Weissman, J.S. (2008). The GET Complex Mediates Insertion of Tail-Anchored Proteins into the ER Membrane. Cell *134*, 634–645.

Sevier, C.S., Qu, H., Heldman, N., Gross, E., Fass, D., and Kaiser, C.A. (2007). Modulation of Cellular Disulfide-Bond Formation and the ER Redox Environment by Feedback Regulation of Ero1. Cell *129*, 333–344.

Shaner, N.C., Campbell, R.E., Steinbach, P.A., Giepmans, B.N.G., Palmer, A.E., and Tsien, R.Y. (2004). Improved monomeric red, orange and yellow fluorescent proteins derived from Discosoma sp. red fluorescent protein. Nat. Biotechnol. *22*, 1567–1572.

Shao, S., and Hegde, R.S. (2011). Membrane Protein Insertion at the Endoplasmic Reticulum. Annu. Rev. Cell Dev. Biol. *27*, 25–56.

Shimura, H., Hattori, N., Kubo, S., Mizuno, Y., Asakawa, S., Minoshima, S., Shimizu, N., Iwai, K., Chiba, T., Tanaka, K., et al. (2000). Familial Parkinson disease gene product, parkin, is a ubiquitin-protein ligase. Nat. Genet. *25*, 302–305.

Shin, C.-S., Kim, S.Y., and Huh, W.-K. (2009). TORC1 controls degradation of the transcription factor Stp1, a key effector of the SPS amino-acid-sensing pathway in Saccharomyces cerevisiae. J. Cell Sci. *122*, 2089–2099.

Sidrauski, C., and Walter, P. (1997). The Transmembrane Kinase Ire1p Is a Site-Specific Endonuclease That Initiates mRNA Splicing in the Unfolded Protein Response. Cell *90*, 1031–1039.

Sidrauski, C., Cox, J.S., and Walter, P. (1996). tRNA Ligase Is Required for Regulated mRNA Splicing in the Unfolded Protein Response. Cell *87*, 405–413.

Simpson-Lavy, K.J., and Johnston, M. (2013). SUMOylation regulates the SNF1 protein kinase. Proc. Natl. Acad. Sci. *110*, 17432–17437.

Singan, V.R., Handzic, K., Curran, K.M., and Simpson, J.C. (2012). A method for improved clustering and classification of microscopy images using quantitative co-localization coefficients. BMC Res. Notes *5*, 281.

Sipos, G., Puoti, A., and Conzelmann, A. (1994). Glycosylphosphatidylinositol membrane anchors in Saccharomyces cerevisiae: absence of ceramides from complete precursor glycolipids. EMBO J. *13*, 2789.

Soboleski, M.R., Oaks, J., and Halford, W.P. (2005). Green fluorescent protein is a quantitative reporter of gene expression in individual eukaryotic cells. FASEB J. *19*, 440–442.

Sopko, R., Huang, D., Preston, N., Chua, G., Papp, B., Kafadar, K., Snyder, M., Oliver, S.G., Cyert, M., Hughes, T.R., et al. (2006). Mapping Pathways and Phenotypes by Systematic Gene Overexpression. Mol. Cell *21*, 319–330.

Stathopoulos-Gerontides, A., Guo, J.J., and Cyert, M.S. (1999). Yeast calcineurin regulates nuclear localization of the Crz1p transcription factor through dephosphorylation. Genes Dev. *13*, 798–803.

Sternberg, S.R. (1983). Biomedical image processing. Computer *16*, 22–34.

Stockwell, B.R. (2000). Chemical genetics: ligand-based discovery of gene function. Nat. Rev. Genet. *1*, 116–125.

Stouffer, S.A., Suchman, E.A., DeVinney, L.C., Star, S.A., and Williams Jr, R.M. (1949). The American soldier: adjustment during army life.(Studies in social psychology in World War II, Vol. 1.).

Sullivan, J.A., Lewis, M.J., Nikko, E., and Pelham, H.R.B. (2007). Multiple Interactions Drive Adaptor-Mediated Recruitment of the Ubiquitin Ligase Rsp5 to Membrane Proteins In Vivo and In Vitro. Mol. Biol. Cell *18*, 2429–2440.

Takemaru, K., Harashima, S., Ueda, H., and Hirose, S. (1998). Yeast Coactivator MBF1 Mediates GCN4-Dependent Transcriptional Activation. Mol. Cell. Biol. *18*, 4971.

Teixeira, M.C., Monteiro, P., Jain, P., Tenreiro, S., Fernandes, A.R., Mira, N.P., Alenquer, M., Freitas, A.T., Oliveira, A.L., and Sá-Correia, I. (2006). The YEASTRACT database: a tool for the analysis of transcription regulatory associations in Saccharomyces cerevisiae. Nucleic Acids Res. *34*, D446–D451.

Teixeira, M.C., Monteiro, P.T., Guerreiro, J.F., Goncalves, J.P., Mira, N.P., dos Santos, S.C., Cabrito, T.R., Palma, M., Costa, C., Francisco, A.P., et al. (2013). The YEASTRACT database: an upgraded information system for the analysis of gene and genomic transcription regulation in Saccharomyces cerevisiae. Nucleic Acids Res. *42*, D161–D166.

Tkach, J.M., Yimit, A., Lee, A.Y., Riffle, M., Costanzo, M., Jaschob, D., Hendry, J.A., Ou, J., Moffat, J., Boone, C., et al. (2012). Dissecting DNA damage response pathways by analysing protein localization and abundance changes during DNA replication stress. Nat. Cell Biol. *14*, 966–976.

Toikkanen, J., Fatal, N., Hildén, P., Makarow, M., and Kuismanen, E. (2006). YET1, YET2 and YET3 of Saccharomyces cerevisiae encode BAP31 homologues with partially overlapping functions. J Biol Sci *6*, 446–456.

Tong, A.H.Y., and Boone, C. (2007). High-Throughput Strain Construction and Systematic Synthetic Lethal Screening in. Methods Microbiol. *36*, 369–707.

Tong, A.H., Evangelista, M., Parsons, A.B., Xu, H., Bader, G.D., Pagé, N., Robinson, M., Raghibizadeh, S., Hogue, C.W., Bussey, H., et al. (2001). Systematic genetic analysis with ordered arrays of yeast deletion mutants. Science *294*, 2364–2368.

Tong, A.H.Y., Lesage, G., Bader, G.D., Ding, H., Xu, H., Xin, X., Young, J., Berriz, G.F., Brost, R.L., Chang, M., et al. (2004). Global mapping of the yeast genetic interaction network. Sci. STKE *303*, 808.

Travers, K.J., Patil, C.K., Wodicka, L., Lockhart, D.J., Weissman, J.S., and Walter, P. (2000). Functional and genomic analyses reveal an essential coordination between the unfolded protein response and ER-associated degradation. Cell *101*, 249–258.

Tringe, S.G., Willis, J., Liberatore, K.L., and Ruby, S.W. (2006). The WTM Genes in Budding Yeast Amplify Expression of the Stress-Inducible Gene RNR3. Genetics *174*, 1215–1228.

Tsien, R.Y. (1998). The Green Fluorescent Protein. Annu. Rev. Biochem. *67*, 509–544.

Tu, B.P., and Weissman, J.S. (2002). The FAD- and O2-Dependent Reaction Cycle of Ero1-Mediated Oxidative Protein Folding in the Endoplasmic Reticulum. Mol. Cell *10*, 983–994.

Tu, B.P., and Weissman, J.S. (2004). Oxidative Protein Folding in Eukaryotes Mechanisms and Consequences. J. Cell Biol. *164*, 341–346.

Tu, B.P., Ho-Schleyer, S.C., Travers, K.J., and Weissman, J.S. (2000). Biochemical basis of oxidative protein folding in the endoplasmic reticulum. Science *290*, 1571–1574.

Turner, M.R. (1986). Texture discrimination by Gabor functions. Biol. Cybern. *55*, 71–82.

Uetz, P., Giot, L., Cagney, G., Mansfield, T.A., Judson, R.S., Knight, J.R., Lockshon, D., Narayan, V., Srinivasan, M., Pochart, P., et al. (2000). A comprehensive analysis of protein–protein interactions in Saccharomyces cerevisiae. Nature *403*, 623–627.

Urano, F., Bertolotti, A., and Ron, D. (2000). IRE1 and efferent signaling from the endoplasmic reticulum. J. Cell Sci. *113*, 3697–3702.

Uzunova, K., Göttsche, K., Miteva, M., Weisshaar, S.R., Glanemann, C., Schnellhardt, M., Niessen, M., Scheel, H., Hofmann, K., Johnson, E.S., et al. (2007). Ubiquitin-dependent Proteolytic Control of SUMO Conjugates. J. Biol. Chem. *282*, 34167–34175.

Vizeacoumar, F.J., Chong, Y., Boone, C., and Andrews, B.J. (2009). A picture is worth a thousand words: Genomics to phenomics in the yeast Saccharomyces cerevisiae. FEBS Lett. *583*, 1656–1661.

Vizeacoumar, F.J., Dyk, N. van, S.Vizeacoumar, F., Cheung, V., Li, J., Sydorskyy, Y., Case, N., Li, Z., Datti, A., Nislow, C., et al. (2010). Integrating high-throughput genetic interaction mapping and high-content screening to explore yeast spindle morphogenesis. J. Cell Biol. *188*, 69–81.

Van Vliet, L.J., Young, I.T., and Verbeek, P.W. (1998). Recursive Gaussian derivative filters. In Fourteenth International Conference on Pattern Recognition, 1998. Proceedings, pp. 509–514 vol.1.

Walter, P., and Blobel, G. (1981). Translocation of proteins across the endoplasmic reticulum III. Signal recognition protein (SRP) causes signal sequence-dependent and site-specific arrest of chain elongation that is released by microsomal membranes. J. Cell Biol. *91*, 557–561.

Wang, C.-W. (2002). The Ccz1-Mon1 Protein Complex Is Required for the Late Step of Multiple Vacuole Delivery Pathways. J. Biol. Chem. *277*, 47917–47927.

Ward, C.L., and Kopito, R.R. (1994). Intracellular turnover of cystic fibrosis transmembrane conductance regulator. Inefficient processing and rapid degradation of wild-type and mutant proteins. J. Biol. Chem. *269*, 25710–25718.

Welihinda, A.A., Tirasophon, W., Green, S.R., and Kaufman, R.J. (1997). Gene Induction in Response to Unfolded Protein in the Endoplasmic Reticulum is Mediated through Ire1p Kinase Interaction with a Transcriptional Coactivator Complex Containing Ada5p. Proc. Natl. Acad. Sci. U. S. A. *94*, 4289–4294.

Welihinda, A.A., Tirasophon, W., and Kaufman, R.J. (2000). The Transcriptional Co-activator ADA5 Is Required for HAC1 mRNA Processing in Vivo. J. Biol. Chem. *275*, 3377–3381.

Whitlock, M.C. (2005). Combining probability from independent tests: the weighted Z-method is superior to Fisher's approach. J. Evol. Biol. *18*, 1368–1373.

Wilson, J.D., and Barlowe, C. (2010). Yet1p and Yet3p, the Yeast Homologs of BAP29 and BAP31, Interact with the Endoplasmic Reticulum Translocation Apparatus and Are Required for Inositol Prototrophy. J. Biol. Chem. *285*, 18252–18261.

Wilson, J.D., Thompson, S.L., and Barlowe, C. (2011). Yet1p–Yet3p interacts with Scs2p–Opi1p to regulate ER localization of the Opi1p repressor. Mol. Biol. Cell *22*, 1430–1439.

Winzeler, E.A., Shoemaker, D.D., Astromoff, A., Liang, H., Anderson, K., Andre, B., Bangham, R., Benito, R., Boeke, J.D., Bussey, H., et al. (1999). Functional Characterization of the S. cerevisiae Genome by Gene Deletion and Parallel Analysis. Science *285*, 901–906.

Wolinski, H., Petrovič, U., Mattiazzi, M., Petschnigg, J., Heise, B., Natter, K., and Kohlwein, S.D. (2009). Imaging-Based Live Cell Yeast Screen Identifies Novel Factors Involved in Peroxisome Assembly. J. Proteome Res. *8*, 20–27.

Wright, M.B., Ramos, J., Gomez, M.J., Moulder, K., Scherrer, M., Munson, G., and Gaber, R.F. (1997). Potassium Transport by Amino Acid Permeases in Saccharomyces cerevisiae. J. Biol. Chem. *272*, 13647–13652.

Wu, W.-S., and Li, W.-H. (2008). Identifying gene regulatory modules of heat shock response in yeast. BMC Genomics *9*, 439.

Xie, W., and Ng, D.T.W. (2010). ERAD substrate recognition in budding yeast. Semin. Cell Dev. Biol. *21*, 533–539.

Xie, Y., and Varshavsky, A. (2001). RPN4 is a ligand, substrate, and transcriptional regulator of the 26S proteasome: A negative feedback circuit. Proc. Natl. Acad. Sci. *98*, 3056–3061.

Xie, W., Kanehara, K., Sayeed, A., and Ng, D.T.W. (2009). Intrinsic Conformational Determinants Signal Protein Misfolding to the Hrd1/Htm1 Endoplasmic Reticulum–associated Degradation System. Mol. Biol. Cell *20*, 3317–3329.

Xie, Y., Kerscher, O., Kroetz, M.B., McConchie, H.F., Sung, P., and Hochstrasser, M. (2007). The Yeast Hex3·Slx8 Heterodimer Is a Ubiquitin Ligase Stimulated by Substrate Sumoylation. J. Biol. Chem. *282*, 34176–34184.

YaDeau, J.T., Klein, C., and Blobel, G. (1991). Yeast signal peptidase contains a glycoprotein and the Sec11 gene product. Proc. Natl. Acad. Sci. U. S. A. *88*, 517.

Yorimitsu, T., Nair, U., Yang, Z., and Klionsky, D.J. (2006). Endoplasmic Reticulum Stress Triggers Autophagy. J. Biol. Chem. *281*, 30299–30304.

Yoshida, H. (2007). ER stress and diseases. FEBS J. *274*, 630–658.

Yoshimoto, H., Saltsman, K., Gasch, A.P., Li, H.X., Ogawa, N., Botstein, D., Brown, P.O., and Cyert, M.S. (2002). Genome-wide Analysis of Gene Expression Regulated by the Calcineurin/Crz1p Signaling Pathway in Saccharomyces cerevisiae. J. Biol. Chem. *277*, 31079–31088.

7 APPENDICES

7.1 Nuclei recognition script

```
Proc nuclei_identification (image nucleus in "nuclei stained image", bool
ShowIllustrations=YES in, objectlist nuclei out "objectlist containing
detcted nuclei", bool ShowTestingImages=NO in, bool
IntensityFilterHigh=No in "filters nuclei + input # of  stdev away of the
mean RFP intensity", bool IntensityFilterLow=No in "filters nuclei -
input # of  stdev away of the mean RFP intensity", string
RoundnessFilter="Radii" in "method for roundness filtering uses either a
radius ratio method or form factor method: possible values 'Radii' 'Form'
or 'none'")
Peters "Detects and outputs an objectlist contianing nuclei"
{
Input(th, 1.4, "initial mask threshold : Nuclei", "d", "Controls the size
of the initial masking threshold, larger values increase mask, adjust
based on nuclei mask")
input(rom, 5, "range of maximums: Nuclei", "d", "Nuclei splitting, Point
is discarded as object center if within a disk with radius
RangeOfMaximums it has the highest pixel value")
input(sdr, 2, "smoothing disk radius: Nuclei", "d", "Nuclei splitting,
Radius for smoothing disk")
Input(minar, 20, "minimum area of initial nuclei : Nuclei", "d", "Nuclei
splitting, minimum allowed area of split nuclei")
input(minar2, 20, "minimum area of final nuclei : Nuclei", "d", "minimum
allowed area of detected nuclei")
Input(sm, 4, "smoothing : Nuclei", "d", "Controls the smoothness of
detected nuclei. adjust based on nuclei")
Input(tun, 2.5, "individual threshold : Nuclei", "d", "Controls the
individual threshold around which individual nuclei are adjusted. adjust
based on nuclei")
input(rd, 0.5, "roundness : Nuclei", "d", "Controls the filtering of
nuclei based on roundness. values closer to one should be more round.
adjust based on nuclei")
input(ed, 2, "border edge: Nuclei", "d", "distance for border removal")
input(ar, 150, "maximum area: Nuclei", "d", "max area filter for
identified nuclei")
input(ih, 2, "Intensity High: Nuclei", "d", "number of stddev away from
mean for high intensity filter")
input(il, 2, "Intensity Low: Nuclei", "d", "number of stddev away from
mean for low intensity filter")
ThresholdXX(th, Image=nucleus)
Mask(threshold, image=nucleus)
if(ShowIllustrations)
imageview(mask, "nuclei mask", image=nucleus)
end()
//Bright_Mask(nucleus, 2)
//imageview(mask, "bright mask")
mask2stencil(mask)
      stencil2objects()
      set(cells_initial=objects)
if(ShowTestingImages)
      imageview(cells_initial.border, "initial objects", image=nucleus,
gamma=2.6)
end()
      split_stuck_objects(nucleus, RangeOfMaximums=rom, minarea=minar,
SmoothingDiskRadius=sdr)
set(cells_initial_split=objects)
if(ShowTestingImages)
imageview(cells_initial_split.border, "split objects", image=nucleus,
gamma=2.6)
end()
calczone(10, ZoneType="equidistant", objects=objects)
CalcZoneIntensity(Image=nucleus)
set(initialzone_objects=objects)
if(ShowTestingImages)
imageview(initialzone_objects.zone, "nuclei inner zones", image=nucleus)
imageview(initialzone_objects.outerzone, "nuclei outer zones",
image=nucleus)
end()
IndividualThreshold(MinArea=minar2, MaxMode="maxpoint", Smooth=sm,
Image=nucleus, Tune=tun)
set(unfiltered_objects=objects)
if(ShowTestingImages)
imageview(unfiltered_objects.border, "individual threshold",
image=nucleus, gamma=2.6)
end()

//    Object Filters
RemoveBorderObjects(ed)
Calcarea() //Calculates area
Calcarea(border) //Calculates border area
if(RoundnessFilter=="Form")
Calcattr(Roundness_corrected, 3.544*sqrt(area-
border_area/2.0)/border_area-0.1)
objectfilter(Roundness_corrected>rd)
else()
      if(RoundnessFilter=="Radii")
      Radii_Ratio_internal(objects=objects, stainedcells=nucleus)
      objectfilter(radius_ratio>rd)
      end()
end()
objectfilter(area<ar)
CalcIntensity(Image=nucleus)
renameattr(RFP=intensity)
calcattr(intensity, "ln(RFP)", objects=objects)
set(intensitylow=objects.intensity.mean-(il*objects.intensity.stddev))
set(intensityhigh=objects.intensity.mean+(ih*objects.intensity.stddev))
if(IntensityFilterHigh)
objectfilter(intensity<intensityhigh)
end()
if(IntensityFilterLow)
objectfilter(intensity>intensitylow)
end()
Stencil2Objects(objects.body)
CalcAttr(area, objects=objects)
set(nuclei=objects)
if(ShowIllustrations)
imageview(nuclei.border, "nuclei", image=nucleus, gamma=2.6)
end()
}
```

7.2 Cytoplasm recognition script from nuclei

```
Proc Cytoplasm_from_nuclei (image stainedcells in "nuclei and cytoplasm
stained image", objectlist nuclei in "nuclei objects", bool
CellDensity=YES in "changes cytoplasm identification based on cell
desnsity", string standard_nuclei_removal="mean" in "other option
```

```
'convolution': this option will be used for low cell counts if
'CellDensity' method is used", string high_nuclei_removal="convolution"
in "only used if CellDensity is used: other option 'mean'", string
standard_cell_detection="threshold" in "other option 'density':  this
option will be used for low cell counts if CellDensity is used", string
high_cell_detection="density" in "other option 'threshold'", string
brightmask="none" in "use a bright mask instead of quantile for high/low
quadrants: options are 'high' 'low' 'both' or 'none'", bool minarray=YES
in "use a minarray for individual cell thresholding", bool maxarray=YES
in "use a maxarray for individual cell thresholding",  bool
IntensityFilterHigh=No in "filters nuclei + input # of  stdev away of the
mean RFP intensity", bool IntensityFilterLow=No in "filters nuclei -
input # of  stdev away of the mean RFP intensity", bool
ShowIllustrations=YES in, bool ShowTestingImages=NO in, string
RoundnessMethod="Radii" in "roundness filtering method uses either Radius
Ratio or Form Factor: possible values 'radii' 'Form'", objectlist
wholecells out, image nonucmean out, image highlowmask out)

Peters "cytoplasm detection from nuclei, the number of nuclei in does doe
always equal the number of cells out"
{
Input(bth, 12, "initial mask threshold: cytoplasm",
description="threshold adjustment for initial bright mask if not using
CellDensity, adjust based on initial wholecell mask")
Input(nuclei_high_count, 300, "nuclei high count:cytoplasm",
description="cutoff for a high density cell count")
Input(high_low_cut, 10, "high low cutoff: cytoplasm",
description="high/low cutoff for grid nuclei count, only for density
method, set to '0' if a low method is not wanted")
Input(high_quantile, 0.4, "high quantile: cytoplasm",
description="quantile for quadrants with high cell counts")
Input(low_quantile, 0.4, "low quantile: cytoplasm", description="quantile
for quadrants with low cell counts")
Input(minar12, 12, "minimum area: cytoplasm")
input(sm1, 4, "smoothing adjustment: cytoplasm")
input(tun1, 3, "individual threshold tuning: cytoplasm")
Input(zm, 3, "trimming threshold: cytoplasm", description="higher values
will trim more protrusions from cell, values too high will decrease
accuracy of cell borders. suggested range 1-4")
Set(zm2=-(zm+1))
Input(rd, 0.5, "roundness filter: cytoplasm")
Input(ar, 1000, "max area filter: cytoplasm")
Input(bd, 2, "border removal distance: cytoplasm")
input(ih, 2, "Intensity High: cytoplasm", "d", "number of stddev away
from mean for high intensity filter")
input(il, 2, "Intensity Low: cytoplasm", "d", "number of stddev away from
mean for low intensity filter")
Input(rt, 0.5, "nuclei/wholecell area ratio filter: cytoplasm")
///////////////////////////////////////////////////////////////////
//  sets methods for nuclei removal (n_meth) and cell thresholding
(cell_meth)
////////////////////////////////////////////////////////////////////////
If(!CellDensity)
      Set(n_meth=standard_nuclei_removal)
      Set(cell_meth=standard_cell_detection)
Else()
      If(CellDensity and nuclei.@count<nuclei_high_count)
            Set(n_meth=standard_nuclei_removal)
            Set(cell_meth=standard_cell_detection)
      Else()
            If(CellDensity and nuclei.@count>=nuclei_high_count)
```

```
                  Set(n_meth=high_nuclei_removal)
                  Set(cell_meth=high_cell_detection)
            End()
      End()
End()
////////////////////////////////////////////////////////////////////
// nuclei removal
////////////////////////////////////////////////////////////////////////
CalcMassCentre(objects=nuclei)
Mask(image=objects.masscentre.mask.image)
set(nucmass=mask)

If(n_meth=="mean")
      CalcZone(5, objects=nuclei, ZoneType="Equidistant")
      ZoneMask(-3, -2, objects=objects)
      CalcIntensity(zonemask, stainedcells, objects=objects)
      ZoneMask(-1, objects=objects
      CarryObjects(objects.zonemask, objects.zonemask_intensity,
image=stainedcells)
      set(nonuc=image)
      Mean(Image=nonuc)
      set(nonucmean=image)
Else()
      If(n_meth=="convolution")
            CalcZone(5, objects=nuclei, ZoneType="Equidistant")
            ZoneMask(-3, -2, objects=objects)
            CalcIntensity(zonemask, stainedcells, objects=objects)
            ZoneMask(-1, objects=objects)
            CarryObjects(objects.zonemask, objects.zonemask_intensity,
image=stainedcells)
            set(nonuc=image)
            Mean(Image=nonuc)
            Set(mean=image)
      set(convolutionkernel=toimage(vec(1,4,6,9,6,4,1,4,16,24,36,24,16,4,
6,24,36,54,36,24,6,9,36,54,81,54,36,9,6,24,36,54,36,24,6,4,16,24,36,24,16
,4,1,4,6,9,6,4,1),7,7).image)
            set(convolutionkernelfactor=convolutionkernel.sum)
            convolution(image=nonuc)
            set(nonucmean=image)
      Else()
            Error("cytoplasm: high_nuclei_removal or
standard_nuclei_removal is not correct, must be 'mean' or 'convolution'")
      End()
End()
//
If(cell_meth=="threshold")
      Bright_Mask(nonucmean, bth)
      //Or(m_bright, image=nuclei.body.mask.image)
      Set(bright_nuc=m_bright)
      set(highlowmask=bright_nuc)
      Mask2Stencil(bright_nuc)
      Stencil2Objects()
      split_stuck_objects(nonucmean, objects)
      FillObjects()
      set(initialobjects=objects)
      Set(nucgridimage=nuclei.body.image)
Else()
      If(cell_meth=="density")
            Blank(stainedcells.width, stainedcells.height)
            Set(blank=image)
            Bright_Mask(nonucmean, bth)
            ReadImage("C:\Documents and Settings\Opera\My Documents\Peter
acapella\Images\grid_dots.bmp")
            Redimension(stainedcells.width, stainedcells.height)
            Mask2Stencil(image)
            Stencil2Objects()
            CalcStencilFromCenters()
            CalcIntensity(stencilfromcenters, nucmass, Total=yes)
            Set(gridobj=objects) // objects from griddots image that
cover the full field of view

      Set(nucgridimage=gridobj.stencilfromcenters_border.mask.image)
// high objects
            ObjectFilter(stencilfromcenters_intensity>=high_low_cut,
objects=gridobj)
      If(brightmask=="high" or brightmask=="both")
            CarryPixels(objects.StencilFromCenters, m_bright,
image=blank)
            Set(highmask=image)
      Else()
            CalcStat("quantile",high_quantile, Image=stainedcells,
Stencil=stencilfromcenters)
            calcThreshMask(stencilfromcenters, quantile, Image=nonucmean,
Inverse=no)
            Set(highmask=objects.threshmask.mask.image)
      End()
            Set(high=objects)

// low objects
            ObjectFilter(stencilfromcenters_intensity<high_low_cut,
objects=gridobj)
      If(brightmask=="low" or brightmask=="both")
            CarryPixels(objects.StencilFromCenters, m_bright,
image=blank)
            Set(lowmask=image)
      Else()
            CalcStat("quantile",low_quantile, Image=nonucmean,
Stencil=stencilfromcenters)
            calcThreshMask(stencilfromcenters, quantile,
Image=stainedcells, Inverse=no)
            Set(lowmask=objects.threshmask.mask.image)
      End()
            Set(low=objects)
// combine objects
Or(lowmask, image=highmask)
//Or(image, image=nuclei.body.mask.image)
Set(highlowmask=image)
            Mask2Stencil(highlowmask)
            Stencil2Objects()
            split_stuck_objects(nonucmean, objects)
            FillObjects()
            set(initialobjects=objects)
      Else()
      Error("cytoplasm_from_nuclei, cell_method must equal 'general' or
'cell_conc'")
      End()
End()
// trimming ends off split object
Set(zmask=objects.body.mask.image)
CalcZone()
ZoneMask(zm, oo)
Stencil2Objects(objects.zonemask)
```

```
CalcZone(Stencil=zmask)
ZoneMask(zm2, oo)
Stencil2Objects(objects.zonemask)
calczone(ZoneType="equidistant", objects=objects)
CalcZoneIntensity(Image=nonucmean)
calcintensity(image=nonucmean,autorecalc=no)
set(initialzone_objects=objects)
If(minarray)
      set(objects=initialzone_objects)
      calczone(ZoneType="equidistant", Autorecalc=no)
      zonemask(-15,-4.0,autorecalc=no)
      calcstat("median",stencil=zonemask,image=nonucmean, autorecalc=no,
AttrName="attribute_median")
      threshmask(zonemask,Threshold=attribute_median, autorecalc=no,
image=stainedcells)
      calcintensity(threshmask,image=nonucmean,autorecalc=no)
      Set(minarrayxx=objects.threshmask_intensity)
End()

If(maxarray)
      set(objects=initialzone_objects)
      calczone(ZoneType="equidistant", Autorecalc=no)
      ZoneMask(2, oo, AutoRecalc=no)
      calcstat("median",stencil=zonemask,image=nonucmean, autorecalc=no,
AttrName="attribute2_median")
      threshmask(zonemask,Threshold=attribute2_median, autorecalc=no,
image=stainedcells)
      calcintensity(threshmask,image=nonucmean,autorecalc=no)
      Set(maxarrayxx=objects.threshmask_intensity)
End()
Set(objects=initialzone_objects)
If(minarray)
      If(maxarray)
            individualThreshold(MinArea=minar12, MaxMode="maxpoint",
Smooth=sm1, Image=nonucmean, Tune=tun1, MinArray=minarrayxx,
MaxArray=maxarrayxx)
      Else()
            IndividualThreshold(MinArea=minar12, MaxMode="maxpoint",
Smooth=sm1, Image=nonucmean, Tune=tun1, MinArray=minarrayxx)
      End()
Else()
      If(maxarray)
            IndividualThreshold(MinArea=minar12, MaxMode="maxpoint",
Smooth=sm1, Image=nonucmean, Tune=tun1, MaxArray=maxarrayxx)
      Else()
            IndividualThreshold(MinArea=minar12, MaxMode="maxpoint",
Smooth=sm1, Image=nonucmean, Tune=tun1)
      End()
End()
set(unfiltered_objects=objects)
////////////////////////////////////////// trimming cells
Set(zmask=objects.body.mask.image)
CalcZone()
ZoneMask(zm, oo)
Stencil2Objects(objects.zonemask)
CalcZone(Stencil=zmask)
ZoneMask(zm2, oo)
Stencil2Objects(objects.zonemask)
Set(trimmed_objects=objects)
split_stuck_objects(stainedcells)
////////////////////////////////////////// object filters
```

```
CalcIntensity(objects=objects, Image=nucmass, Total=yes)
ObjectFilter(intensity==1, objects=objects)
//filters out objects with more or less than one nuclei




// adding nuclei body to wholecell body to keep nuclei within cytoplasm
mask
Set(prenuc=objects)
Mask2Stencil(nucmass)
Stencil2Objects()
CalcIntensity(Image=prenuc.body.vector.image)
ObjectFilter(intensity>0, objects=objects)
CalcZone(50, Stencil=nuclei.body)
ZoneMask(-50)
Set(newnuc=objects)
CarryObjects(newnuc.zonemask, newnuc.intensity,
image=prenuc.body.vector.image)
Stencil2Objects(image)
Set(postnuc=objects)
Stencil2Objects(newnuc.zonemask.vector)
Set(cyto=objects)
Set(objects=postnuc)
Calcarea() //Calculates area
Calcarea(border) //Calculates border area
if(RoundnessMethod=="Radii")
      Radii_Ratio_internal(objects=objects, stainedcells=stainedcells)
      objectfilter(radius_ratio>rd)
else()
      if(RoundnessMethod=="Form")
            Calcattr(Roundness_corrected, 3.544*sqrt(area-
border_area/2.0)/border_area-0.1)
      objectfilter(Roundness_corrected>rd)
      else()
      Error("cytoplasm_from_nuclei: method type must be Radii or Form and
within quoatations")
      end()
end()
RemoveBorderObjects(bd)
objectfilter(area<ar)
CalcIntensity(Image=stainedcells)
calcattr(RFPintensity, "ln(intensity)", objects=objects)
set(RFPintensityhigh=objects.RFPintensity.mean+(ih*objects.RFPintensity.s
tddev))
set(RFPintensitylow=objects.RFPintensity.mean-
(il*objects.RFPintensity.stddev))
if(IntensityFilterHigh)
objectfilter(RFPintensity<RFPintensityhigh)
end()
if(IntensityFilterLow)
objectfilter(RFPintensity>RFPintensitylow)
end()
deleteattr(RFPintensity)
set(preWholeCells=objects)
CarryPixels(objects.body.mask.image, 0, image=nucmass)
//removing nuclei that have no cytoplasm
CalcIntensity(objects=cyto, Image=image, Total=yes)
ObjectFilter(intensity<1)
CalcIntensity(body, prewholecells.body.vector.image)
CarryObjects(objects.body, objects.intensity, image=objects.body.image)
```

```
Stencil2Objects(image)
CalcZone(50, ZoneType="equidistant", Stencil=postnuc.body)
ZoneMask(-50, -1)
CalcBorder(zonemask)
set(newcyto=objects)
set(objects=preWholeCells)
CalcArea()
//adding final attributes
SetAttr(centers, newcyto.body)
SetAttr(centers_border, newcyto.border)
SetAttr(cytoplasm, newcyto.zonemask)
SetAttr(cytoplasm_border, newcyto.zonemask_border)
CalcArea(centers)
CalcAttr(centers_body_ratio, "centers_area/area")
CalcMassCentre()
ObjectFilter(centers_body_ratio<rt)
/////////////////
MembraneRegion("body", no, WholeCells=objects)
CalcIntensity(Image=wholecells.body.mask.image, objects=wholecells)
CalcThreshMask(body, Image=wholecells.membraneregion.mask.image,
objects=objects, Inverse=yes, Threshold=intensity)
RenameAttr(inside=threshmask)
/////////////////
Set(WholeCells=objects)
if(showillustrations)
imageview(nuclei.body, "nuclei quadrants", image=nucgridimage, gamma=2.6)
end()
if(showtestingimages)
imageview(nonucmean, "Nuclei removal", gamma=2.6)
If(cell_meth=="density")
imageview(highmask, "high mask")
imageview(lowmask, "low mask")
end()
end()
if(showillustrations)
imageview(highlowmask, "initial cell mask")
end()
if(showtestingimages)
imageview(initialobjects.border, "initial objects", image=stainedcells,
gamma=2.6)
imageview(initialzone_objects.border, "pre-individual threshold",
image=stainedcells, gamma=2.6)
imageview(unfiltered_objects.border, "post-individual threshold",
image=stainedcells, gamma=2.6)
end()
if(showillustrations)
imageview(wholecells.border, "testwholecells", image=stainedcells,
gamma=2.6)
end()
}
```

7.3 Cytoplasm recognition script without nuclei

```
proc cytoplasm_without_nuclei (image cytoplasm in "cytoplasm stained
image to detect wholecells", bool Ploi_is_a_horrible_little_shit=YES
in, string ThreshMethod="Bright" in "method of thresholding: possible
values 'bright' or 'standard'", bool minarray=YES in, bool
maxarray=YES in, bool IntensityFilterHigh=No in "filters nuclei +
input # of  stdev away of the mean RFP intensity", bool
```

```
IntensityFilterLow=No in "filters nuclei - input # of  stdev away of
the mean RFP intensity", image RFP=none in "image for RFP intensity
filtering", string RoundnessMethod="Radii" in "roundness filtering
method uses either Radius Ratio or Form Factor: possible values
'radii' 'Form'", bool ShowTestingImages=NO in, bool
ShowIllustrations=YES in, objectlist WholeCells out "Output
objectlist of WholeCells")
Peters "Detects and outputs an objectlist contianing wholecells
without the use of nuclei"
{
Input(br1, 10, "threshold tuning bright: cytoplasm", "d", "only
required if method is set to bright")
Input(th1, 1.4, "threshold tuning standard: cytoplasm", "d", "only
required if method is set to standard")
input(rom1, 8, "range of maximums: cytoplasm", "d", "Cell splitting,
Point is discarded as object center if within a disk with radius
RangeOfMaximums it has the highest pixel value")
input(sdr1, 4, "smoothing disk radius: cytoplasm", "d", "Cell
splitting, Radius for smoothing disk")
Input(minar1, 80, "minimum area of initial wholecells : cytoplasm",
"d", "Cell splitting, minimum allowed area of split Cells")
input(minar12, 80, "minimum area of individual threshold adjusted
cells : cytoplasm", "d", "minimum allowed area of detected Cells")
Input(sm1, 4, "smoothing : cytoplasm", "d", "Controls the smoothness
of detected Cells. adjust based on WholeCells")
Input(tun1, 5, "individual threshold : cytoplasm", "d", "Controls the
individual threshold around which individual Cells are adjusted.
adjust based on WholeCells")
Input(zm, 3, "trimming threshold: cytoplasm", description="higher
values will trim more protrusions from cell, values too high will
decrease accuracy of cell borders. suggested range 1-4")
Set(zm2=-(zm+1))
input(rd1, 0.5, "roundness : cytoplasm", "d", "Controls the filtering
of Cells based on roundness. values closer to one should be more
round. adjust based on WholeCells")
input(ed1, 2, "border edge: cytoplasm", "d", "distance for border
removal")
input(ar1, 700, "maximum area: cytoplasm", "d", "max area filter for
identified cells")
Input(ar2, 80, "minum area: cytoplasm", "d", "mimimum area filter for
identified cells")
input(ih1, 2, "Intensity High: cytoplasm", "d", "number of stddev
away from mean for high intensity filter")
input(il1, 2, "Intensity Low: cytoplasm", "d", "number of stddev away
from mean for low intensity filter")



if(ThreshMethod=="Bright")
      Bright_Mask(cytoplasm, br1)
      set(mask=M_Bright)
            Else()
                  if(ThreshMethod=="standard")
                        ThresholdXX(th1, Image=cytoplasm)
                        Mask(threshold, image=cytoplasm)
                  else()
                        Error("Wholeobjectsonly: method type must be
Bright or Standard and within quoatations")
```

```
                end()
End() //if ThreshMethod

mask2stencil(mask)
      stencil2objects()
      set(cells_initial=objects)
fillobjects()
      split_stuck_objects(cytoplasm, RangeOfMaximums=rom1,
minarea=minar1, SmoothingDiskRadius=sdr1)
set(cells_initial_split=objects)
// trimming ends off split object
Set(zmask=objects.body.mask.image)
CalcZone()
ZoneMask(zm, oo)
Stencil2Objects(objects.zonemask)
CalcZone(Stencil=zmask)
ZoneMask(zm2, oo)
Stencil2Objects(objects.zonemask)
calczone(ZoneType="equidistant", objects=objects)
CalcZoneIntensity(Image=cytoplasm)
calcintensity(image=cytoplasm,autorecalc=no)
set(initialzone_objects=objects)
If(minarray)
      set(objects=initialzone_objects)
      calczone(ZoneType="equidistant", Autorecalc=no)
      zonemask(-15,-4.0,autorecalc=no)
      calcstat("median",stencil=zonemask,image=cytoplasm,
autorecalc=no, AttrName="attribute_median")
      threshmask(zonemask,Threshold=attribute_median, autorecalc=no,
image=cytoplasm)
      calcintensity(threshmask,image=cytoplasm,autorecalc=no)
      Set(minarrayxx=objects.threshmask_intensity)
End()
If(maxarray)
      set(objects=initialzone_objects)
      calczone(ZoneType="equidistant", Autorecalc=no)
      ZoneMask(2, oo, AutoRecalc=no)
      calcstat("median",stencil=zonemask,image=cytoplasm,
autorecalc=no, AttrName="attribute2_median")
      threshmask(zonemask,Threshold=attribute2_median, autorecalc=no,
image=cytoplasm)
      calcintensity(threshmask,image=cytoplasm,autorecalc=no)
      Set(maxarrayxx=objects.threshmask_intensity)
End()

Set(objects=initialzone_objects)
If(minarray)
      If(maxarray)
            individualThreshold(MinArea=minar12, MaxMode="maxpoint",
Smooth=sm1, Image=cytoplasm, Tune=tun1, MinArray=minarrayxx,
MaxArray=maxarrayxx)
      Else()
            IndividualThreshold(MinArea=minar12, MaxMode="maxpoint",
Smooth=sm1, Image=cytoplasm, Tune=tun1, MinArray=minarrayxx)
      End()
Else()
      If(maxarray)
```

```
            IndividualThreshold(MinArea=minar12, MaxMode="maxpoint", 
Smooth=sm1, Image=cytoplasm, Tune=tun1, MaxArray=maxarrayxx)
      Else()
            IndividualThreshold(MinArea=minar12, MaxMode="maxpoint", 
Smooth=sm1, Image=cytoplasm, Tune=tun1)
      End()
End()
set(unfiltered_objects=objects)
/////// trimming cells
Set(zmask=objects.body.mask.image)
CalcZone()
ZoneMask(zm, oo)
Stencil2Objects(objects.zonemask)
CalcZone(Stencil=zmask)
ZoneMask(zm2, oo)
Stencil2Objects(objects.zonemask)
Set(trimmed_objects=objects)
split_stuck_objects(cytoplasm)
//          Object Filters
RemoveBorderObjects(ed1)
Calcarea()
Calcarea(border)
if(RoundnessMethod=="Radii")
      Radii_Ratio_internal(objects=objects, stainedcells=cytoplasm)
      objectfilter(radius_ratio>rd1)
else()
      if(RoundnessMethod=="Form")
            Calcattr(Roundness_corrected, 3.544*sqrt(area-
border_area/2.0)/border_area-0.1)
      objectfilter(Roundness_corrected>rd1)
      else()
      Error("cytoplasm_without_nuclei: method type must be Radii or 
Form and within quoatations")

      //roundness
      end()
end()
objectfilter(area<ar1)
ObjectFilter(area>ar2)
set(filtered_objects=objects)

      //intensity filter
CalcIntensity(Image=cytoplasm)
calcattr(RFPintensity, "ln(intensity)", objects=objects)
set(RFPintensityhigh=objects.RFPintensity.mean+(ih1*objects.RFPintens
ity.stddev))
set(RFPintensitylow=objects.RFPintensity.mean-
(il1*objects.RFPintensity.stddev))
if(IntensityFilterHigh)
objectfilter(RFPintensity<RFPintensityhigh)
end()
if(IntensityFilterLow)
objectfilter(RFPintensity>RFPintensitylow)
end()
Stencil2Objects(objects.body)
fillobjects()
CalcAttr(area, objects=objects)
set(WholeCells=objects)
```

```
if(ShowIllustrations)
      imageview(mask, "initial wholecell mask")
End()

if(ShowTestingImages)
      imageview(cells_initial.border, "initial wholecell objects",
image=cytoplasm, gamma=2.6)
      imageview(cells_initial_split.border, "split wholecell
objects", image=cytoplasm, gamma=2.6)
      imageview(initialzone_objects.zone, "inner wholecell zones",
image=cytoplasm)
      imageview(initialzone_objects.outerzone, "outer wholecell
zones", image=cytoplasm)
      imageview(unfiltered_objects.border, "individual wholecell
threshold", image=cytoplasm, gamma=2.6)
end()
if(ShowIllustrations)
      imageview(Wholecells.border, "WholeCells", image=cytoplasm,
gamma=2.6)
End()

}
```

7.4 Combine stack script (mid-section selection)

```
Proc CombineStack (double NumberOfChannels=2 in "number of channels",
double NumberOfFields=1 in "number of fields", double
NumberOfZplanes=5 in "number of Z planes", bool AllImages=YES in "use
all images in stack, if No then select stacks to use in ImagePrep
inputs", bool ImageAddition=YES in "adds images in a stack together",
bool MaxProjection=YES in "Maxprojection of images in a stack", bool
FocusImage=YES in "finds the most in focus image within a stack",
bool BGC=NO in "Rollingball Background correction", bool
ShowIllustrations=YES in, Memblock CombinedPack out)
Peters "Combines images in a stack by adding together, images are
output within a Package by channel and field as
ComImage_method_C#_F#. Module must be follwed by the module:
unpack(Combinedpack)"
{
singlewell()
Set(num_ch=NumberOfChannels)
Set(num_planes=NumberOfZplanes)
Set(num_fields=NumberOfFields)
input(er, 15, "BGC edge: ImagePrep", "d", "edge size for background
correction erosion")
input(firstplane, 1, "first plane: ImagePrep", "d", "first plane to
use from stack")
input(lastplane, 5, "last plane: ImagePrep", "d", "last plane to use
from stack")
if(AllImages)
set(firstplane=1)
set(lastplane=num_planes)
end()
//////////////// ADDITION
if(ImageAddition)
set(m="Plus")
```

```
foreach(1..num_fields, "f")

Foreach(1..num_ch, "c")
set(i2=1)
Foreach(firstplane..lastplane, "p")
      if(i2==1)
            set(zz=c+(f-1)*num_planes*num_ch)
            plus(_["image" &zz], _["image" &zz+num_ch])
set(i2=i2+1)

      else()
            while(i2<=lastplane-firstplane)
                  set(i3=zz+i2*num_ch)
                  plus(result, _["image" &i3])
                  set(i2=i2+1)
            end() //while
      end() //if(i2==1)
set(_["ComImage_"&m&"_C"&c&"_F"&f]=result)

end() //planes
if(ShowIllustrations)
      imageview(_["ComImage_"&m&"_C"&c&"_F"&f],
"Plus_Field"&f&"Channel" & c)
end()
                        //rollingballing background correction plus
      if(BGC)
            resize(1/4, image=result, yfactor=1/4)
            eros(edge=er)
            resize(4, image=image, yfactor=4)
                  set(backval = image.mean)
                  set(xslip = (result.width - image.width)/2)
                  set(yslip = (result.height - image.height)/2)
                  redimension(result.width, result.height, xslip,
yslip, backval, image=image, BackgroundMethod="mirror")
                        set(BGimage = image)
                  minus(result, BGimage, neg_method="zero")
                        set(_["ComImage_"&m&"_BGC_C"&c&"_F"&f]=result)
                  if(ShowIllustrations)
                  imageview(_["ComImage_"&m&"_BGC_C"&c&"_F"&f],
"BGC_"&m&"_Field_"&f&"Channel" & c)
                  end()
      end()
end() //channels
end() //fields
end()

////////////////////////////////////////////////// MAX PROJECTION
if(MaxProjection)
set(m="Max")
foreach(1..num_fields, "f")
set(CubeDepth=Lastplane-Firstplane+1)
foreach(1..num_ch, "c") // For each channel
            create("datacube",image1.width,image1.height,CubeDepth,
"unsigned short")
            set(j=(c-num_ch)+(firstplane*num_ch))
            set(i2=0)
            foreach(Firstplane..Lastplane, "p")  // For each plane
```

```
                  Set(zz=(c-num_ch)+(p*num_ch)+((f-
1)*num_planes*num_ch))
                        set(datacube[i2]=_["image" & zz ])
                        set(i2=i2+1)
                  set(j=j+num_ch)
      end() // For each plane
            Maximums3D(0,image=datacube)
            StencilFrom3DTo2D(stencil=maximums,datacube=datacube)
//Projects the found maximums from 3-dim to plane
            delete(datacube)
            set(_["ComImage_"&m&"_C"&c&"_F"&f] = ValueImage)
                  if(ShowIllustrations)
                  imageview(_["ComImage_"&m&"_C"&c&"_F"&f],
"Max_Field"&f&"Channel" & c)
                  end()
//    foreach(1..num_ch, "c") // For each channel
//          create("datacube",image1.width,image1.height,CubeDepth,
"unsigned short")
//          set(j=c+(f-1)*num_planes*num_ch)
//          set(i2=0)
//          foreach(Firstplane..Lastplane, "p")  // For each plane
//                if(i2==0)
//                      set(datacube[i2]=_["image" & c+(f-
1)*num_planes*num_ch ])
//                else()
//                      set(datacube[i2]=_["image" & j ])
//                end()
//                set(i2=i2+1)
//                set(j=j+num_ch)
//          end() // For each plane
//          Maximums3D(0,image=datacube)
//          StencilFrom3DTo2D(stencil=maximums,datacube=datacube)
//Projects the found maximums from 3-dim to plane
//          delete(datacube)
//          set(_["ComImage_"&m&"_C"&c&"_F"&f] = ValueImage)
//
//
//                if(ShowIllustrations)
//                imageview(_["ComImage_"&m&"_C"&c&"_F"&f],
"Max_Field"&f&"Channel" & c)
//                end()
                        //rollingballing background correction max
      if(BGC)
            resize(1/4, image=ValueImage, yfactor=1/4)
            eros(edge=er)

            resize(4, image=image, yfactor=4)
                  set(backval = image.mean)
                  set(xslip = (ValueImage.width - image.width)/2)
                  set(yslip = (ValueImage.height - image.height)/2)

                  redimension(ValueImage.width, ValueImage.height,
xslip, yslip, backval, image=image, BackgroundMethod="mirror")
                        set(BGimage = image)
                  minus(ValueImage, BGimage, neg_method="zero")
                        set(_["ComImage_"&m&"_BGC_C"&c&"_F"&f]=result)
                  if(ShowIllustrations)
                  imageview(_["ComImage_"&m&"_BGC_C"&c&"_F"&f],
"BGC_"&m&"_Field_"&f&"Channel" & c)
                  end()
      end()
      end() // For each channel
end()
end()
//////////////////////////////// Focus Image
if(FocusImage)
if(!Allimages)
Set(num_planes=lastplane-firstplane+1)

end()
Set(m="Focus")
Foreach(1..num_fields, "f")
      Foreach(1..num_ch, "c")
      Set(i2=1)
            //set(zz=c+(f-1)*num_planes*num_ch)
            Foreach(firstplane..lastplane, "p")
                  Set(zz=(c-num_ch)+(p*num_ch)+((f-
1)*num_planes*num_ch))
                  If(i2==1)
                        mean(image=_["image"&zz])
                        Gradient(image=image)
                        Set(temp=image.stddev)
                        //Set(_["stddev"&zz]=temp)
                        Set(temp2=_["image"&zz])
                        Set(temp3="image"&zz)
                        set(temp4=zz)
                        Set(i2=i2+1)
                        //Set(zz=zz+num_ch)
                  Else()
                        if(i2<=lastplane)
                              mean(image=_["image"&zz])
                              Gradient(image=image)
                              Set(_["stddev"&zz]=image.stddev)
                              If(image.stddev > temp)
                                    Set(temp=image.stddev)
                                    Set(temp2=_["image"&zz])
                                    Set(temp3="image"&zz)
                                    Set(temp4=zz)
                              End() //if image.stddev
                              Set(i2=i2+1)
                              //Set(zz=zz+num_ch)
                        End() // while
                  End()//if i2==1 else
            Set(_["ComImage_"&m&"_C"&c&"_F"&f]=temp2)
            Set(_["ComImage_"&m&"_C"&c&"_F"&f&"_Image_name"]=temp3)
            set(_["ComImage_"&m&"_C"&c&"_F"&f&"_Image_no"]=temp4)
            //Set(_["C"&c&"F"&f&"mean"]=temp)
            End()
      //num planes
if(ShowIllustrations)
      imageview(_["ComImage_"&m&"_C"&c&"_F"&f],
"Focus_Field"&f&"Channel" &c)
end()
                        //rollingballing background correction plus
      if(BGC)
```

```
            resize(1/4, image=temp2, yfactor=1/4)
            eros(edge=er)

            resize(4, image=image, yfactor=4)

                  set(backval = image.mean)
                  set(xslip = (temp2.width - image.width)/2)
                  set(yslip = (temp2.height - image.height)/2)

                  redimension(temp2.width, temp2.height, xslip, yslip,
backval, image=image, BackgroundMethod="mirror")
                        set(BGimage = image)

                  minus(temp2, BGimage, neg_method="zero")
                        set(_["ComImage_"&m&"_BGC_C"&c&"_F"&f]=result)

                  if(ShowIllustrations)
                  imageview(_["ComImage_"&m&"_BGC_C"&c&"_F"&f],
"BGC_"&m&"_Field"&f&"Channel"&c)
                  end()
      end()
      End()
      //num ch
End()
      //num fields
end()
/////////////////// Package
pack(CombinedPack, ComImage_*)
}
```

7.5 Z-score script

```
Proc Z_score (objectlist control_objects in "control objects",
objectlist treated_objects in "treated objects", string
variable="nan" in "attribute list to be tested'", string Z_score out
"Z-score")
Peters "Z score = treated_median-control_median/control_MAD"
{
Set(values=_["control_objects."&variable])
company ( "abs(values - values.median)" )
Set(z_score=(_["treated_objects."&variable&".median"]-
_["control_objects."&variable&".median"])/(result.median * 1.4826))
}
```

7.6 Mann-Whitney U test script

```
Proc Mann_Whitney (objectlist obj1 in "object group one", objectlist
obj2 in "object group two", string data="nan" in "attribute list to
be tested'", double Z_ratio out "Z-score")
Peters "Mann-Whitney test"
{
///////////////combining object lists
CalcAttr(label, "1", objects=obj1)
Rename(obj1=objects)
CalcAttr(label, "2", objects=obj2)
Rename(obj2=objects)
```

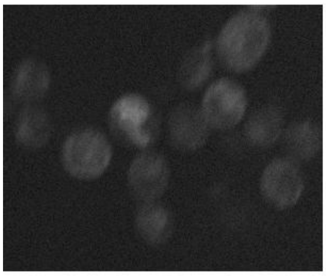
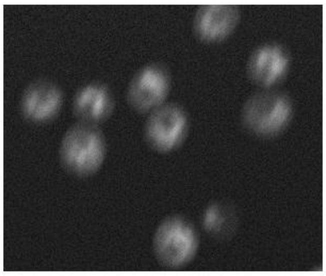
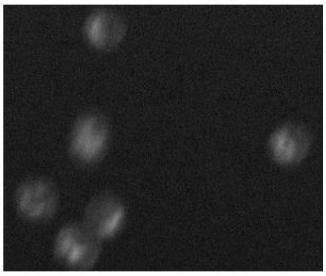
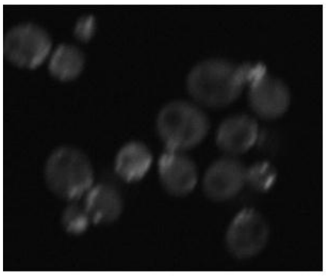
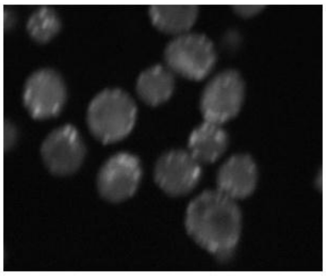
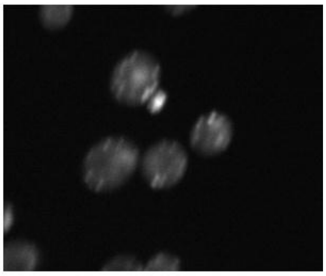
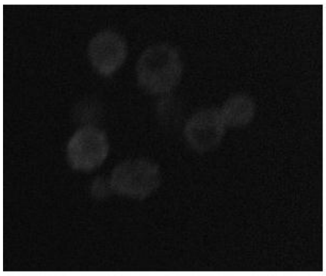
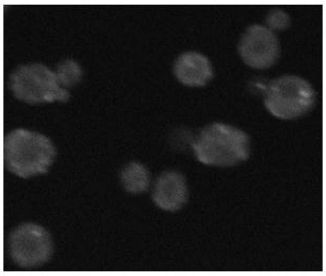
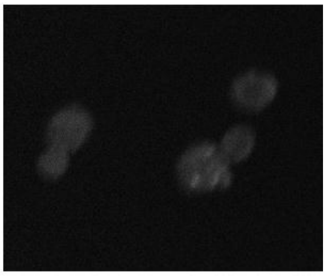
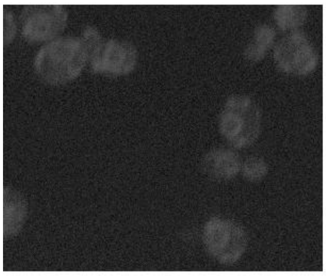
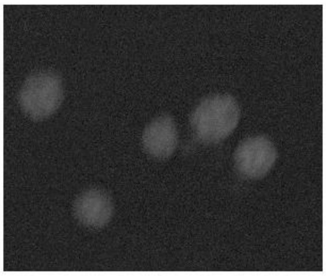
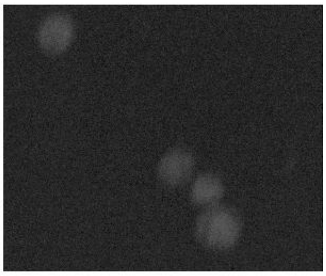
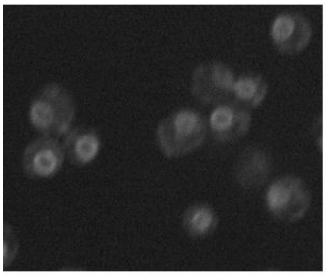
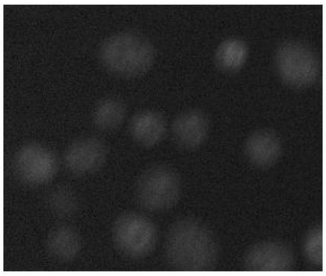
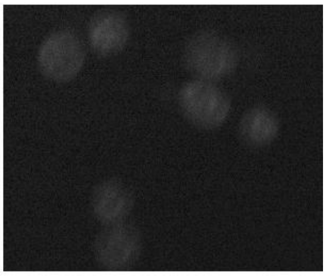
```
AddObjects(obj1, objects=obj2, CheckOverlap=no)
Rename(obj_all=objects)

////////////////////////////ranking
Sort(_["obj_all."&data])
Set(data_orig=_["obj_all."&data])
Sort_Prepare(data_orig, 0)

Set(v1=vec())
Foreach(0..obj_all.@count-2)
If(result[i]==result[i+1])
Set(same=1)
Else()
Set(same=0)
End()
Push(v1, same)
End()
Push(v1,0)
ConvElems(result, "float")
Set(result1=result)
Set(x="nan")
Foreach(0..obj_all.@count-1)
If(i<x+1)
Else()
      If(v1[i]==1)
      Set(med=i+1+i+1+1)
      Set(x=i+1)
            While(v1[x])
            Set(x=x+1)
            Set(med=med+x+1)
                  End()
            Foreach(i..x,"j")
            Set(result[j]=med/(x+-i+1))
            Set(test=med/(x-i+1))
            End()
Else()
Set(result[i]=i+1)
End()
End()
End()
////// Mann-Whitney Test
Set(resultx=result)
Create("vector", "float", data_orig.size, 0)
BackOrder(resultx, target=vector)
SetAttr(rank, result, objects=obj_all)
If(obj1.@count<obj2.@count)
      Set(obj_small=obj1)
      Set(lab=1)
      Else()
      Set(lab=2)
      Set(obj_small=obj2)
End()
ObjectFilter("label==lab")
Set(sqrt=sqrt(obj1.@count*obj2.@count*(obj_all.@count+1)/12))
Set(uT=obj_small.@count*(obj_all.@count+1)/2)
Set(Z_ratio=(objects.rank.sum-uT+.5)/sqrt)
if(lab==1)
set(z_ratio=-z_ratio)
```

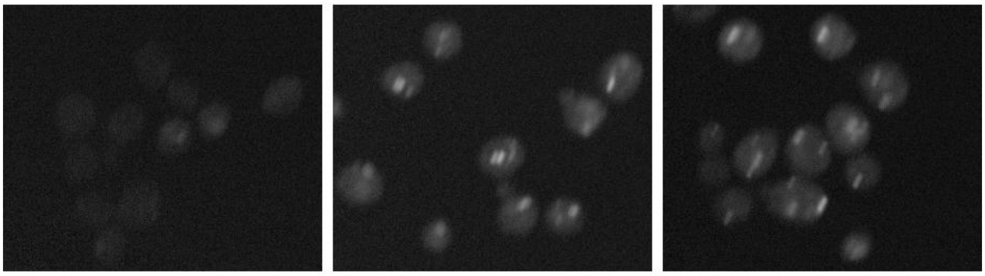
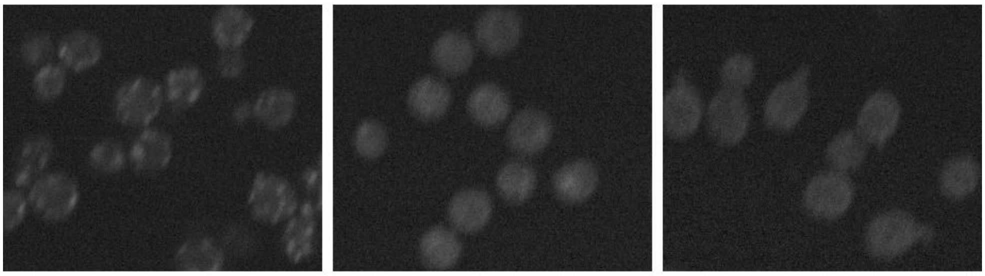
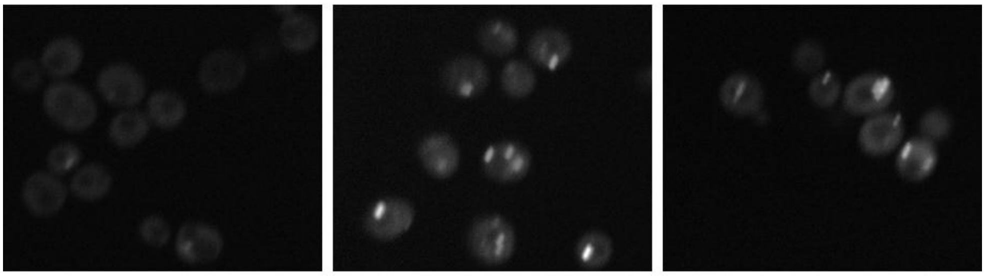
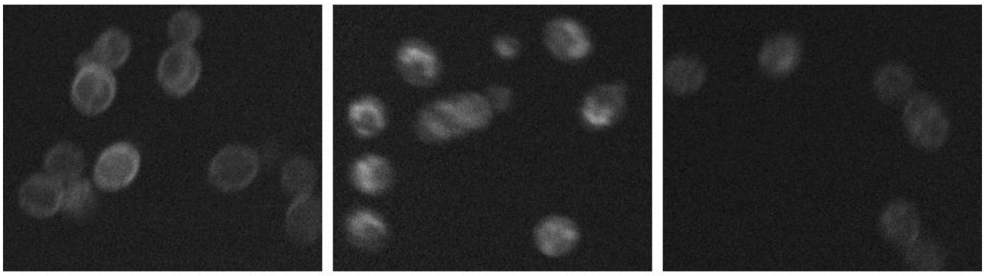
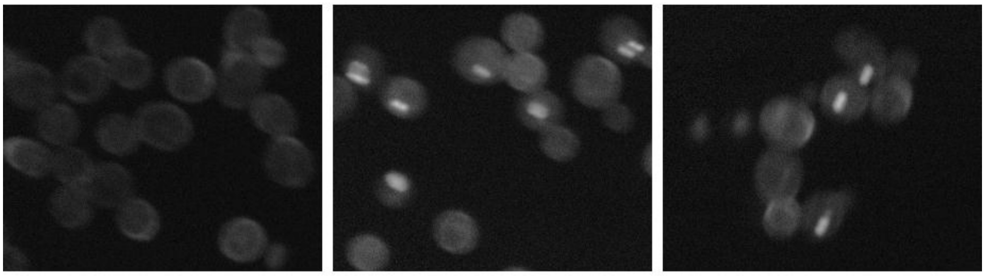
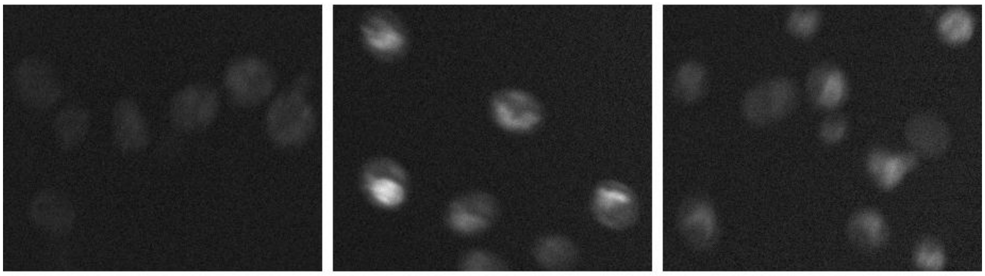
```
end()
}
```

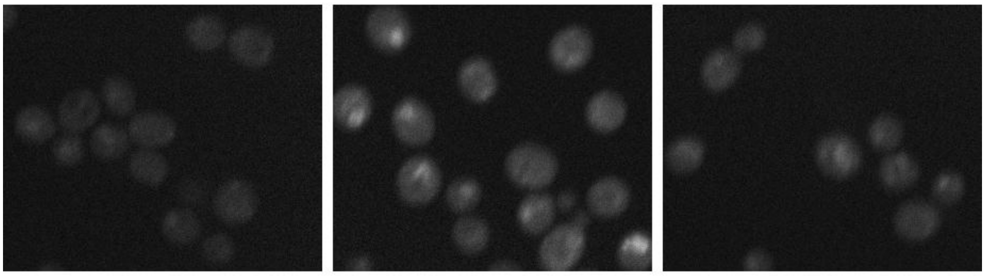
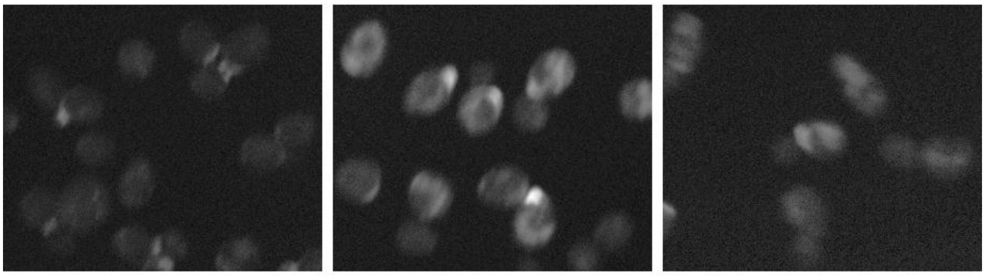
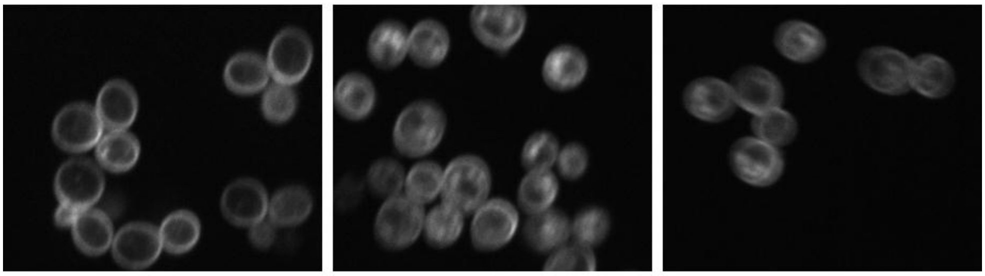
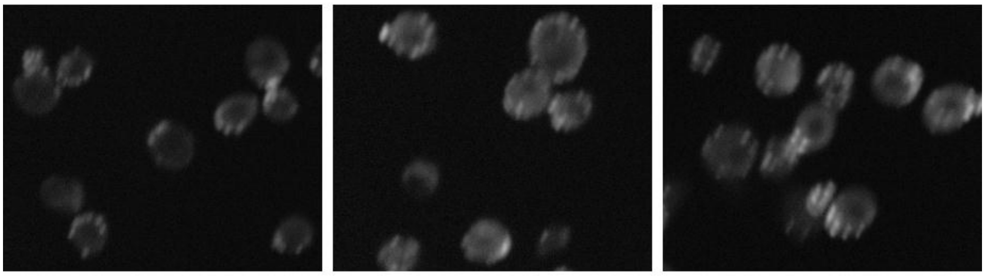
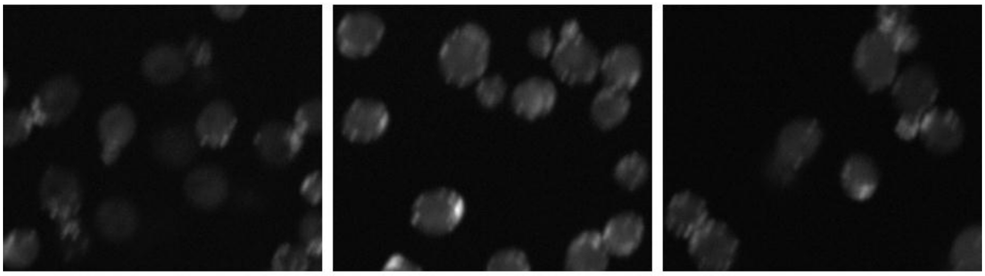
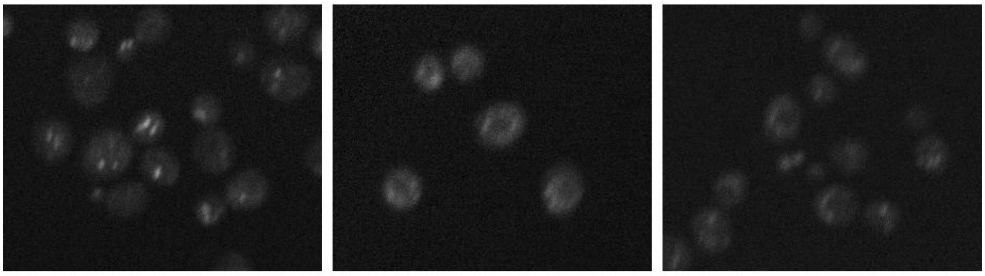
7.7 Radii ratio script

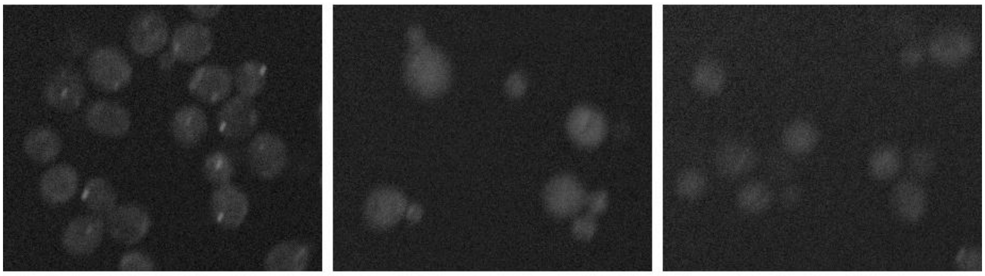
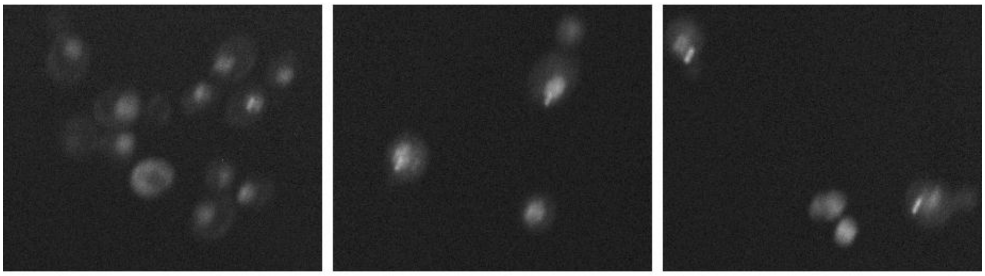
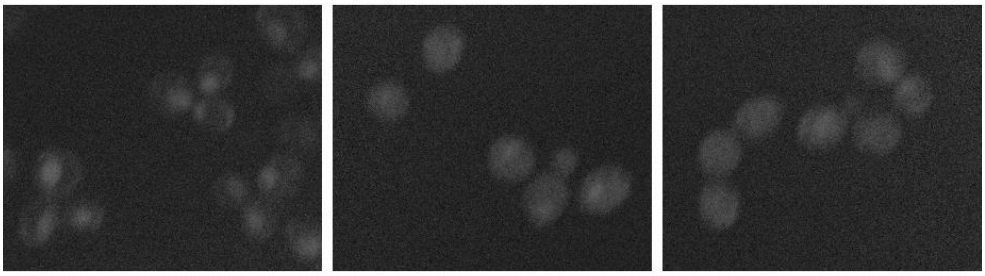
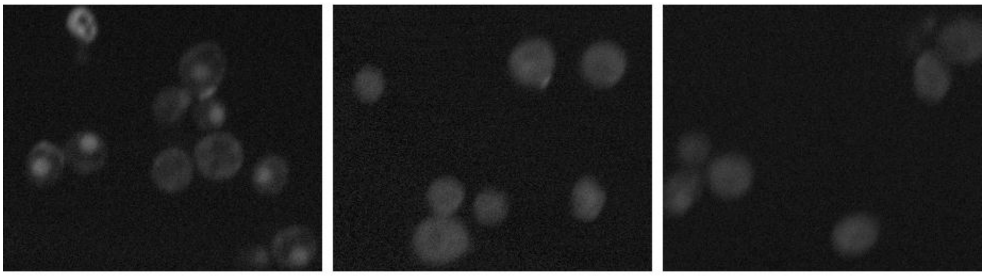
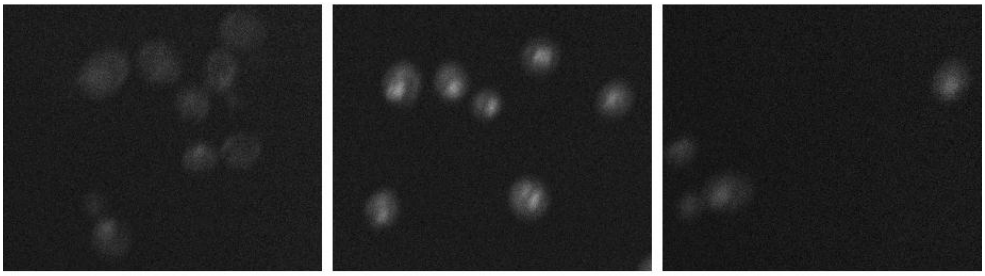
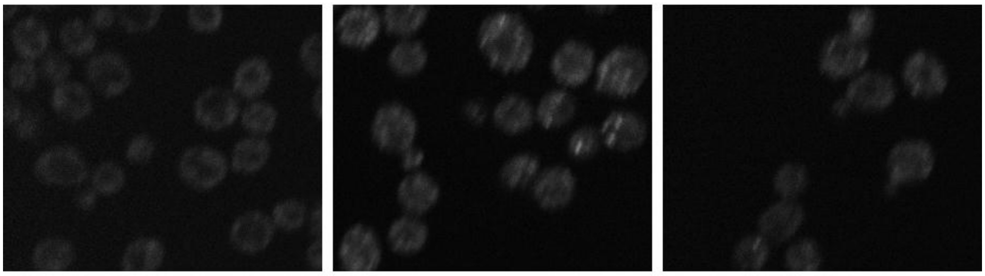
```
Proc Radii_Ratio (objectlist objects in "wholecells, body will be
used to measure roundness", bool ShowIllustrations=NO in, image
StainedCells=none in, bool OutputResults=YES in, objectlist objects
out)
Peters "Calculates the roundness of cells using the radii ratio
method, generally more robust and resolustion independent than form
factor methods used in the roundness_corrected attribute"
{
set(initial_cells=objects)
Input(rr, 0, "roundness ratio high: object_attributes",
description="roundness filter, objects with roundness above this
threshold will be kept. range 0-1")
Input(rr2, 1, "roundness ratio low: object_attributes",
description="roundness filter, objects with roundness below this
threshold will be kept. range 0-1")
CalcMassCentre(objects=objects)
Stencil2Objects(objects.masscentre)
CalcZone(50, ZoneType="equidistant", Stencil=initial_cells.body)
ZoneImage(ZoneType="outerZones")
DistanceImage()
CalcStat("min", Stencil=border, Image=distanceimage,
objects=initial_cells)
CalcStat("max", Stencil=border, Image=distanceimage, objects=objects)
CalcAttr(radius_ratio, "min/max")
ObjectFilter(radius_ratio>rr, objects=objects)
ObjectFilter(radius_ratio<rr2, objects=objects)
If(ShowIllustrations)
ImageView(objects.body, "radius ratio", image=StainedCells)
End()
if(OutputResults)
output(objects.radius_ratio.mean, "radius_ratio")
end()
}
```

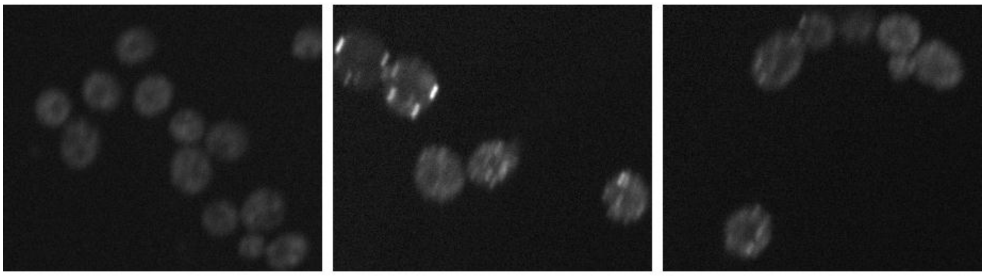
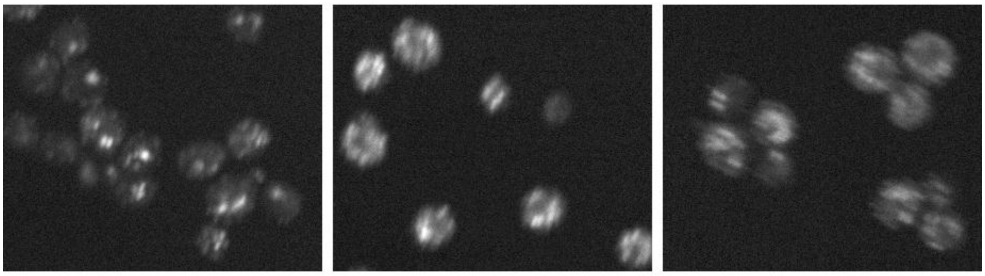
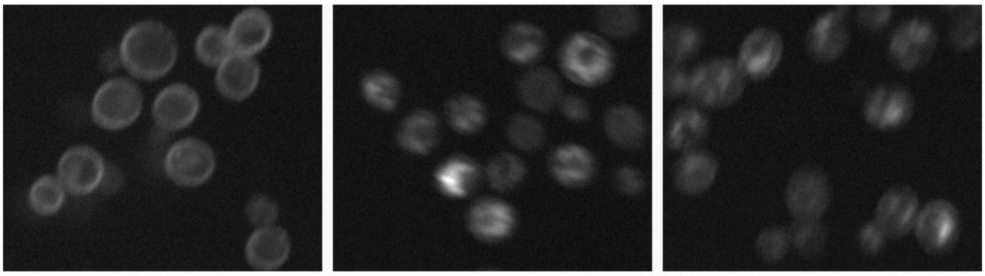
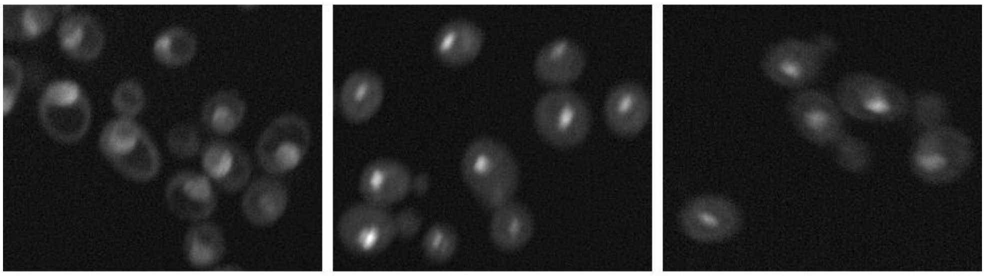
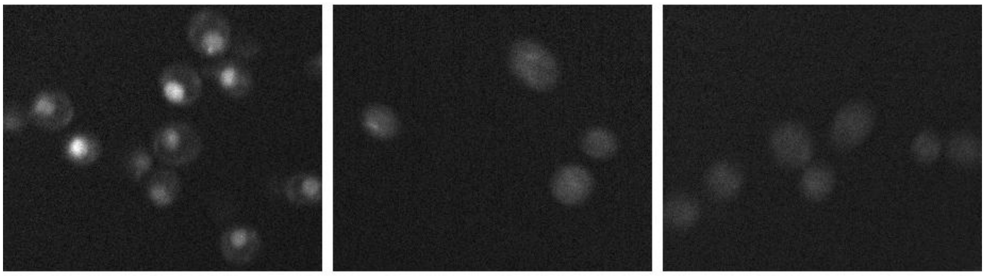
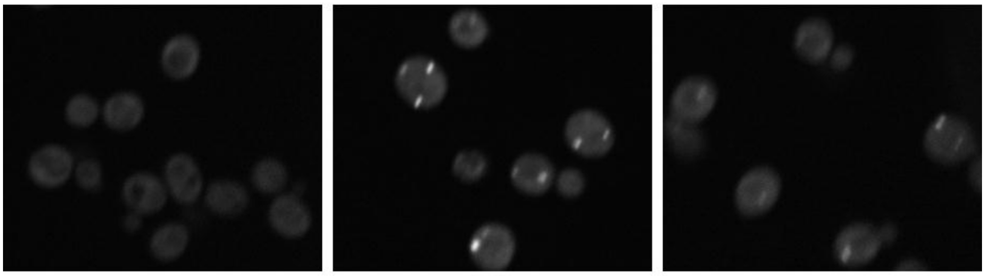
7.8 DTT and TM induced localisation changes

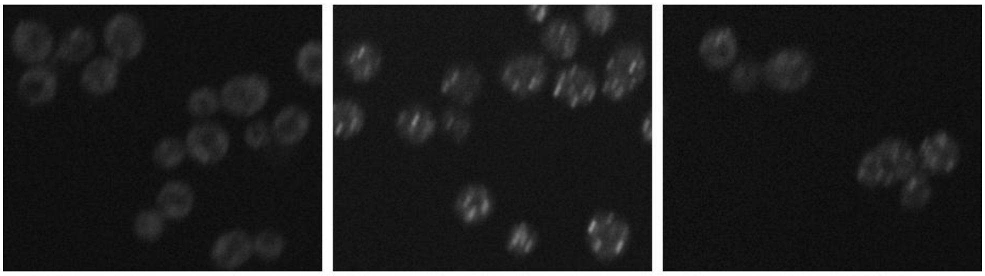
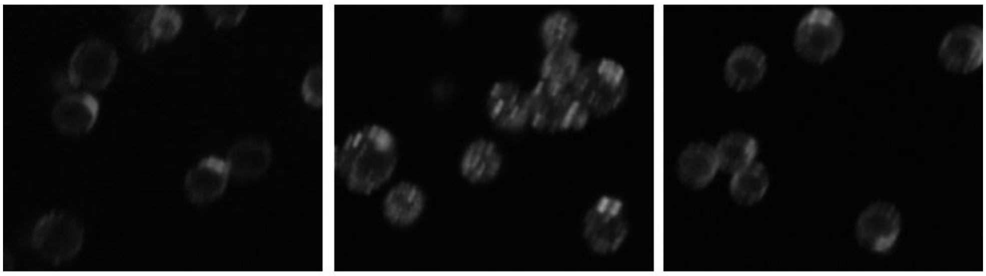
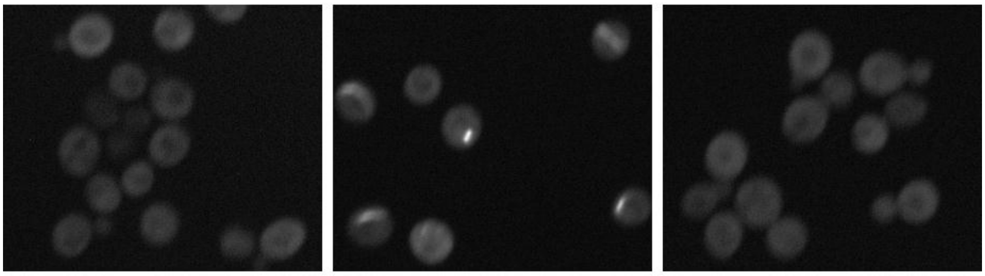
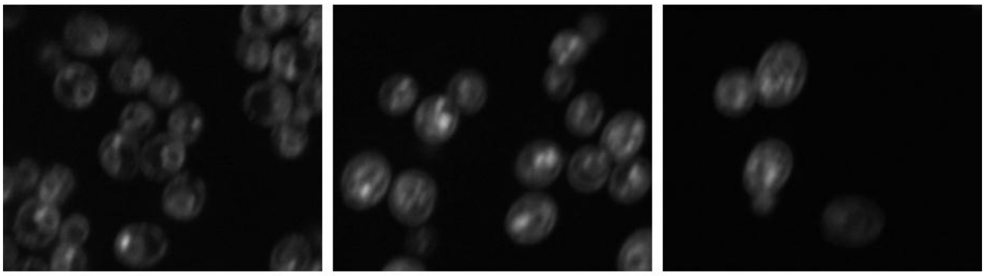
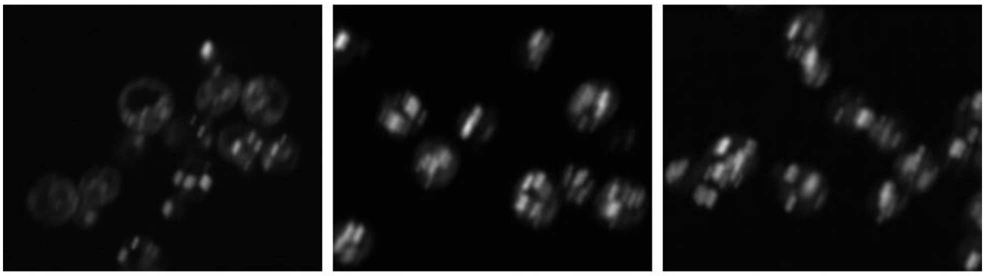
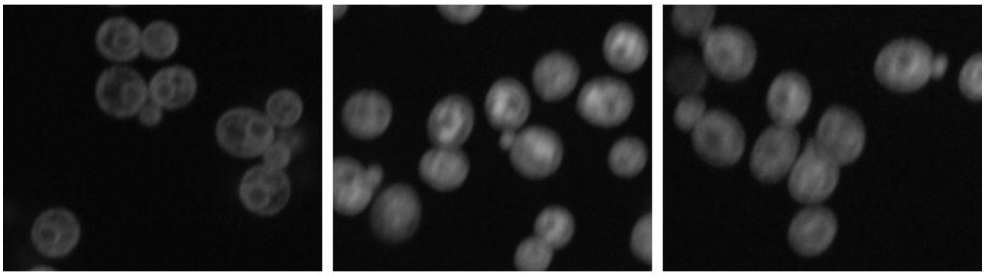
Control	DTT	TM	Gene/ORF notes
			AGP2 YBR132C
			AIP1 YMR092C
			AKL1 YBR059C
			ALG2 YGL065C
			ALG9 YNL219C

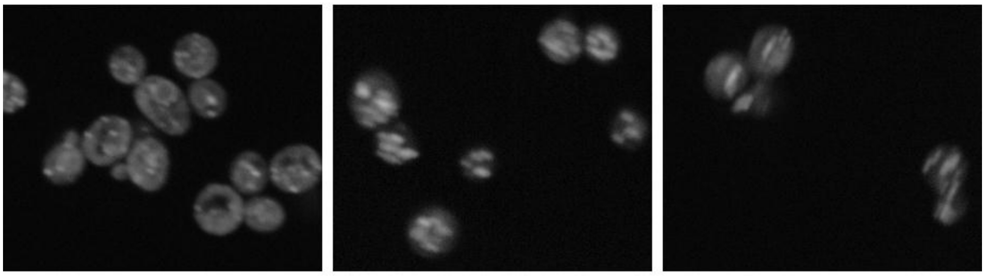
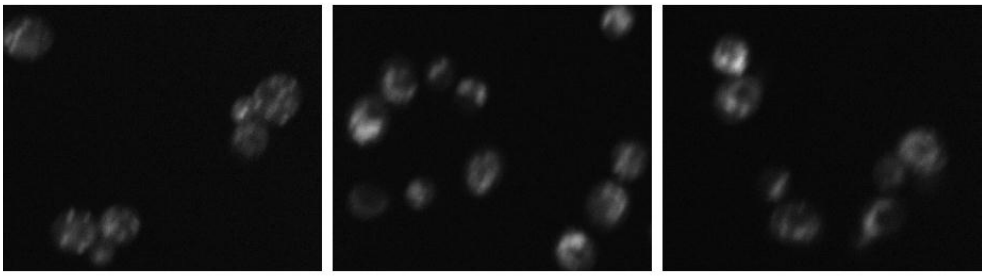
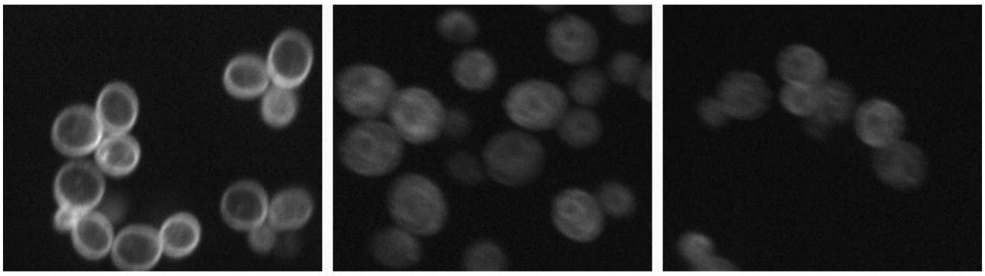
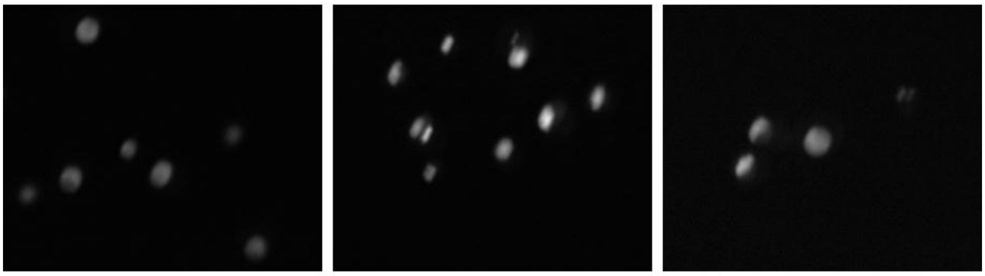
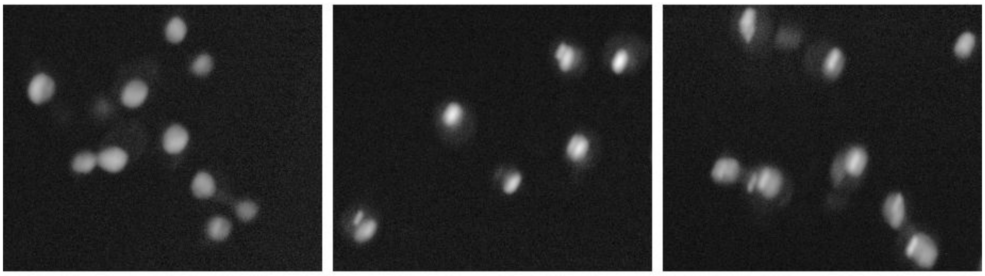
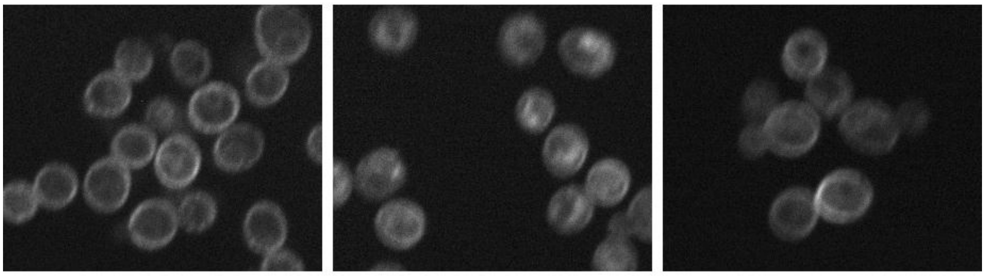
	AMS1 YGL156W
	ANP1 YEL036C
	APE1 YKL103C
	AQR1 YNL065W
	ARF3 YOR094W
	ARN1 YHL040C

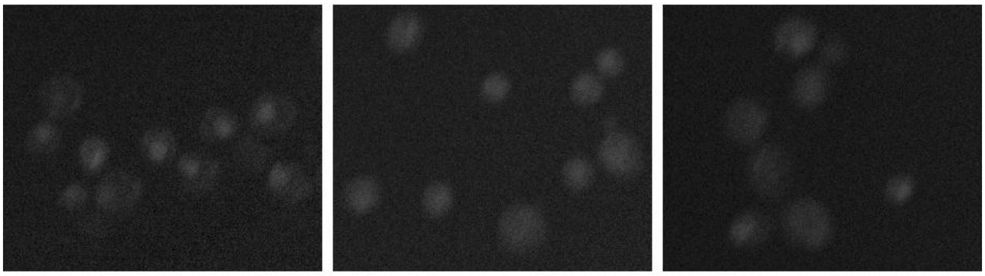
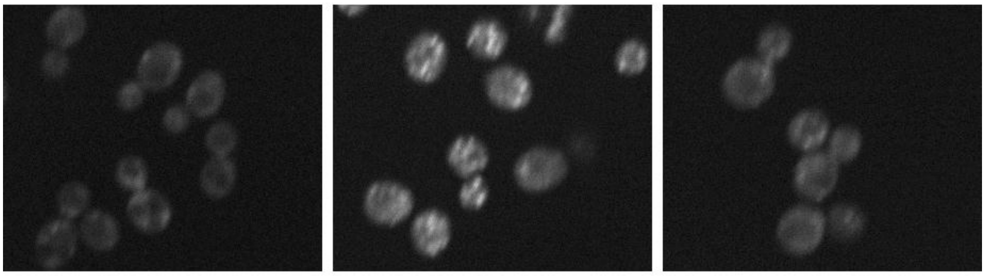
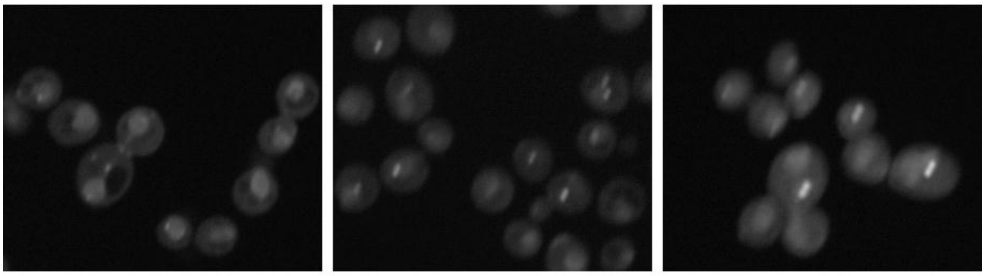
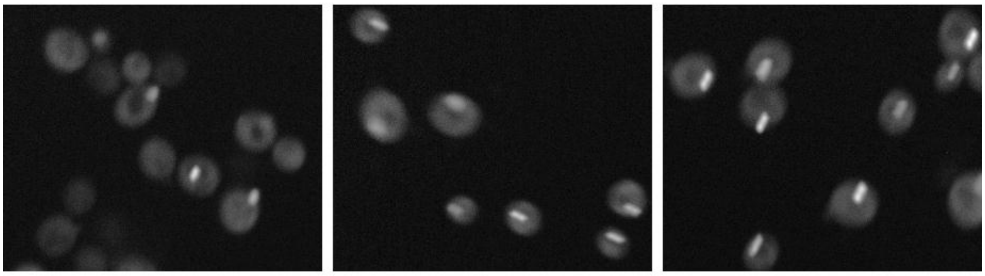
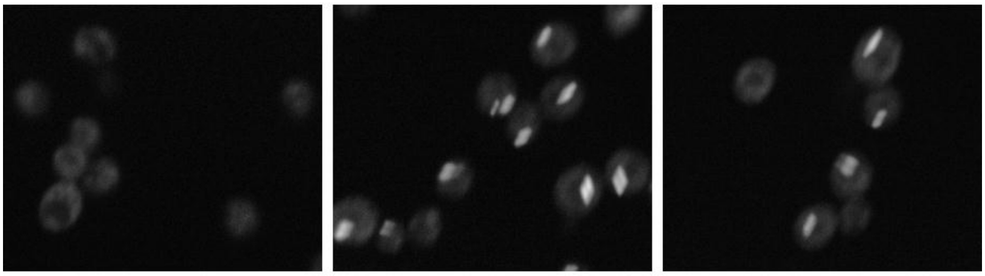
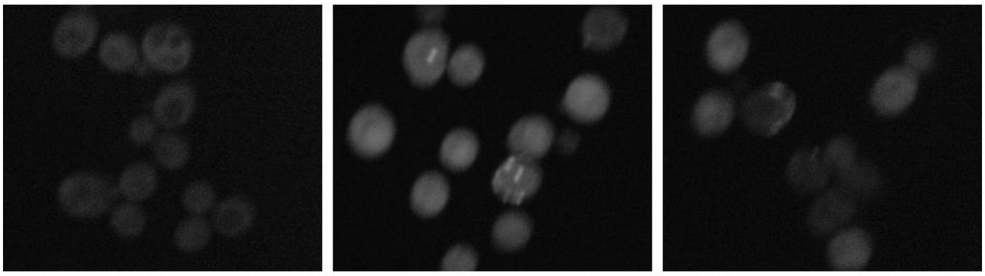
	ATG18 YFR021W
	AXL2 YIL140W
	BAP2 YBR068C
	CAP1 YKL007W
	CAP2 YIL034C
	CCZ1 YBR131W

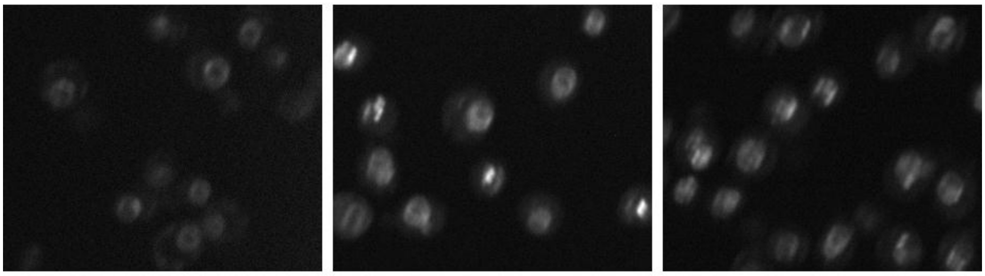
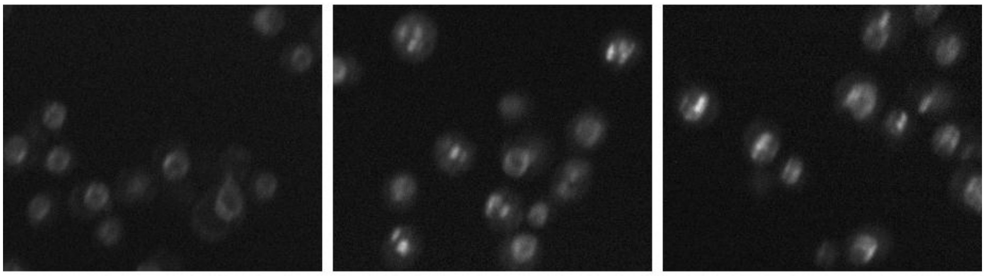
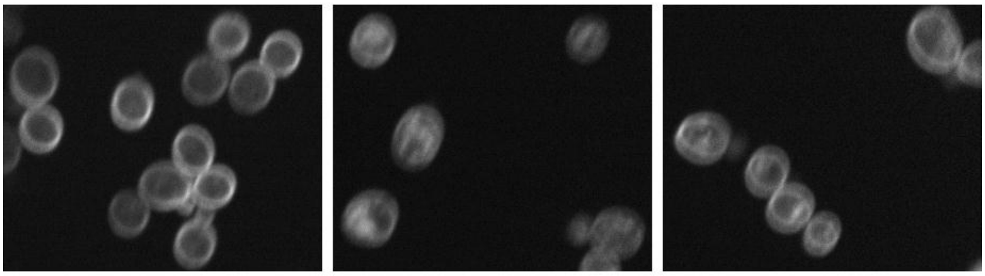
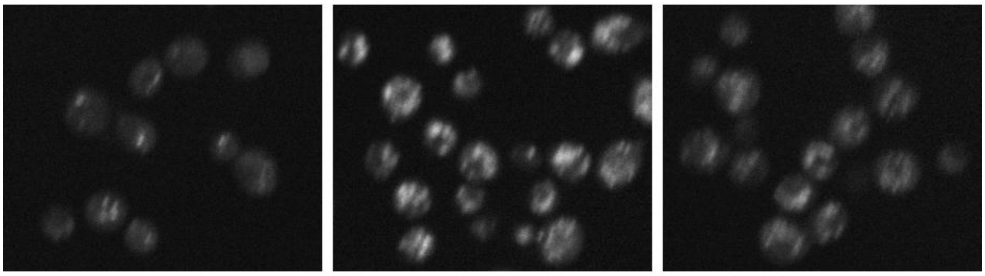
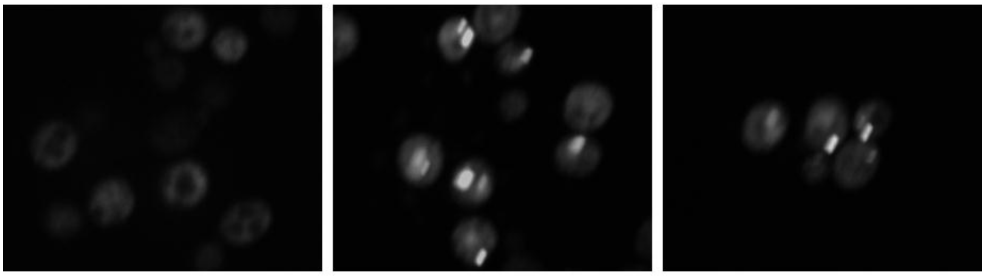
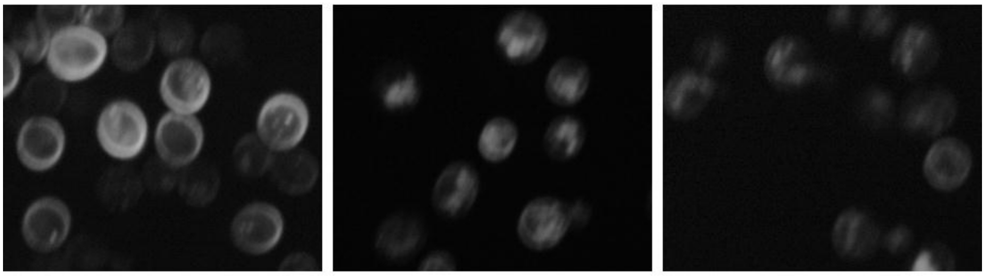
	CDC15 YAR019C
	CDC16 YKL022C
	CDC20 YGL116W
	CDC24 YAL041W
	CRZ1 YNL027W
	DCP1 YOL149W DTT only

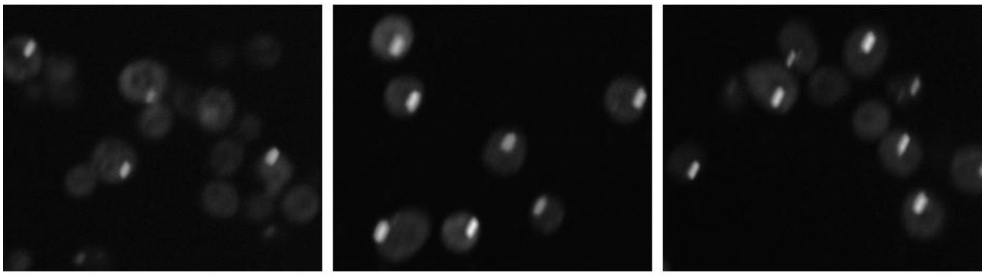
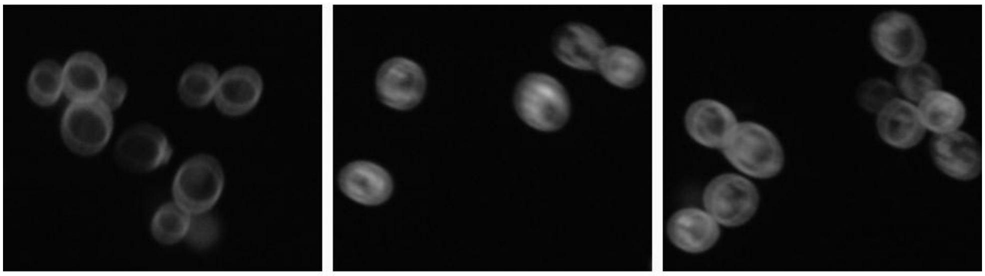
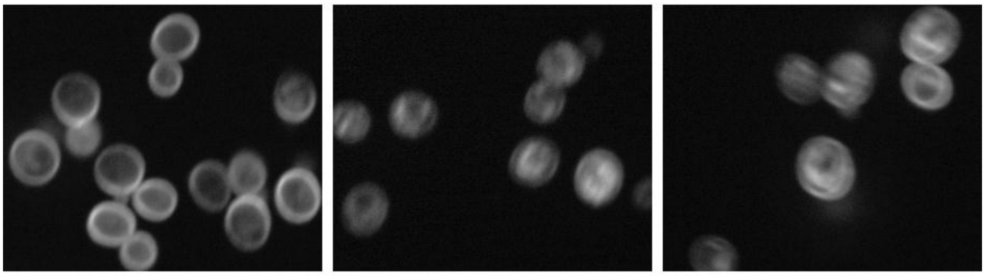
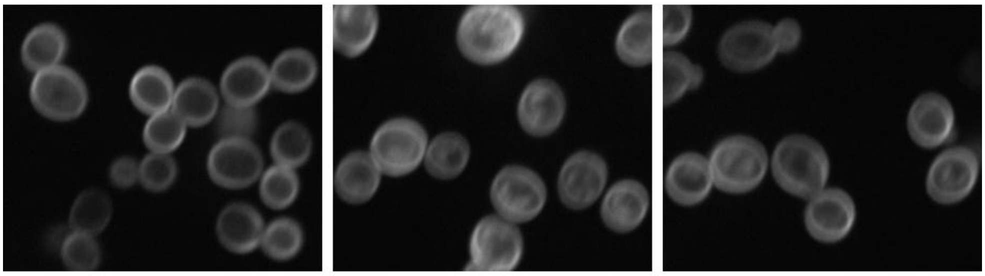
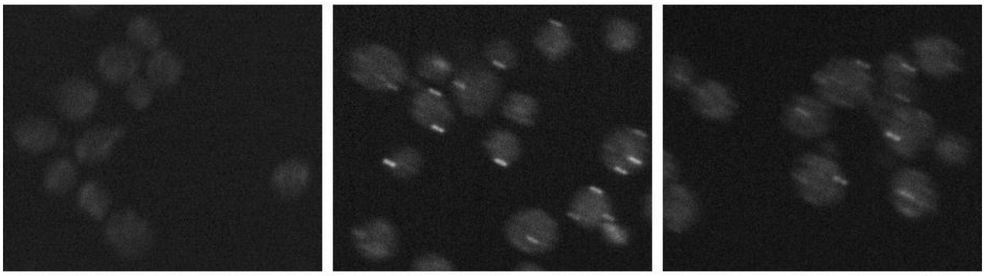
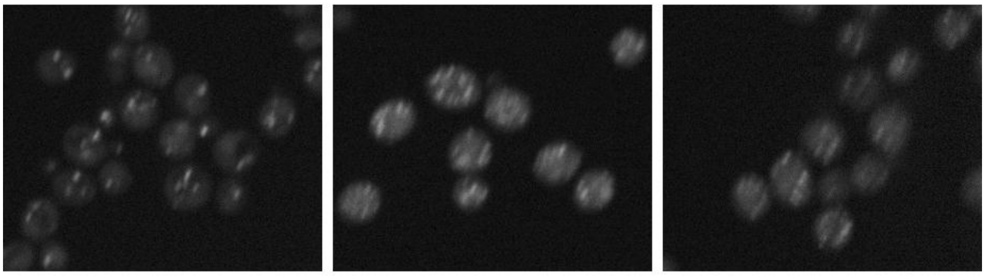
	DCP2 YNL118C
	DID4 YKL002W
	DIP5 YPL265W
	DRS1 YLL008W
	DUS3 YLR401C
	DYS1 YHR068W

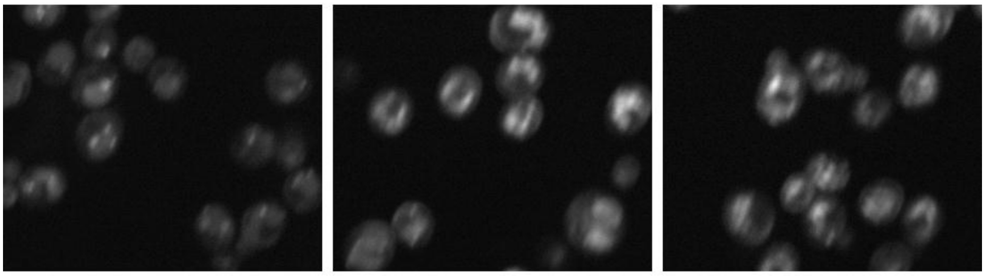
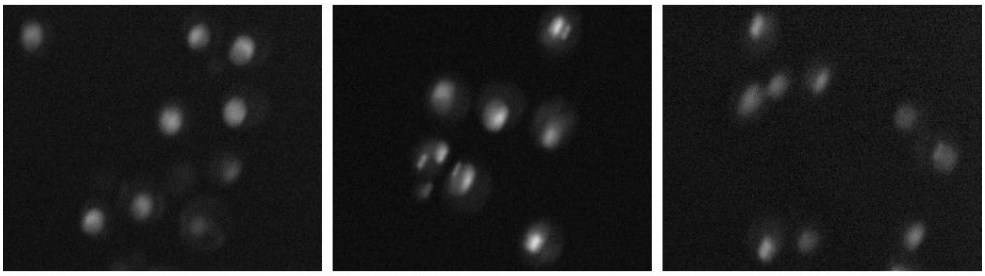
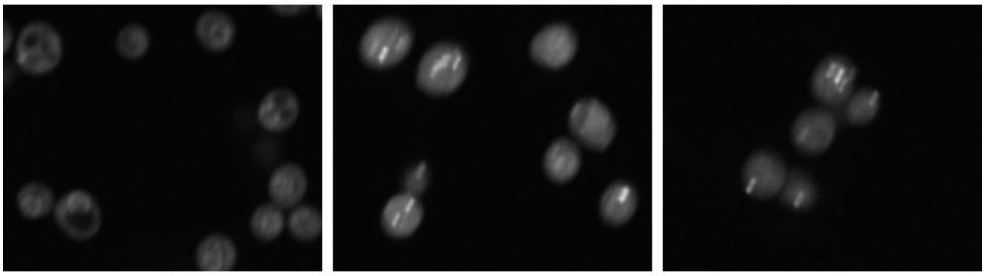
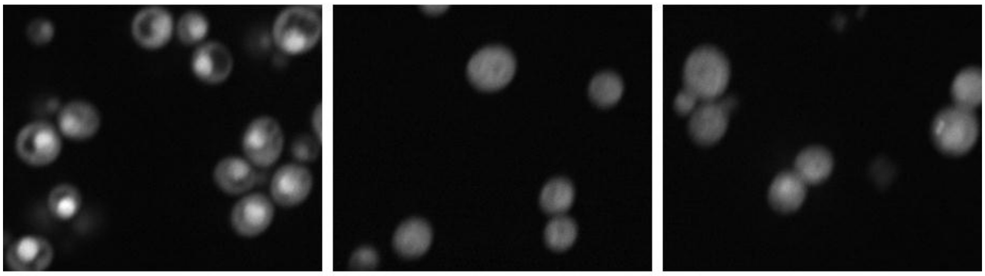
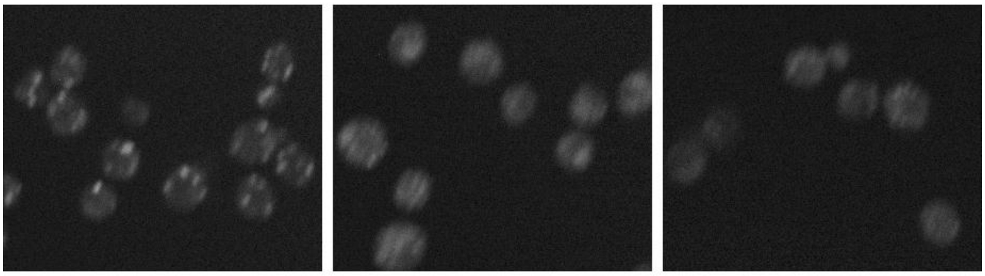
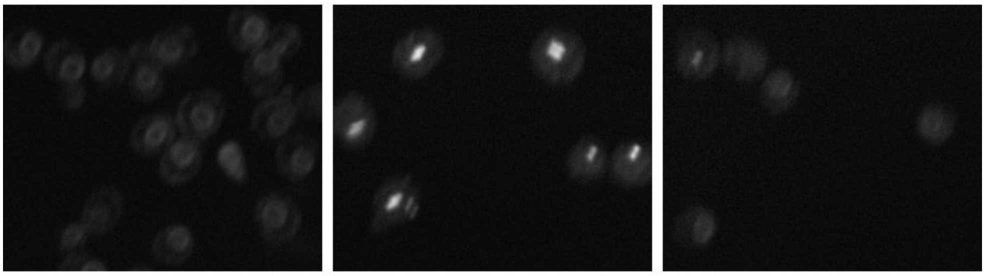
	EDC3 YEL015W
	EDE1 YBL047C
	EMI2 YDR516C DTT only
	ERG1 YGR175C
	ERG6 YML008C
	FAA1 YOR317W

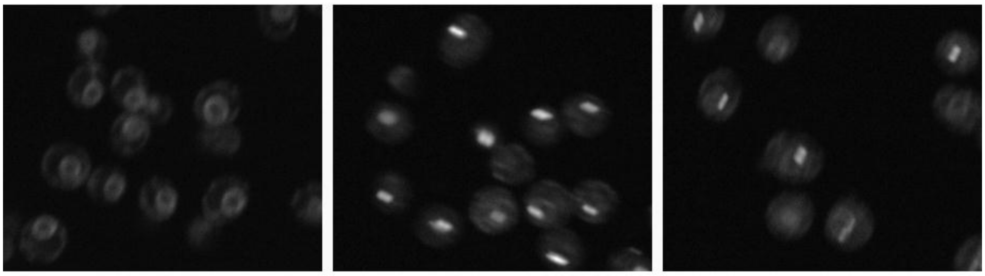
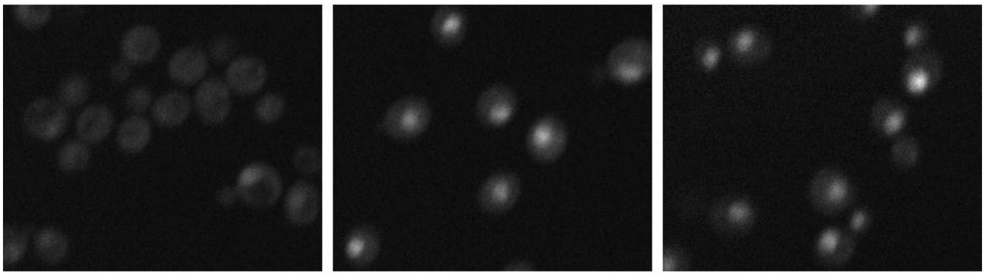
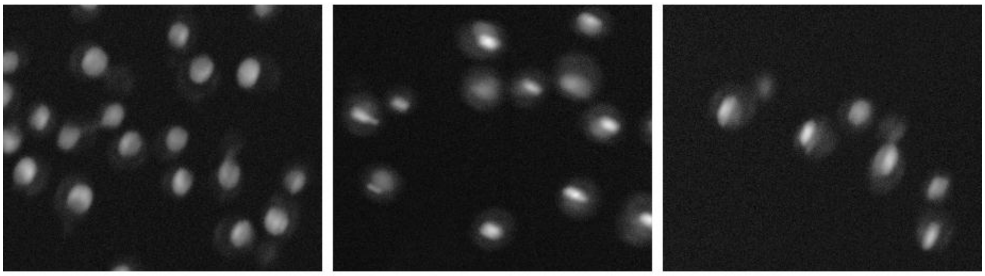
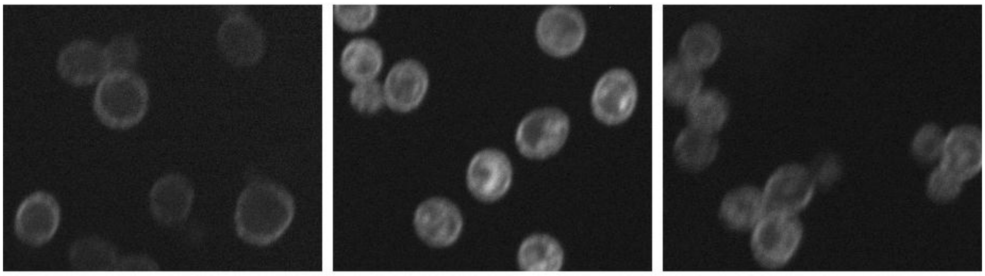
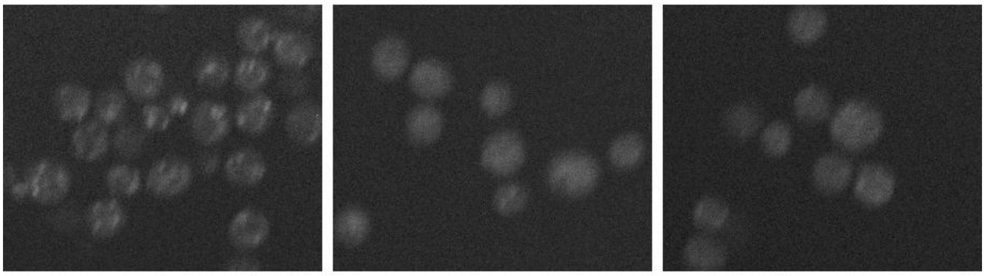
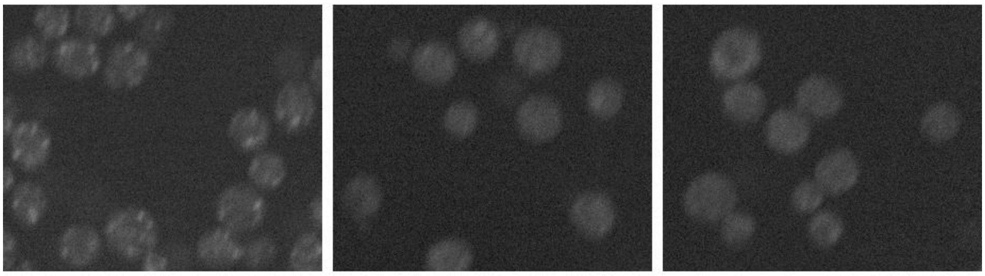
	FAA4 YMR246W
	FCY2 YER056C
	FLC1 YPL221W
	FPR3 YML074C
	FPR4 YLR449W
	FUI1 YBL042C

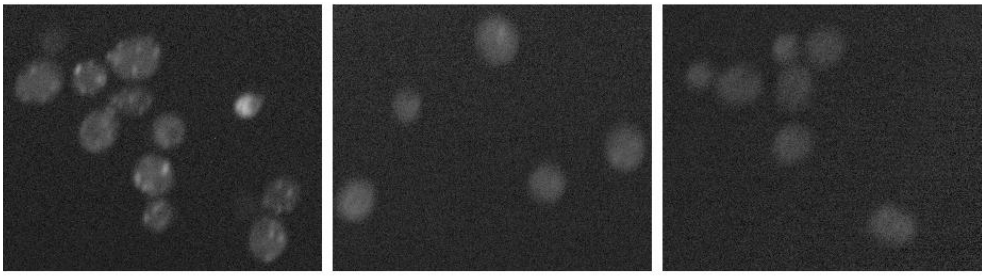
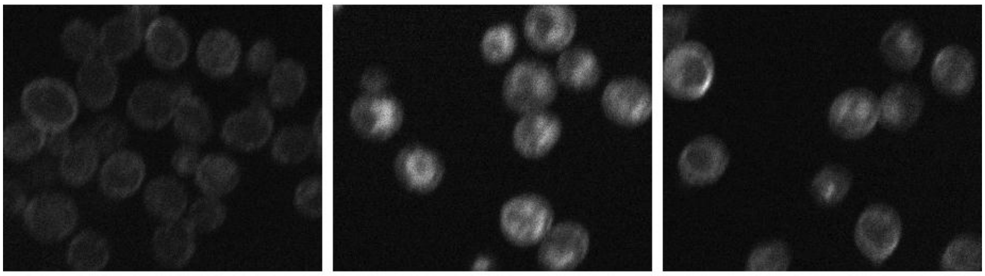
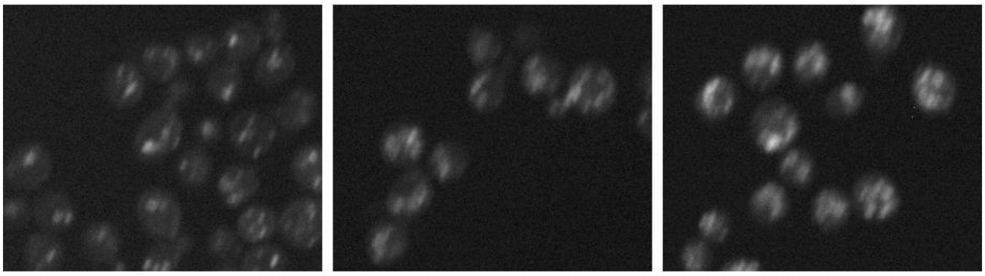
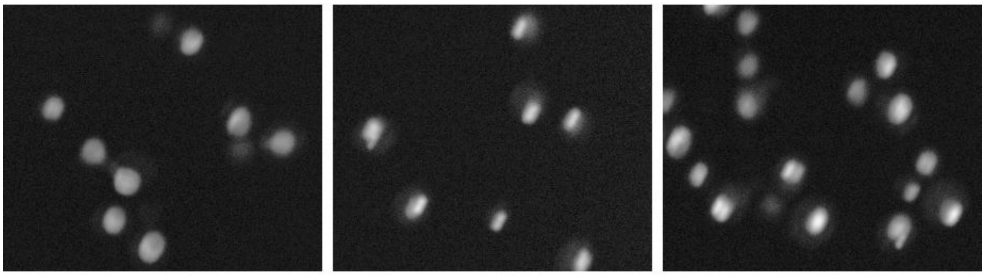
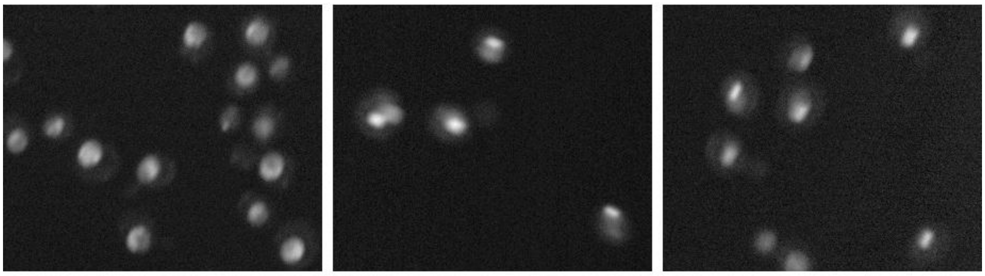
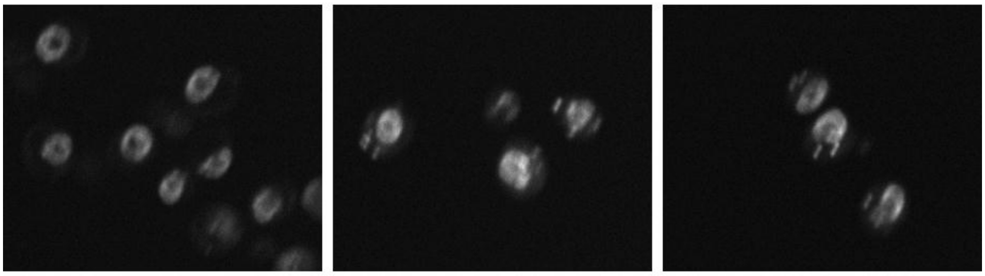
	GCN4 YEL009C
	GGA1 YDR358W DTT only
	GLC7 YER133W
	GLK1 YCL040W
	GLT1 YDL171C
	GSY1 YFR015C DTT only

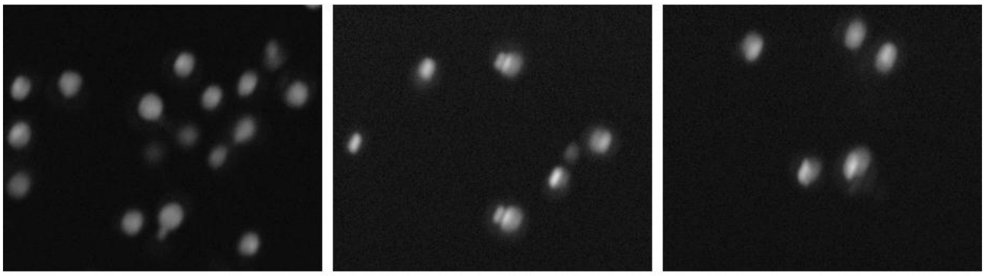
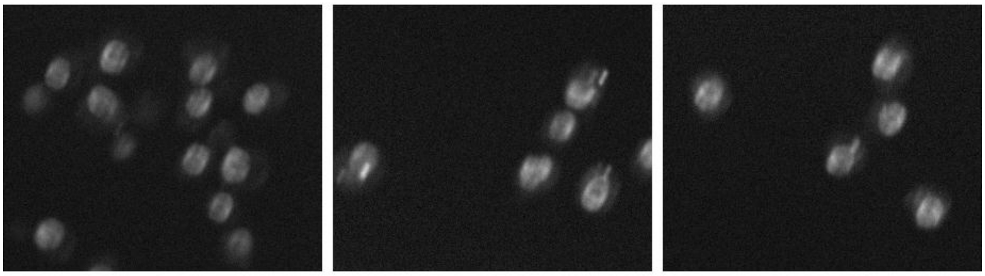
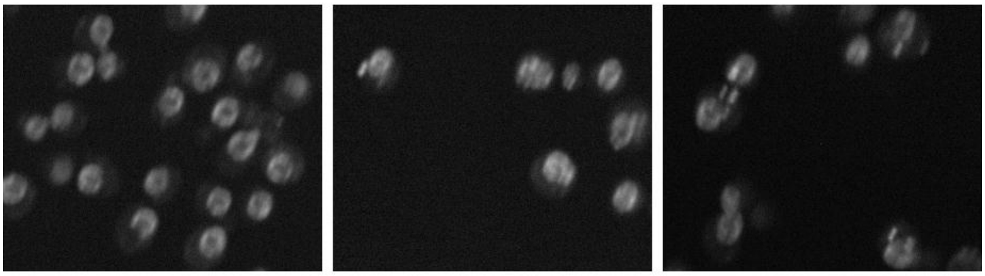
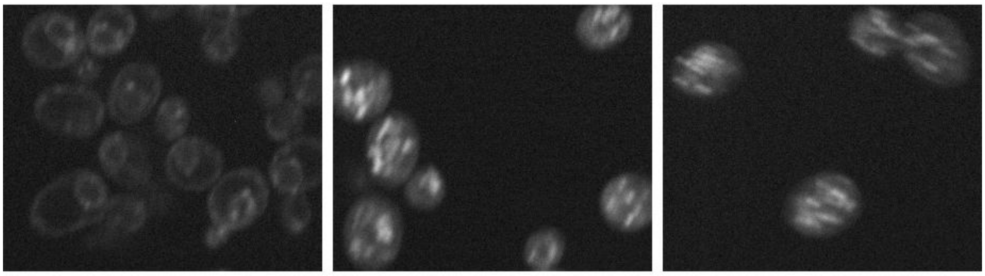
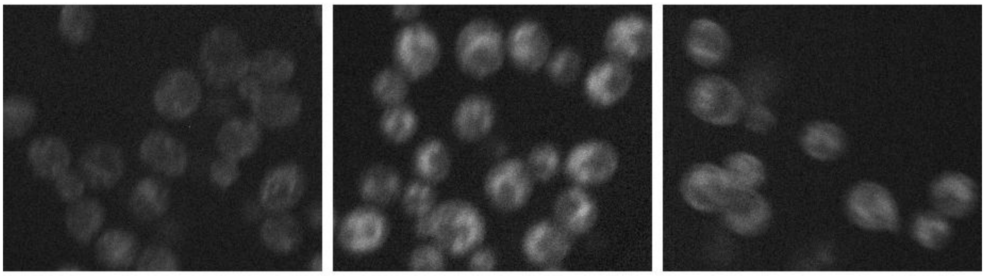
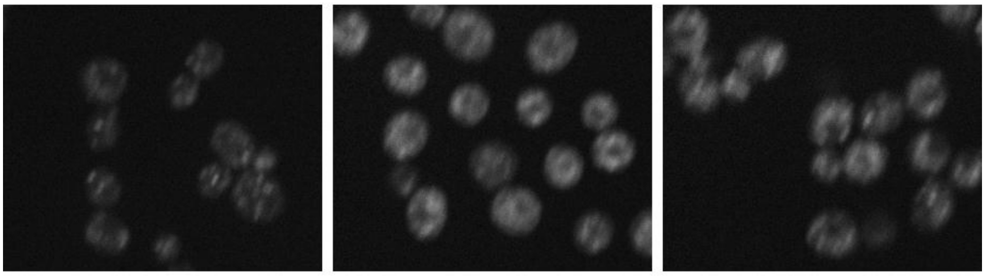
	HMG1 YML075C
	HMG2 YLR450W
	HNM1 YGL077C
	HSE1 YHL002W
	HSP26 YBR072W
	HSP30 YCR021C

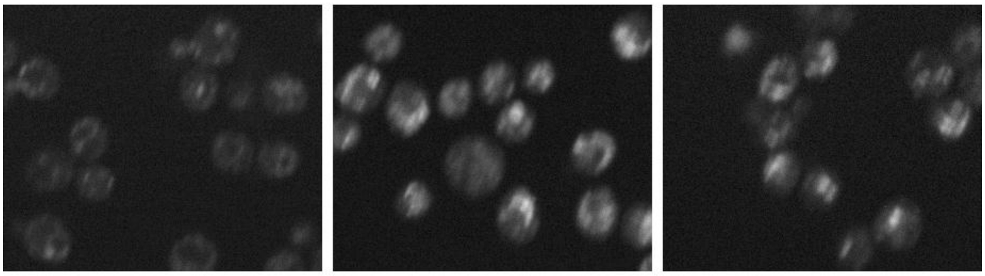
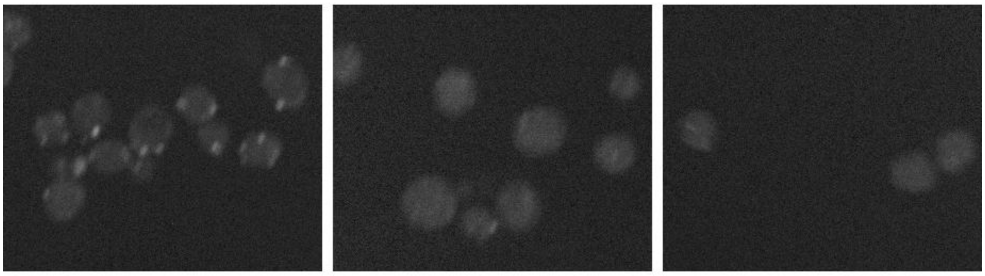
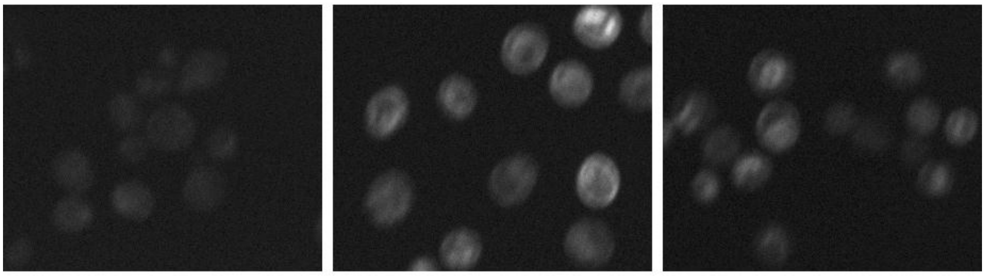
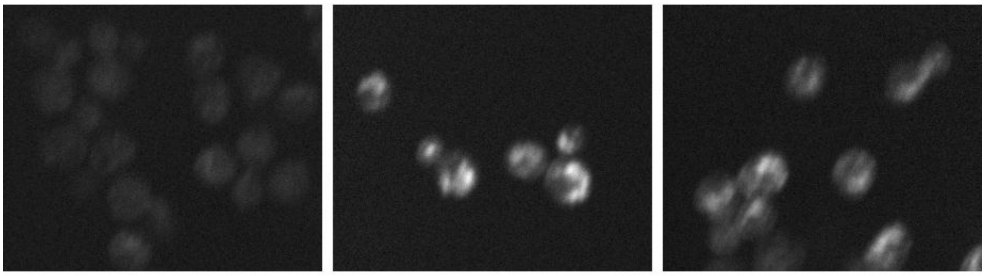
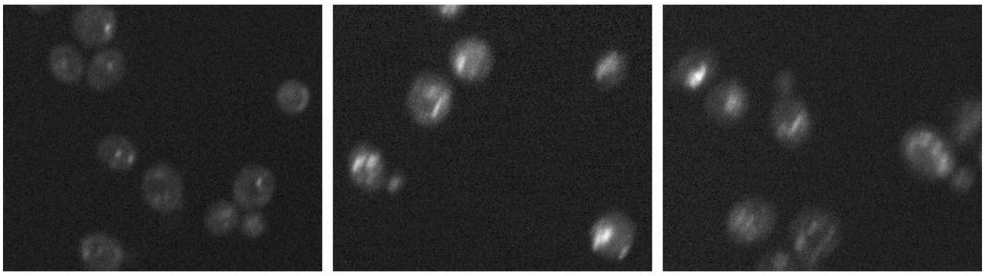
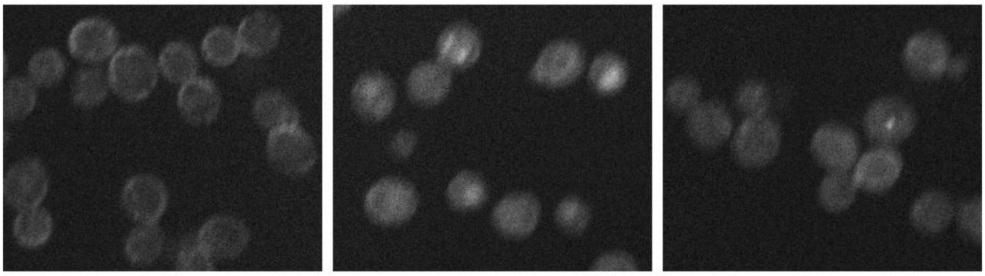
	HSP42 YDR171W
	HXT2 YMR011W
	HXT3 YDR345C
	HXT6 YDR343C
	IRE1 YHR079C
	IST1 YNL265C

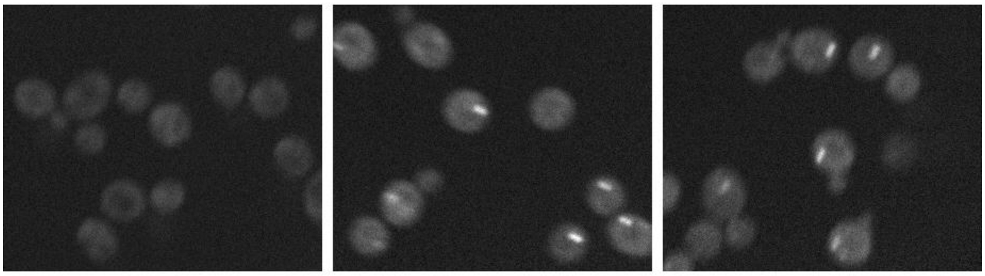
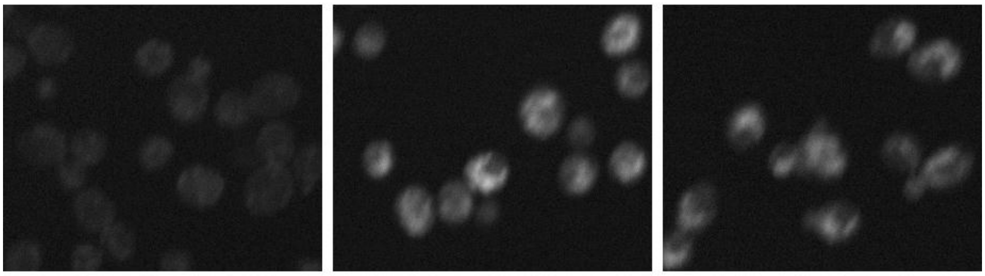
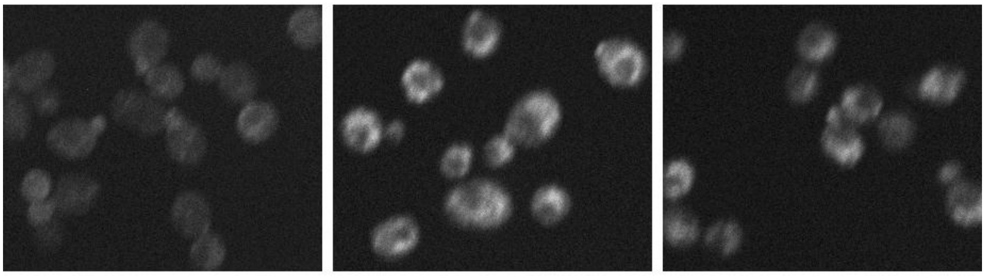
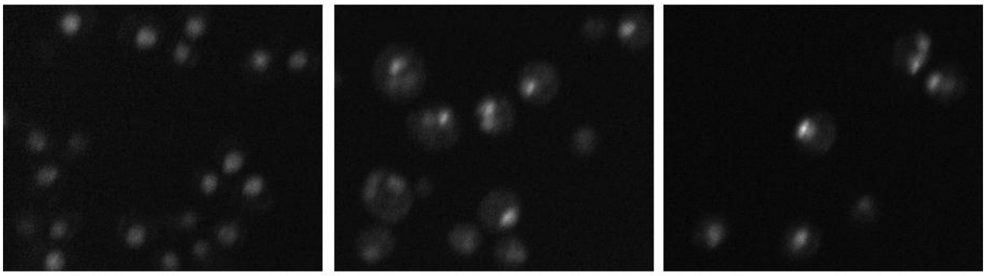
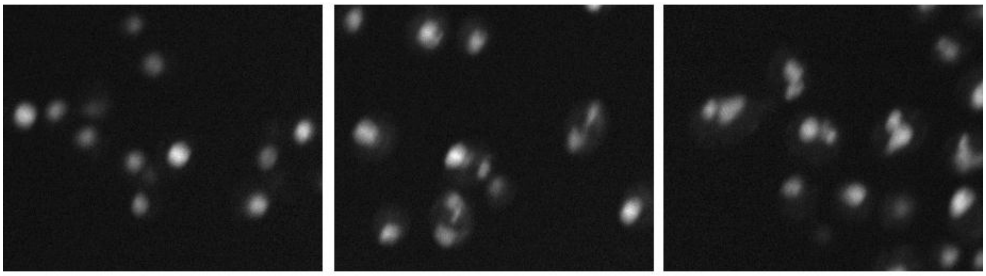
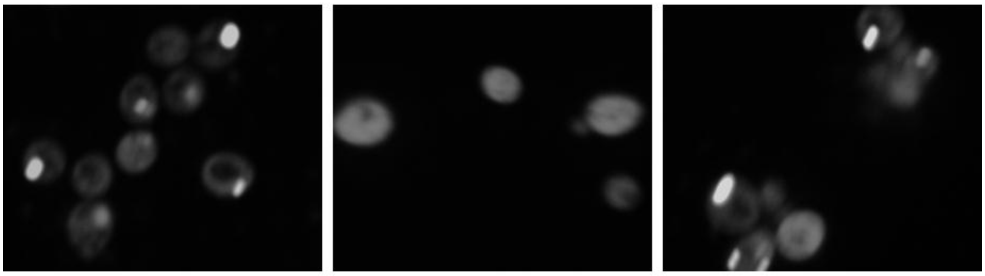
	ITR1 YDR497C
	JIP5 YPR169W
	KAP95 YLR347C
	KAP123 YER110C
	KEX2 YNL238W
	LAC1 YKL008C

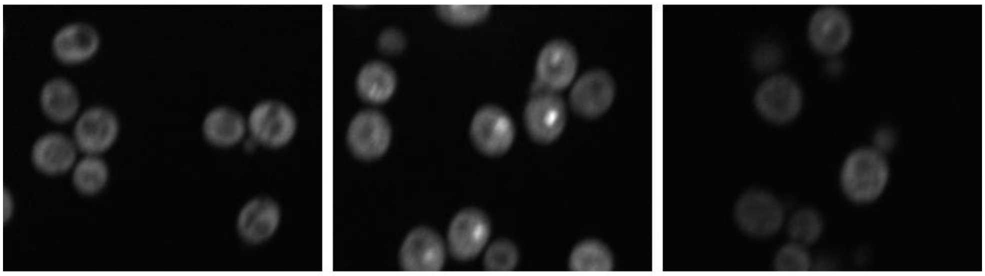
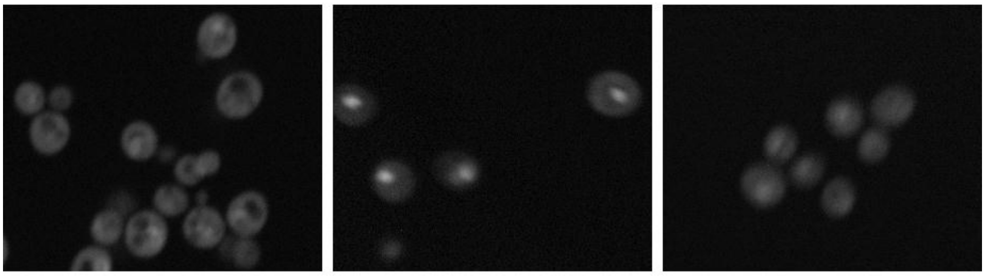
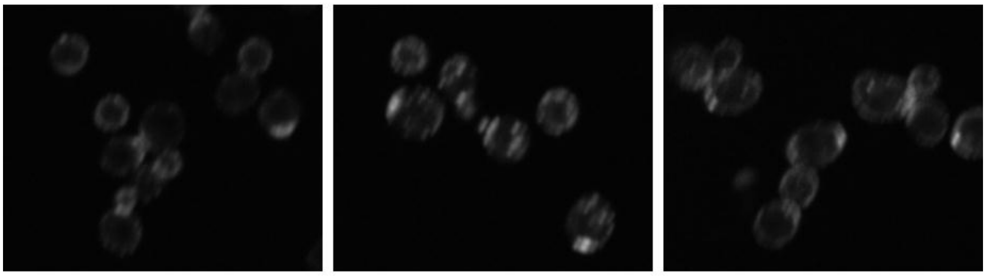
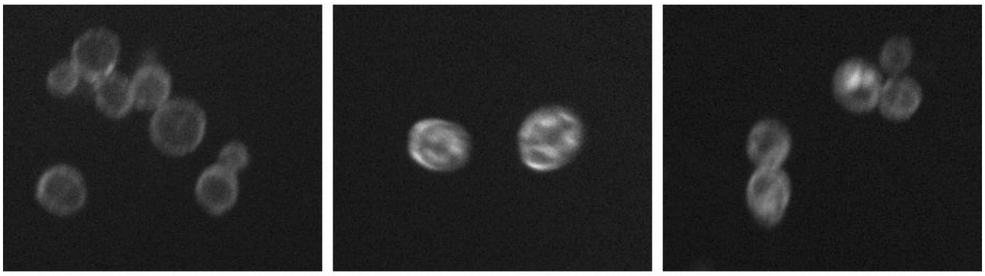
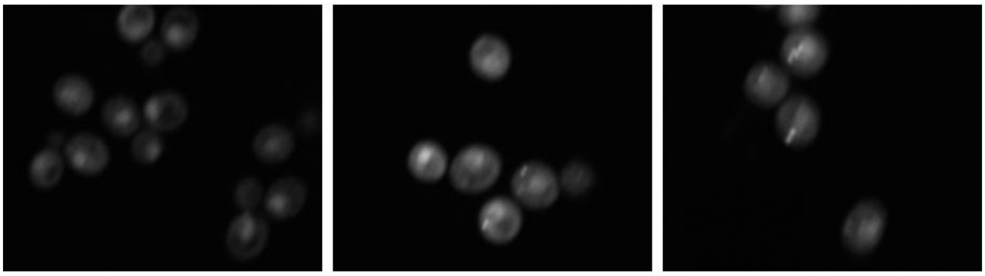
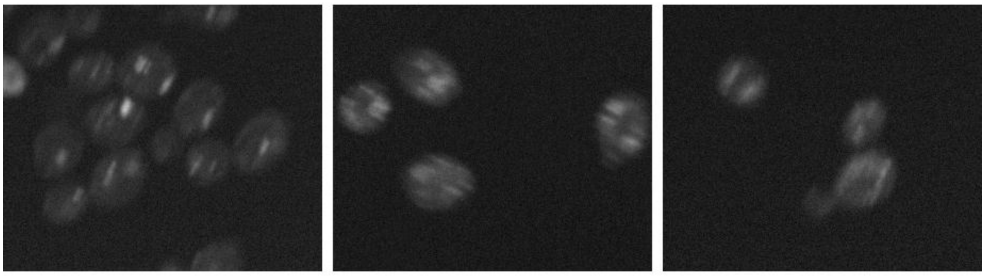
	LAG1 YHL003C
	LCB5 YLR260W
	MAK11 YKL021C
	MEP1 YGR121C
	MNN10 YDR245W
	MNN11 YJL183W

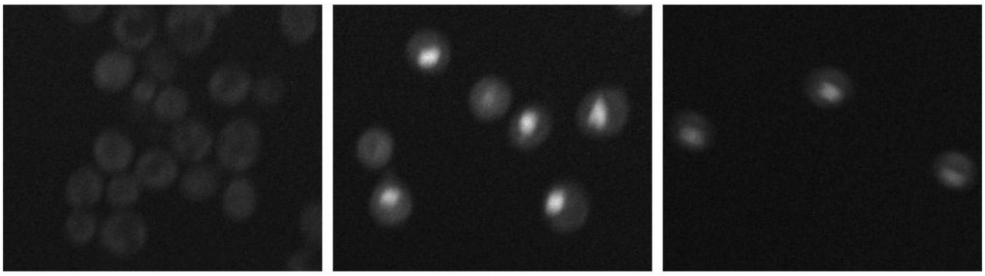
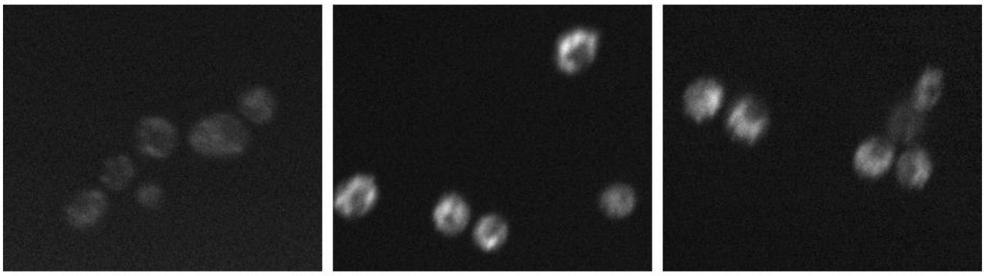
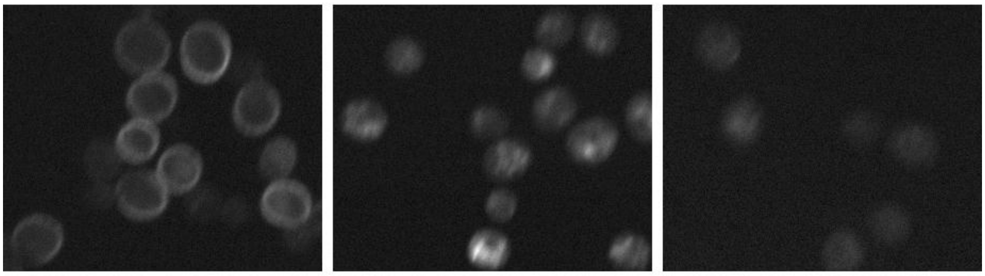
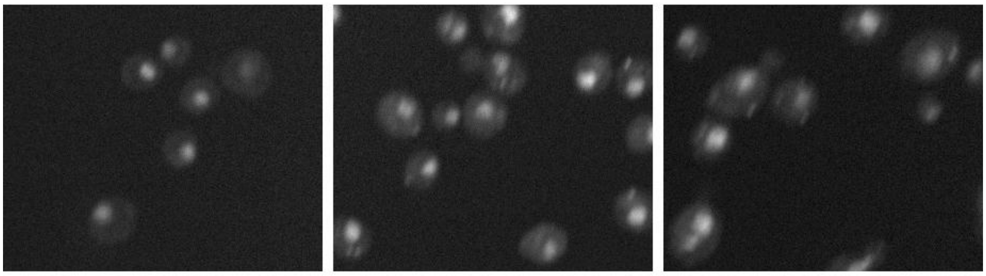
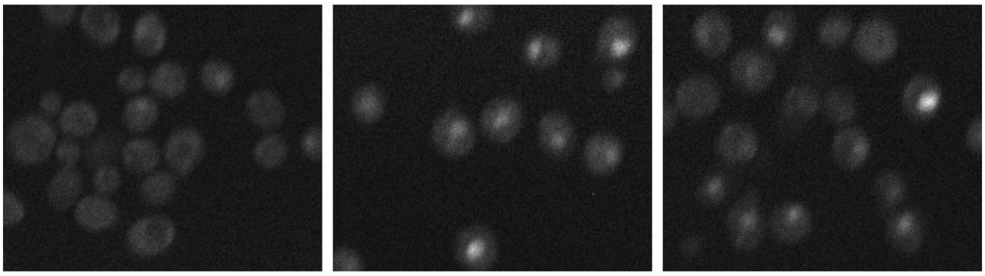
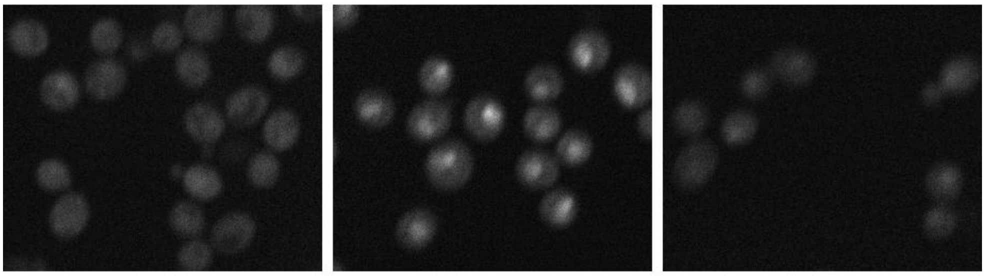
	MNN5 YJL186W
	MRPL6 YHR147C
	MVP1 YMR004W
	NOP16 YER002W
	NSA1 YGL111W
	NSP1 YJL041W

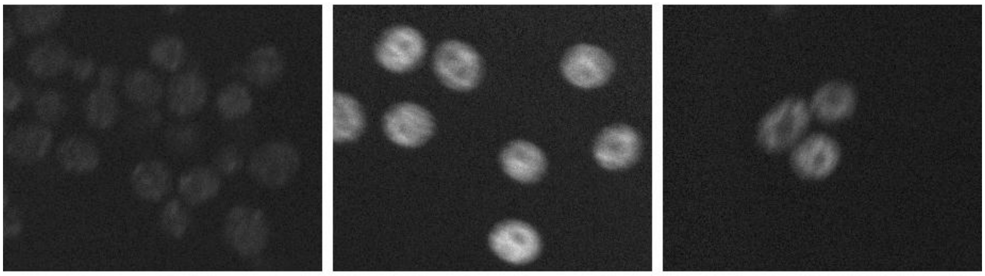
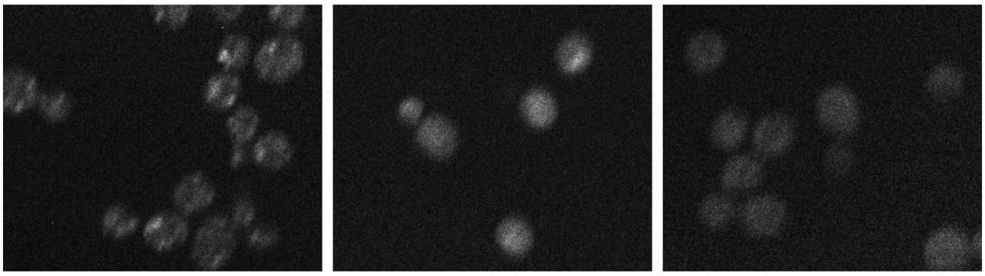
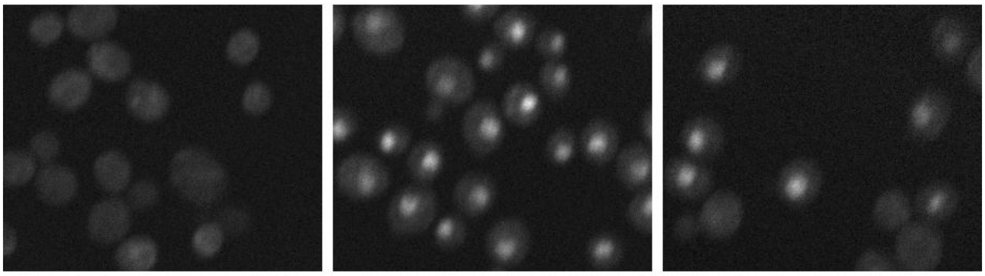
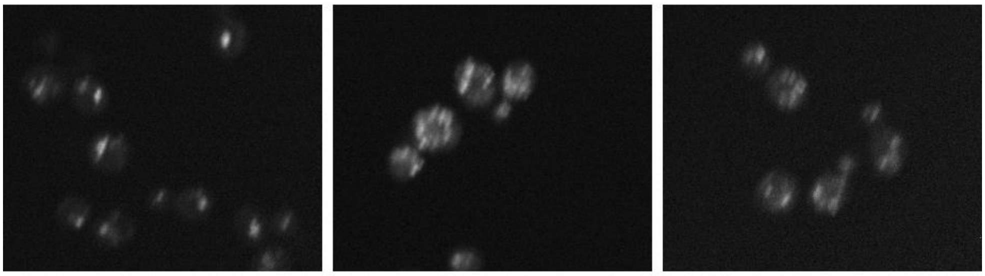
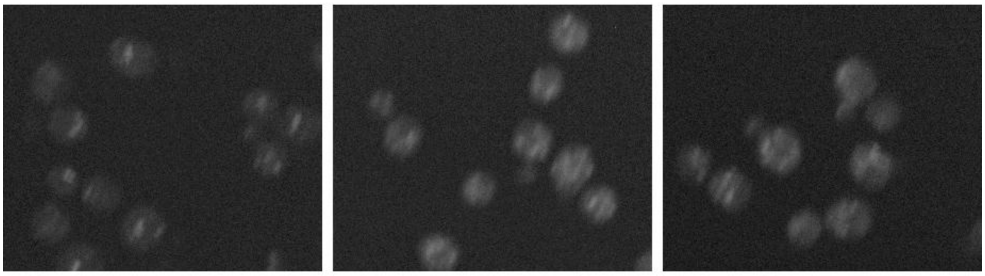
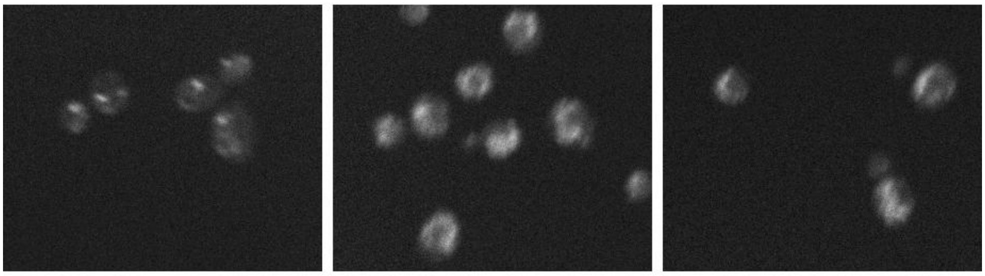
	NUG1 YER006W
	NUP82 YJL061W
	NUP159 YIL115C
	NUS1 YDL193W
	OPY2 YPR075C
	PEP1 YBL017C

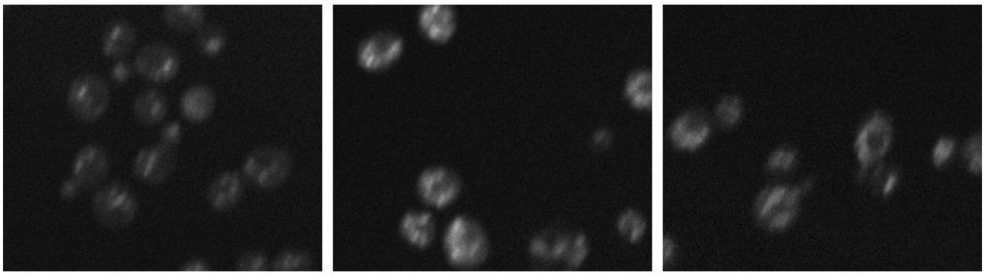
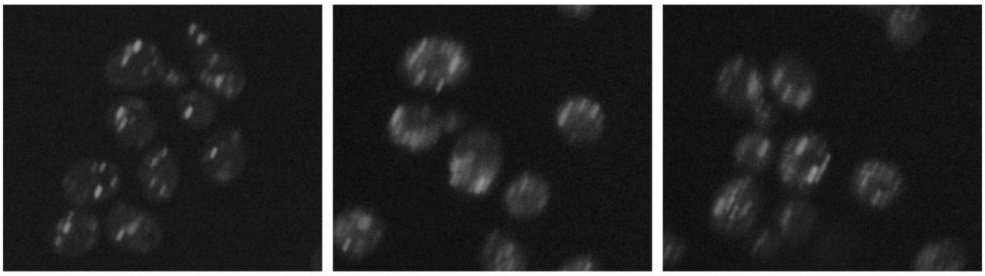
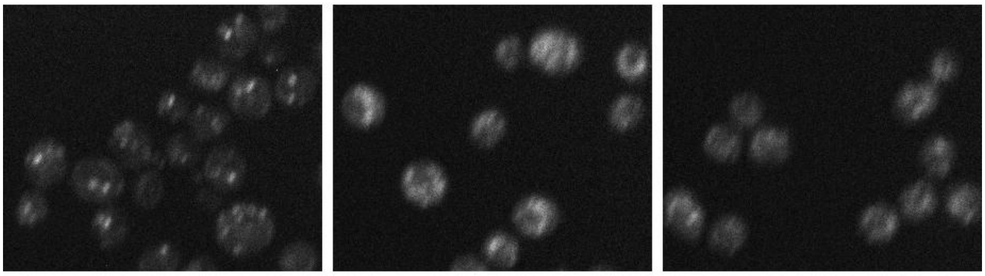
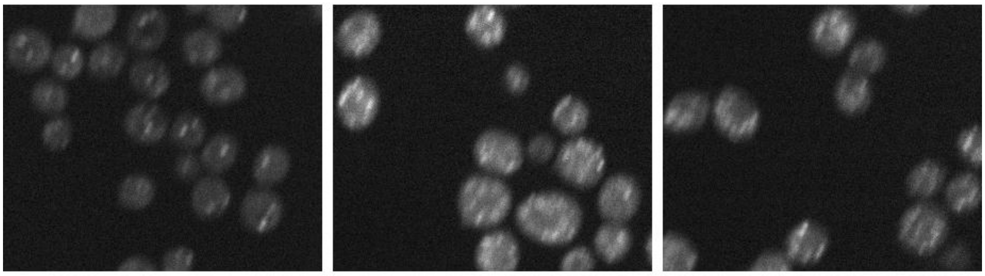
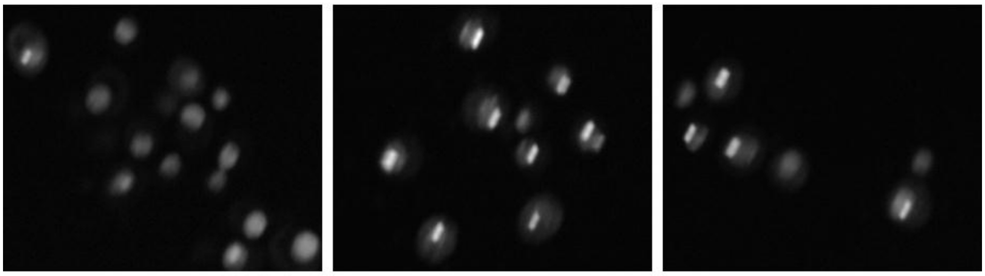
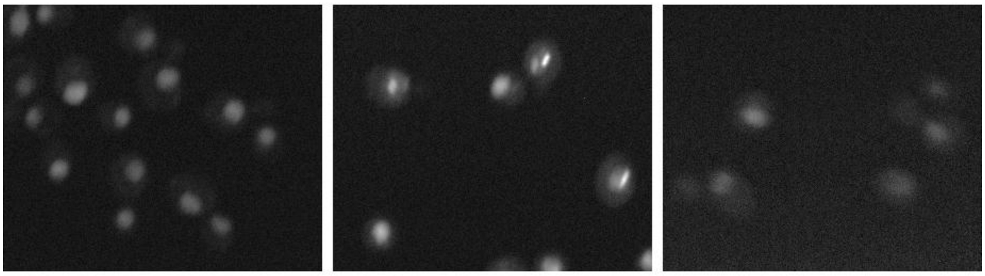
	PEP8 YJL053W
	PEX21 YGR239C
	PHM7 YOL084W
	PRM5 YIL117C
	PSO2 YMR137C
	QDR3 YBR043C

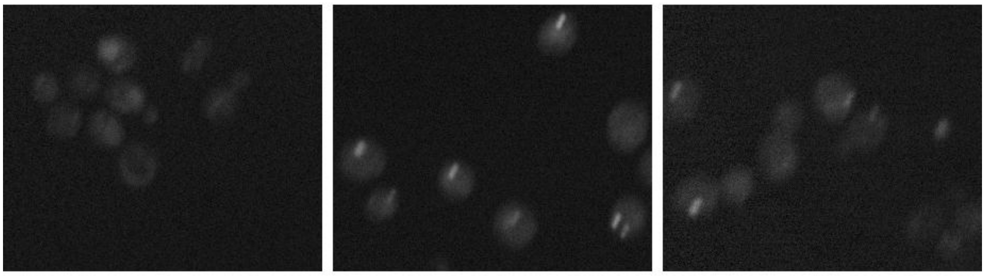
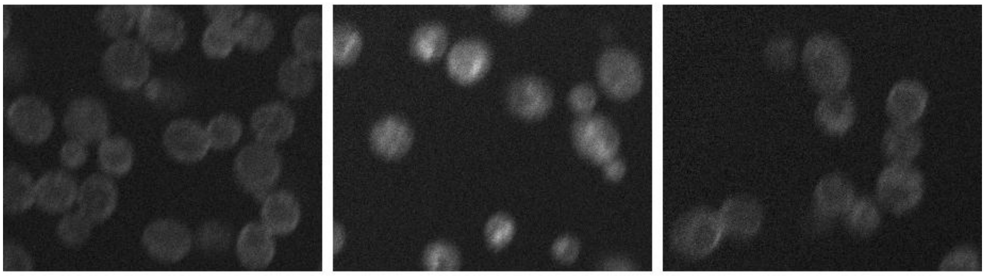
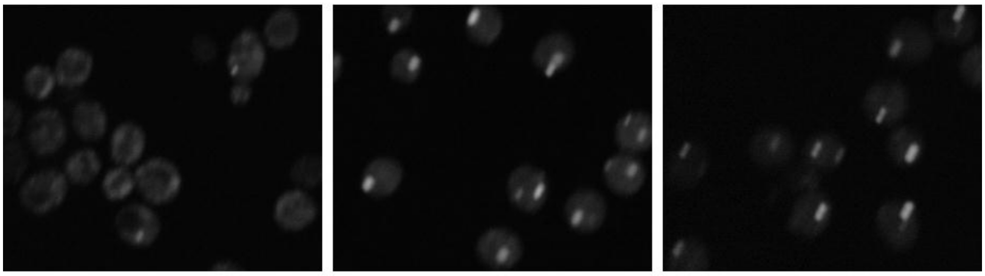
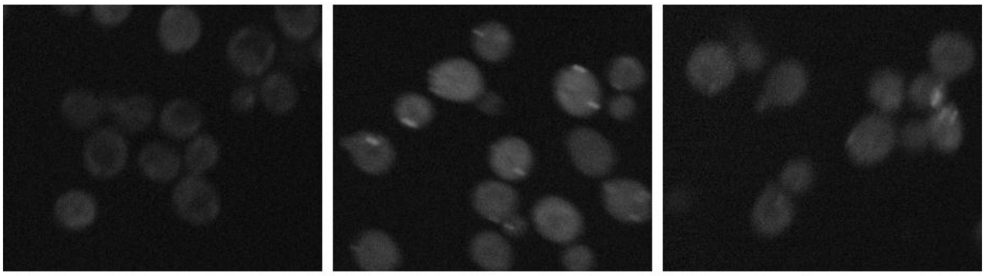
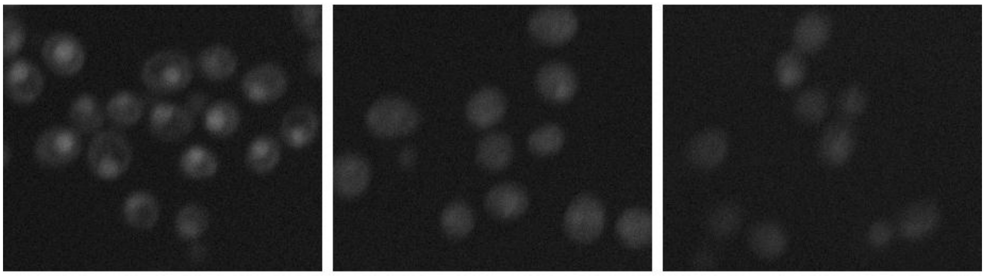
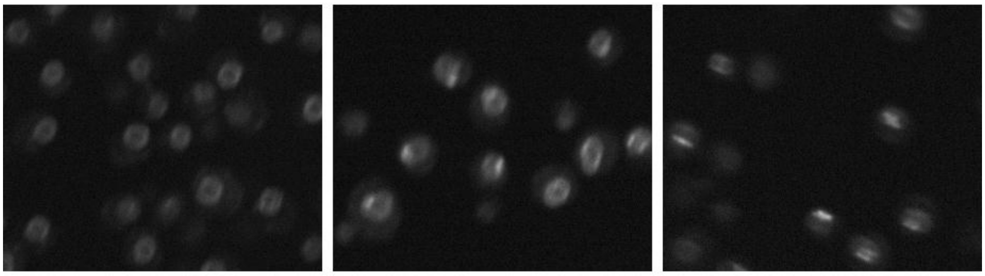
	RBS1 YDL189W
	RCR1 YBR005W
	RCR2 YDR003W
	RDH54 YBR073W
	RIX1 YHR197W
	RNR1 YER070W DTT only

	RPL18B YNL301C DTT only
	RRP12 YPL012W
	RSE1
	RSN1 YMR266W
	SIS1 YNL007C
	SNF7 YLR025W

	STB3 YDR169C
	SUR1 YPL057C
	THI7 YLR237W
	THO2 YNL139C
	TOD6 YBL054W
	TSR3 YOR006C

	UIP3 YAR027W
	VAN1 YML115C
	VHR1 YIL056W
	VPS4 YPR173C
	VPS8 YAL002W
	VPS16 YPL045W

	VPS17 YOR132W
	VPS24 YKL041W
	VPS38 YLR360W
	VTA1 YLR181C
	WTM1 YOR230W
	WTM2 YOR229W

	YAR009C
	YDR090C DTT only
	YDR170W-A
	YIL108W
	YOR342C
	YPR174C

7.9 Systematically removed GFP strains

ORF	Reason for exclusion
YAL014C	Tail-anchored protein
YBR016W	Tail-anchored protein
YBR067C	GPI-anchored protein
YBR093C	cell wall
YBR162W-A	Tail-anchored protein
YBR222C	PTS1
YCL043C	HDEL protein
YCR005C	PTS1
YCR067C	HDEL protein
YDL012C	Tail-anchored protein
YDL065C	fatty acid acylation
YDL078C	PTS1
YDL195W	GPI-anchored protein
YDR055W	GPI-anchored protein
YDR077W	GPI-anchored protein
YDR086C	Tail-anchored protein
YDR144C	GPI-anchored protein
YDR200C	Tail-anchored protein
YDR234W	PTS1
YDR261C	GPI-anchored protein
YDR281C	Tail-anchored protein
YDR304C	HDEL protein
YDR461W	fatty acid acylation
YDR498C	Tail-anchored protein
YDR517W	cell wall
YEL040W	GPI-anchored protein
YER019C-A	Tail-anchored protein
YER150W	GPI-anchored protein
YFL046W	Tail-anchored protein
YGL028C	cell wall
YGL032C	cell wall
YGL067W	PTS1
YGL184C	PTS1
YGR136W	GPI-anchored protein
YGR189C	GPI-anchored protein
YGR279C	cell wall
YGR282C	cell wall
YHR204W	GPI-anchored protein
YIL123W	cell wall
YIL160C	PTS1
YIR034C	PTS1
YJL034W	HDEL protein
YJL078C	GPI-anchored protein
YJL171C	cell wall
YJR086W	fatty acid acylation
YKL020C	Tail-anchored protein
YKL046C	GPI-anchored protein
YKL073W	HDEL protein
YKL096W-A	GPI-anchored protein
YKL164C	cell wall
YKL165C	cell wall
YKL175W	GPI-anchored protein
YKR042W	cell wall
YLR042C	GPI-anchored protein
YLR093C	Tail-anchored protein
YLR110C	GPI-anchored protein
YLR120C	GPI-anchored protein
YLR268W	Tail-anchored protein
YLR300W	cell wall
YLR343W	GPI-anchored protein
YLR390W-A	GPI-anchored protein
YMR161W	Tail-anchored protein
YMR215W	GPI-anchored protein
YMR238W	GPI-anchored protein
YMR251W-A	GPI-anchored protein
YMR307W	GPI-anchored protein
YNL064C	fatty acid acylation
YNL070W	Tail-anchored protein
YNL111C	Tail-anchored protein
YNL111C	GPI-anchored protein
YNL154C	fatty acid acylation
YNL300W	GPI-anchored protein
YNL322C	cell wall
YNL327W	GPI-anchored protein
YNR044W	GPI-anchored protein
YNR056C	GPI-anchored protein
YNR067C	GPI-anchored protein
YOL011W	GPI-anchored protein
YOL030W	cell wall
YOL044W	Tail-anchored protein
YOL052C-A	GPI-anchored protein
YOL088C	HDEL protein
YOL154W	GPI-anchored protein
YOR045W	Tail-anchored protein
YOR084W	PTS1
YOR101W	fatty acid acylation
YOR214C	GPI-anchored protein
YOR288C	HDEL protein
YOR324C	Tail-anchored protein
YOR327C	Tail-anchored protein
YOR336W	HDEL protein
YPL192C	Tail-anchored protein
YPL206C	Tail-anchored protein