



HAL
open science

Characterisation of transcriptional properties of the $hid1\Delta$ and $hid3\Delta$ mutants of *Schizosaccharomyces pombe*

Mohammed Alshehri

► **To cite this version:**

Mohammed Alshehri. Characterisation of transcriptional properties of the $hid1\Delta$ and $hid3\Delta$ mutants of *Schizosaccharomyces pombe*. *Genetics*. Université de Bordeaux, 2015. English. NNT : 2015BORD0141 . tel-01617019

HAL Id: tel-01617019

<https://theses.hal.science/tel-01617019>

Submitted on 16 Oct 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Année 2015

université
de BORDEAUX



Thèse n°2015BORD0141

THÈSE

pour le

DOCTORAT DE L'UNIVERSITÉ de BORDEAUX

Ecole Doctorale des Sciences de la Vie et de la Santé

Spécialité: Génétique

Présentée et soutenue publiquement

Le 15 Septembre 2015

Par Mohammed Alshehri

Characterisation of transcriptional properties of the *hid1Δ* and *hid3Δ* mutants of *Schizosaccharomyces pombe*

*RNAseq reveals precise transcriptional changes affecting stress
proteins and those transported through the endomembrane secretory
system*

Membres du jury:

M. Moreau, Patrick
M. Gachet, Yannick
M. McFarlane, Ramsay
M. El Ghouzzi, Vincent
M. Hooks, Mark

Univeristé de Bordeaux
Université Paul Sabatier
Bangor University
INSERM, Hôpital Robert Debré
Univeristé de Bordeaux

Président
rapporteur
rapporteur
Examineur
Directeur de thèse

Characerisation des propriétés transcriptionnelles des mutants *hid1Δ* et *hid3Δ* chez *S. pombe*

RÉCAPITULATIF EN FRANÇAIS

Schizosaccharomyces pombe devient de plus en plus un système modèle pour étudier la régulation de l'expression des gènes et des protéines dans les processus impliqués dans le développement de cancers et de maladies génétiques. Ces travaux peuvent servir à étudier les propriétés putatives de la protéine humaine HID1 à empêcher des tumeurs de se former. J'ai utilisé la technique RNAseq pour révéler les changements d'expression des gènes sur les cellules de *S. pombe* dont sont absentes trois gènes orthologues du gène humain *HID1*: *hid1+*, *hid2+* et *hid3+*. Des mutants ont été créés par remplacement de gènes et testés pour découvrir leurs propriétés de croissance. La croissance du mutant *hid2Δ* semblait meilleure tandis que celle de *hid3Δ* semblait plus lente que les contrôles. La morphologie cellulaire de chaque mutant était normale. La microscopie à transmission électronique a révélé que l'appareil de Golgi était fortement modifié dans *hid3Δ*. RNAseq a montré que plus de 500 gènes étaient exprimés différemment dans *hid3Δ*. Les changements d'expression indiquaient des cellules sous tension. Par ailleurs, un jeu défini de facteurs de transcription et un groupe de gènes encodant des protéines situées et sécrétées ont été introduits, ce qui suggère que la perturbation de la fonction protéique au niveau de la membrane plasmique a un effet feedback sur la régulation de l'expression des gènes. J'émetts l'hypothèse que la croissance lente de *hid3Δ* s'explique par un état cellulaire de quiescence partielle. Aussi, *S. pombe* ont été séquencées et révèlent l'expression de petits ARN non codants spécifiques, dont des introns complets et d'autres ARN non codants non annotés.

MOTS CLÉS: Formation Dauer, Transcriptomique, Quiescence cellulaire

ENGLISH SUMMARY

Schizosacchormyces pombe has become increasingly a model system to study the regulation of gene expression, stress signalling and metabolic and protein changes for processes implicated in the development of cancer and genetic diseases. This work was started to determine if *S. pombe* could be used to study the putative anti-tumour forming properties of the human HID1 protein. I employed the NGS technology RNAseq to reveal gene expression changes to the cells of *S. pombe* lacking three orthologues of the human *HID1* gene, *hid1*⁺, *hid2*⁺ and *hid3*⁺. Mutants lacking these genes were created by gene replacement and tested for growth properties. The mutant *hid2Δ* appeared to grow better and *hid3Δ* grew more slowly than WT controls. The cell morphology of each mutant was normal and lengths and widths were unchanged. Transmission electron microscopy revealed that the Golgi apparatus was greatly modified in *hid3Δ* but not in other genotypes. RNAseq showed that under standard growth conditions more than 500 genes were differentially expressed in *hid3Δ*. Expression changes were indicative of cells under stress. Also, a defined set of transcription factors and a group of genes encoding proteins located in the plasma membrane and secreted were induced, suggesting that disruption of protein function at the plasma membrane feeds back to regulate gene expression. I hypothesize that slow growth of *hid3Δ* is due to a partial quiescent cell state. To start investigating mechanisms regulating gene expression, small RNA libraries of the size of *S. pombe* introns were sequenced and revealed the expression of specific small non-coding RNAs, including full introns and other un-annotated non-coding RNAs.

KEY WORDS: Dauer formtion, transcriptomics; cellular quiescence

Table of Contents

Récapitulatif en Français & Mots Clés	II
English Summary & Key Words	III
Table of Contents	IV
List of Tables and Figures	VII
Acknowledgement	X
List of Abbreviations	XI
Récapitulatif du projet en Français	XIII
I. Chapter 1: General Introduction	1
1.1. Nucleic acids and general transcription	1
1.2 Gene expression and its regulation	2
1.2.1 The transcriptional machinery	2
1.2.2 Chromatin structure	3
1.3. The basic biology of cancer	3
1.3.1 Genes involved in carcinogenesis: Oncogenes and tumour-suppressor genes	4
1.4 Gene expression in cancer	5
1.4.1 Regulation of transcriptional by Ubiquitin	6
1.4.2 Ubiquitination and Transcription Factor activity	7
1.5 Introduction to the Golgi apparatus	8
1.5.1 Golgi apparatus in yeast	9
1.5.2 Signaling through the Sterol receptor element binding protein (SREBP) pathway	10
1.5.3 GOLPH3	11
1.5.4 GOLPH3 function and cancer	12
1.6. Human HID1 and its role in cancer development	13
1.7 Introduction to non-coding RNAs	14
1.7.1 Classes of ncRNA	14
1.7.1.1 Ribosomal and transfer RNAs	14
1.7.1.2 Regulatory ncRNAs	15
1.7.1.2.1 .LncRNAs	16
1.7.1.2.2 sncRNAs	17

1.7.1.2.2.1 miRNAs, siRNAs and piRNAs	17
1.7.1.2.3 snoRNAs	19
1.7.2 General features of ncRNA function	19
1.7.3 ncRNAs and cancer	20
1.7.4 Modeling small RNA structures and determining expression	21
1.8 RNAseq as a gene expression tool	22
1.9 <i>S. pombe</i> as a model organism	24
1.9.1 The cell cycle in <i>S. pombe</i>	25
1.10 Objectives of the work and strategy of the thesis	26
II. Chapter 2: Materials and Methods	28
2.1 Bioinformatic analysis and expression of <i>Psc3</i> intronic sequences	28
2.1.1 Prediction of structural features within <i>Psc3</i> introns	28
2.1.1.1 Identification of orthologous proteins in fungi	28
2.1.1.2 Identification of orthologous introns in fungal cohesins	28
2.1.1.3 Prediction of structure within introns of <i>Psc3</i>	29
2.1.2 Detection of ncRNA transcripts from <i>Psc3</i> intronic sequences by PCR	29
2.1.2.1 Growth of <i>S. pombe</i>	29
2.1.2.2. Isolation of total RNA	29
2.1.2.3 Quality and quantity of RNA	30
2.1.2.4 Reverse Transcription	30
2.1.2.5 Design of primers for PCR	31
2.1.2.6 Polymerase Chain Reaction	31
2.1.2.7 Visualisation of PCR products by UREA-PAGE	31
2.2 Generation of mutant strains of <i>S. pombe</i>	31
2.2.1 Biological Material	31
2.2.2 Isolation of plasmid DNA from <i>E. coli</i>	32
2.2.3 Design of primers for amplification of transformation templates	33
2.2.4 Amplification and preparation of templates for gene replacement by homologous recombination.	33
2.2.5 Transformation of <i>S. pombe</i> cells	34
2.2.6 Verification of gene replacements	34

2.2.6.1 Isolation of genomic DNA from <i>S. pombe</i>	34
2.2.6.2 Mutant genotyping and expression analysis by PCR	35
2.2.6.3 Transmission electron microscopy	35
2.3 Characterisation of gene expression	36
2.3.1 Preparation of material and isolation of mRNA	36
2.3.2 Calculation of gene correlations	36
2.3.3 Primer design for quantitative PCR of genes correlated to the Hids	36
2.3.4 Gene expression by RT-qPCR	37
2.3.4.1 Analysis of RT-qPCR data	37
2.3.5 RNAseq and differential expression analysis	37
2.3.6 Small size selected RNAseq and novel ncRNA transcript detection	38
2.3.7 snoRNA prediction from novel ncRNA transcripts	38
2.4 General Data Analysis.	39
III. Chapter 3: Production of <i>hidΔ</i> mutants of <i>S. pombe</i>	40
3.1 Introduction	40
3.2 Results	41
3.2.1 Creation of mutants by gene replacement	41
3.2.2 Verification of gene deletion and replacement by PCR-genotyping and RT-PCR:	41
3.2.3 Construction of single mutants	42
3.2.3.1 Confirmation of strains lacking <i>Hid1</i>	42
3.2.3.2 Confirmation of strains lacking <i>Hid2</i>	42
3.2.3.3. Confirmation of strains lacking <i>Hid3</i>	43
3.2.4 Identification of Vector Control strains	43
3.2.5 Construction of double mutants	45
3.2.6 Initial physiological studies of gene-replacement mutants	45
3.2.6.1 Morphology of single mutants	45
3.2.6.2 Characterisation of growth properties.	46
3.2.4.3 Preliminary ultrastructural characterisation of the mutants	47
3.3 Discussion	47
IV. Chapter 4: Transcriptional properties of the <i>hid1Δ</i> and <i>hid3Δ</i> mutants of <i>S. pombe</i>	50
4.1. Introduction	50

4.2 Results	52
4.2.1 Evaluation of <i>HsHID1</i> as a Tumour Suppressor gene	52
4.2.2 Comparative expression of <i>hid</i> genes using RT-qPCR	52
4.2.3 Evaluating the potential regulation of co-expressed genes by RT-PCR	54
4.2.4 Global analysis of gene expression changes in <i>hid1Δ</i> and <i>hid3Δ</i> by RNAseq	55
4.2.4.1 Strategy of sample selection and evaluation of RNA quality	55
4.2.4.2 RNAseq reveals greatest differences in gene expression for <i>hid3Δ</i>	55
4.2.4.3. Validation of RNAseq data by qRT-PCR	59
4.2.4.4. Analysis of correspondence with genetic interaction data provided in PomBase.	59
4.2.4.5. Finding biological function through gene ontologies	60
4.2.4.6. Changes in expression of specific genes	61
4.3 Discussion	63
V. Chapter 5: Analysis of structural features of introns and their expression in fission yeast based on the <i>pcs3</i> gene family	67
5.1. Introduction	67
5.2. Results	70
5.2.1. Identifying orthologous introns in fungal cohesins	70
5.2.2. Alignment of orthologous introns and secondary structure prediction	70
5.2.3. Investigation of intron expression by RT-PCR	71
5.2.4. Investigating the potential for intron expression by RNAseq	72
5.3 Discussion	75
VI. Chapter 6: General Discussion and Outlook	77
6.1 Outlook and Future Work	83
VII. REFERENCES	85

List of Tables and Figures

Chapter 1: General Introduction:

- Figure 1.1. Diagram of the Golgi apparatus and ERES in *P. pastoris* and *S. cerevisiae*.
Figure 1.2. The scheme of the SREBP signalling pathway in mammals
Figure 1.3. The Sre1 signalling pathway of fission yeast
Figure 1.4 DNA damage leading to Golgi dispersal.
Figure 1.5: The cell cycle of a eukaryotic cell.

Chapter 2: Materials and Methods

- Table 2.1. list of strains used in experimentation
Table 2.2. Primers for amplifying sequences in *psc3*.
Table 2.3 : Primers for amplifying sequences of ncRNA genes.
Table 2.4. PCR primers used to amplify transformation cassettes for gene replacement.
Table 2.5. Primer pairs for verification of *hid*⁺ replacement with marker gene.
Table 2.6. Primer pairs use to determine expression of genes correlated to the *hid* genes.

Chapter 3: Production of *hidΔ* mutants of *S. pombe*

- Table 3.1. Mutant strains created during the course of the study, the selectable marker and the number of independent isolates.
Figure 3.1: Diagram illustrating the process of mutant creation through gene replacement.
Figure 3.2. Location of primers for confirmation of gene replacement by PCR.
Figure 3.3: Molecular characterization of *hid1Δ::natMX6* strains.
Figure 3.4. Molecular characterization of *hid2Δ::kanMX6* mutant strains.
Figure 3.5. Molecular characterization of *hid3Δ* strains.
Figure 3.6. Molecular characterization of Vector Control strains.
Figure 3.7. Molecular characterization of double mutant strains.
Figure 3.8. Visible phenotype of cells.
Figure 3.9 Size properties of mutant strains.
Figure 3.10. Growth curves for *hid1Δ::natMX6* and *hid3Δ::natMX6* strains.
Figure 3.11. Growth profiles of *hid2Δ* mutant strains.
Figure 3.12. Transmission electron microscopy of thin layer sections of *S. pombe* cells.

Chapter 4: Transcriptional properties of the *hid1Δ* and *hid3Δ* mutants of *S. pombe*

- Table 4.1. Genes correlated in expression to each of the *hid* genes.
Table 4.2. Statistics for read mapping.
Table 4.3. DE genes selected for qRT-PCR for *hid3Δ*.
Table 4.4. Transcription Factor genes induced in *hid3Δ*.
Table 4.5. DE genes in *hid3Δ*.
Figure 4.1. Relative transcript levels of *hid* genes in wild-type (□) and VCN (■).
Figure 4.2. Differential expression of genes highly-correlated to the *hid* genes.
Figure 4.3. Scheme of preparation of samples for post-genomic studies.
Figure 4.4. Quality determination of total RNA preparations.
Figure 4.5. Visualisation of mapped reads at *hid* loci.
Figure 4.6 PCA of combined gene expression data from RNAseq.
Figure 4.7. Relative expression of *hid* genes based on read counts.
Figure 4.8. Number of genes differentially expressed in the mutants.
Figure 4.9. Comparison of relative transcript quantification by RNAseq and RT-qPCR.
Figure 4.10. Gene Ontology analysis of DE genes.

Chapter 5: Analysis of structural features of introns and their expression in fission yeast based on the *pcs3*⁺ gene family

Figure 5.1: Phylogenetic tree of cohesion gene orthologues in fungi.

Figure 5.2: Testing for intron orthology using Group 1.

Figure 5.3: Secondary structures for the orthologous introns.

Figure 5.4: Gel-based verification of product amplification by RT-PCR.

Figure 5.5: Intron expression in RNAseq performed on small RNA library.

Figure 5.6: Examples of novel expressed transcripts identified as snoRNAs.

Chapter 6: General Discussion and Outlook

Figure 6.1. The regulation between HID1 and FOXO4.

Figure 6.2. Transcription signalling pathways potentially affected in *hid3Δ*.

Figure 6.3. Analysis of tumour suppressor properties of HID1.

Acknowledgments

I would like to thank my supervisor Professor Mark A. HOOKS for his help throughout my study. I am very grateful to him for spending the time guide my work and correct a thesis on an organism none of us knew anything about at the beginning. Many thanks to Dr Katarzyna HOOKS for I have learned a lot about bioinformatics from her, and also for helping me to analyse the RNAseq data. I would like to thank Christophe Hubert of the Centre Génomique Fonctionnelle de Bordeaux for conducting the RNAseq analyses. Also, I wish to thank my colleague Alasmari Abdulrahman for being so helpful. Finally, I would like to thank my parents and my wife for supporting me during my study.

List of Abbreviations

bp	base pair
<i>C. elegans</i>	<i>Caenorhabditis elegans</i>
cAMP	cyclic-AMP
cDNA	complementary DNA
CESR	core environmental stress response
Ct	Comparative threshold
DE	Differential gene expression
<i>DMC1</i>	Down-regulated in multiple cancers
DNA	Deoxyribonucleic Acid
DNA-PK	DNA Protein Kinase
DR	<i>Down-Regulation</i>
DYM	DYMECLIN
EDTA	Ethylenediaminetetraacetic acid
ER	Endoplasmic reticulum
ERES	ER Exit Sites
G418	Gentamycin 418
GO	Gene Ontology
GOLPH3	Golgi Phosphoprotein 3
HID1	High-temperature-induced dauer-formation protein 1
KAc	Potassium Acetate
Kan	kanamycin
LB	Lysogeny broth
lncRNA	Long/large non-coding RNA
MFE	minimum free energy
min	Minute
miRNAs	micro RNAs
mRNA	messenger RNA
Nat ^R	Nat resistance
ncRNAs	Non-coding RNAs
NGS	Next Generation Sequencing
PCA	Principal component analysis
PCR	Polymerase Chain Reaction
PI4P	Phosphatidylinositol-4-phosphate
piRNA	Piwi-interacting RNAs
PKA	Protein kinase A
<i>P. pastoris</i>	<i>Pichia pastoris</i>
RIN	RNA Integrity Number
RNA	Ribonucleic acid
RNA-seq	RNA sequencing
rRNAs	ribosomal RNAs
RT-PCR	Reverse transcription polymerase chain reaction
RT-qPCR	Quantitative reverse transcription PCR
<i>S. cryophilus</i>	<i>Schizosaccharomyces cryophilus</i>
<i>S. japonicas</i>	<i>Schizosaccharomyces japonicas</i>
<i>S. octosporus</i>	<i>Schizosaccharomyces octosporus</i>
<i>S. pombe</i>	<i>Schizosaccharomyces pombe</i>
SDS	Sodium dodecyl sulphate
Sec	second

siRNAs	Short interfering RNAs
sncRNA	small ncRNAs
snoRNAs	Small nucleolar RNAs
snRNAs	Small nuclear RNAs
SREBP	Sterol receptor element binding protein
TBE	Tris-Boric acid-EDTA
TE	Tris-EDTA
TEMED	Tetramethylethylenediamine
tER	transitional-ER
TF	Transcription factors
TGN	<i>trans</i> -Golgi Network
Tris-HCl	Tris-Hydrochloride
tRNA	transfer RNAs
Ub	Ubiquitin
UCGs	ultraconserved genes
UR	Up-Regulation
Ura	uracil
VCN	Vector controls Nat ^R
VCU	Vector controls Ura ⁴
WT	Wild-type
YES	Yeast extract-based medium

RÉCAPITULATIF DU PROJET EN FRANÇAIS

INTRODUCTION

Les biologistes moléculaires s'intéressent depuis longtemps à l'expression des gènes parce qu'ils dictent les protéines qu'une cellule produit et en dernier lieu la fonction d'une cellule. De plus, l'expression des gènes change dans les cellules à mesure qu'elles sont affectées par des tensions ou touchées par des maladies (Barry et al., 2005). Le transcriptome est une image globale qui se produit de l'expression génique et il est décrit comme le jeu complet de transcrits d'une cellule. Il s'est donc avéré et demeure très intéressant de déterminer le transcriptome afin de comprendre les éléments fonctionnels du génome, mais aussi comment ils évoluent en cas de maladie (Berretta et Moscato, 2010).

La régulation de l'expression des gènes est importante pour les cellules eucaryotes, parce qu'elle sert à créer différents types de cellules (Chen et al., 2013) tout en menant à la différenciation des cellules (Thorrez et al., 2011). La régulation de l'expression des gènes se base sur le concept que les gènes peuvent être divisés en gènes exprimés de façon ubiquitaire et en gènes qui modifient l'expression selon l'ajout d'un signal (Ramsköld et al., 2009 ; Zhu et al., 2008). Depuis plus de dix ans, le profilage de l'expression génique sert à essayer de comprendre du point de vue génétique ce qui fait qu'une cellule cancéreuse se comporte différemment d'une cellule normale (Rapin et al., 2014). Il existe un grand nombre d'études de l'expression de gènes de différents types de cancers. Ces études ont démontré que l'expression des gènes peut servir à différencier entre les tissus normaux et cancéreux et à distinguer les sous-types de cancers, ce qui pourrait mener à des traitements spécialisés. L'expression des gènes sert également fréquemment à essayer d'identifier des marqueurs de tissus cancéreux. Par exemple, il existe des marqueurs génétiques pour plusieurs types de cancer, comme les marqueurs BRCA1 et BRCA2 pour le cancer du sein et CDH1 pour le cancer de l'estomac (Xu et al., 2010). Il est clair que le commencement d'un cancer et sa prolifération ont des effets graves sur l'expression des gènes. Pour bon nombre des changements observés dans l'expression génique, des schémas sont en commun avec d'autres maladies, comme celles d'inflammation chronique ou du métabolisme (Hirsch et al., 2010). Cela suggère qu'il existe des mécanismes fondamentaux par lesquels des cellules modifient leurs programmes transcriptionnels. Ces mécanismes de base sont liés à la réponse au mécanisme de signalement au moyen duquel les facteurs de transcription s'associent ou se dissocient des régions de transcription. Ces processus sont gouvernés par un jeu complexe de modifications protéiques qui se produisent dès le début du signalement jusqu'à la régulation de la transcription. Je vais me concentrer sur l'ubiquitination, un type de modification protéique, parce qu'elle est impliquée dans la fonction de HID1 chez les humains.

Objectifs des travaux et de la stratégie de cette présentation de thèse

La première question survenant de la description de la fonction de HID-1 ci-dessus est de savoir si sa fonction comme protéine de l'appareil de Golgi la rend compatible comme protéine de suppression de tumeur. J'ai présenté des rapports indiquant que la fonction de l'appareil de Golgi est liée aux processus nucléaires en relation à l'expression génique, mais cette preuve semble aller dans le sens que le maintien des processus de l'appareil de Golgi est nécessaire pour que le cancer progresse. Ces travaux visent à obtenir la preuve moléculaire en utilisant pombe comme système modèle pour déterminer si la perte de HID1 pourrait promouvoir une oncogénèse, une prolifération ou l'empêcher. Cette thèse comporte trois chapitres principaux de données et un chapitre final présentant le contexte biologique et les orientations futures des travaux. Chaque partie est résumée brièvement ci-dessous.

Chapitre 3: Production de mutants *hidΔ* de *S. pombe*

Il a été démontré que le gène *HsHID1* comporte une lacune ou que son expression est régulée négativement dans un certain nombre de types différents de cancer. De ce fait, il a été décrit comme un gène de suppression de tumeur de classe II appelé *Down-regulated in Multiple Cancers-1 (DMC1)*, Harada et al., 2001). Afin d'établir le rôle biologique de *HID1/DMC1*, nous utilisons *S. Pombe* comme organisme modèle. Ce chapitre décrit la création d'un mutant dont le gène orthologue Ftp105 du HID1 est absent, appelé Hid3 par le groupe. Ceci s'explique par le fait que pombe a deux autres gènes orthologues de *HID1* : *SPAC27G11.12* et *SPBP19A11.07c*. Des souches de mutants ont été créées pour lesquelles un gène particulier avait été remplacé par un marqueur sélectif (figures 3.1, 3.2). La majorité des travaux comprenait la vérification des bonnes suppressions et la confirmation que l'expression des autres gènes orthologues n'étaient pas affectée dans les souches individuelles du mutant (figures 3.3-3.7). Des mutants simples où un gène *hid+* a été remplacé ont été obtenus pour chacun des gènes (tableau 3.1). Plusieurs souches doubles de mutants ont aussi été créés, mais la création d'un mutant triple n'a pas été possible. L'analyse RT-PCR de l'expression génique a montré que l'expression des deux autres gènes *hid⁺* n'était pas affectée dans les mutants simples. Chacun des mutants simples et certains des mutants doubles ont été testés pour connaître les caractéristiques de croissance et les modifications morphologiques des cellules. Tous les mutants *hidΔ* avaient une morphologie cellulaire normale (figure 3.8) et des dimensions de type sauvage (figure 3.9). Il est apparu que la croissance de la souche *hid3Δ* était plus lente que celle d'autres mutants et des contrôles (figure 3.10), mais celle de *hid2Δ* semblait plus rapide (figure 3.11). Des études ultrastructurelles par microscopie à transmission électronique de cellules mutantes ont révélé que seul *hid3Δ* a des restes de l'appareil de Golgi (figure 3.12). Il est donc probable que la croissance plus lente de *hid3Δ* soit due à la perturbation du tri des protéines en raison d'un appareil de Golgi mal formé.

Chapitre 4: Propriétés transcriptionnelles des mutants *hid1Δ* et *hid3Δ* de *S. pombe*

Tout d'abord, le niveau relatif d'expression de chacun des gènes *hid*⁺ a été mesuré par RT-qPCR dans la souche de type sauvage. Les niveaux des transcrits de *hid1*⁺ et de *hid3*⁺ étaient 10 à 20 fois moins abondants que ceux du gène de contrôle *act2*⁺ et ceux pour *hid2*⁺ étaient presque 500 fois inférieurs à ceux de *act2*⁺ (figure 4.1). Une analyse par corrélation a permis de faire une investigation des gènes dont l'expression dépendait de la présence de chacune des protéines Hid. Pour chaque gène, deux groupes ont été identifiés à partir des données d'expression génique de Chen et al. (2013) : les gènes pour lesquels la corrélation était très positive et ceux pour lesquels elle était très négative à travers tous leurs traitements et échantillons (tableau 4.1). Des paires d'amorces ont été conçues pour des gènes choisis et l'expression a été mesurée par RT-qPCR dans le mutant approprié. Ce n'est que pour *hid3Δ* qu'il y avait une indication que l'expression des gènes correspondants était modifiée comme le montre une régulation négative générale des gènes à corrélation positive (figure 4.2). Un examen détaillé des évolutions de l'expression génique dans les mutants *hid1Δ* et *hid3Δ* a été réalisée par transcriptomique. Les travaux de transcriptomique ont été réalisés à l'aide de la technique RNAseq de séquençage de seconde génération sur des échantillons en parallèle issus de WT, des souches de contrôle exprimant le gène marqueur noursethricin-N-acetylase (*Nat*^R), les souches *hid1Δ* et *hid3Δ*. Les souches de mutants ont été regroupées pour tenir compte de la variation biologique et elles ont donné en tout 12 échantillons recouvrant les quatre génotypes (figure 4.3). L'ARN total a été isolé des échantillons en laboratoire, puis envoyé à la structure de séquençage de GCFB où un analyseur d'ARN en a déterminé la qualité (figure 4.4). Un séquençage a été réalisé sur la fraction des poly A+ ribosomiques sans ARN et son résultat a été analysé par l'intermédiaire du flux de travail Tophat/Cufflinks (tableau 4.2, Trapnell et al., 2012). Les données de séquençage ont montré que le gène *hid*⁺ approprié manquait dans chacun des mutants et que le remplacement des gènes était un processus très précis (figure 4.5). Une comparaison globale des données d'expression génique par l'analyse en composantes principales a indiqué que les évolutions de l'expression génique étaient plus importantes dans *hid3Δ* (figure 4.6). Les données RNAseq ont confirmé la relation dans l'expression génique chez les trois gènes *hid* qui ont été déterminés auparavant par RT-qPCR, *hid1*⁺ et *hid3*⁺ étant plus fortement exprimés que *hid2*⁺ (Figure 4.7). Les gènes dans *hid1Δ* avec des niveaux modifiés de transcrits étaient d'un nombre relativement faible comparé au contrôle négatif VCN, mais par contraste, plus de 500 gènes ont montré une évolution dans les niveaux de transcrits de *hid3Δ* qui est significative d'un point de vue statistique (Figure 4.8). L'évolution de l'expression génique a été déterminée par RT-qPCR pour un groupe de gènes de l'ensemble des données de *hid3Δ* (Table 4.3). Bien que l'amplitude de l'évolution d'expression ne soit pas la même dans les deux techniques, le sens de l'évolution dans les niveaux de transcrits était la même (Figure 4.9), ce qui est

courant avec les deux techniques quantitatives différentes. Les données d'expression des gènes issus des mutants ont été analysées à l'aide de Gene Ontology afin d'assigner des fonctions biologiques aux jeux de gènes en évolution (figure 4.10). Dans *hid1Δ*, la seule évolution apparente était une régulation négative des processus métaboliques. Dans le cas de *hid3Δ*, les groupes de gènes induisaient et réprimaient des cellules suggérées sous tension chronique (Figure 4.9). L'inspection des gènes individuels qui avaient évolué (Table 4.5) a montré une induction d'un grand nombre nécessaire pour que les cellules puissent entrer et rester dans un état quiescent nutritionnellement dépendant au niveau de la croissance (Sajiki et al., 2009). D'une façon intéressante, les gènes les plus fortement induits étaient ceux encodant la membrane plasmique ou les protéines sécrétées que trierait l'appareil de Golgi. Ceci suggère qu'il y a un processus qui se produit au niveau de la membrane plasmique qui est perturbé dans les mutants *hid3Δ*, qui est probablement une dépravation nutritionnelle menant à une régulation génique positive en feedback. L'ensemble induit de facteurs de transcription peut être responsable de cette régulation positive générale de l'expression des gènes (tableau 4.4).

Chapitre 5: Analyse des caractéristiques structurales des introns et leur expression dans la levure à fission sur la base de la famille des gènes *psc3*⁺.

Ce chapitre est volontairement présenté dans la mauvaise séquence. Les travaux de ma thèse ont commencé par une étude de l'expression possible d'introns ou de séquences d'introns avec un rôle fonctionnel possible. La rétention inhabituelle de l'intron terminal du gène *psc3*⁺ de pombe est à l'origine de mes travaux. L'objectif principal de ce chapitre était de donner l'expérience bioinformatique qui me permettrait d'extraire, aligner et manipuler des séquences acides nucléiques pour former des hypothèses fonctionnelles en déterminant la structure. Ces travaux ont été combinés à des expériences pour déterminer s'il est possible que cet intron soit exprimé dans pombe par RT-PCR. Ce chapitre a été mis à la fin parce qu'il présente un mécanisme de régulation de l'expression génique par des ARN non codants. Ceci pourrait constituer une prochaine étape d'investigation de la régulation de l'expression des gènes dans les mutants. Suite à l'étude RT-PCR non concluante, il a été tenté d'examiner l'expression globale d'introns avec de petites RNA libraries de wild-type et de *hid3Δ*. Il existait des lectures de base pour presque tous les introns de *psc3*⁺ et son paralogue *rec11*⁺, mais des introns de certains gènes ont été exprimés. La principale conclusion était que les petits ARN étaient de bons outils pour identifier des snoRNA dont environ 36 non annotés ont été découverts.

Chapitre 6: Discussion générale et perspective

Dans cette partie, je réalise une analyse critique de l'hypothèse de Harada et al. (2001) que le HID1 humain peut fonctionner comme suppresseur de tumeur et que la mutation ou régulation négative de l'expression du gène pourrait mener à la prolifération de la tumeur. La découverte du besoin par la protéine HID-1 du nématode *C. elegans* pour une croissance normale est un argument clair contre cette hypothèse. Il s'agit d'une protéine de tri de neuropeptides et son absence engendre en fait une inhibition de croissance cellulaire. Toutefois, la régulation négative des niveaux de protéines est une façon possible d'engendrer la prolifération de cellules si HID1 fonctionne par la déubiquitinase USP7. Les preuves suggèrent que HID1 séquestre l'USP7 à l'écart du noyau en le maintenant au niveau de l'appareil de Golgi. Quand les niveaux de HID1 baissent, l'USP7 est relâchée et peut migrer vers le noyau pour arrêter l'activation des voies inhibitives de tumeurs. Je fournis un exemple de ceci par le facteur humain de transcription FOXO4, qui est nécessaire pour engendrer des voies apoptotiques (figure 6.1). La régulation négative de HID1 peut également empêcher les gènes de suppression de tumeur p53 et pTEN de fonctionner normalement. Mes travaux avec le gène orthologue pombe de HID1 (SpHid3/Ftp105) suggèrent que le manque de HID1 empêcherait la prolifération cellulaire au lieu de l'augmenter. Les mutants de *S. pombe* qui manquent de Hid3 ont une croissance plus lente, connaissent des tensions plus fortes et présentent des évolutions d'expression génique associées à un état quiescent partiel. L'induction de gènes encodant des transporteurs de la membrane plasmique suggère que les cellules qui manquent de Hid3 sont à court de nutriments. Avec la perturbation probable de l'appareil de Golgi, il n'est pas apparent comment des protéines nouvellement synthétisées par plus de transcrits atteindraient la membrane plasmique. Il existe un certain nombre de voies intéressantes dans lesquelles ces travaux pourraient progresser à l'avenir, en particulier pour comprendre la relation entre l'évolution des facteurs de transcription et pour déterminer si les cellules *hid3Δ* sont dans un état quiescent (figure 6.2). De plus, les fonctions des gènes orthologues *hid1*⁺ et *hid2*⁺ doivent encore être découvertes.

Chapter 1

General Introduction

1.1. Nucleic acids and general transcription

All cellular organisms contain genetic information arranged in distinct macromolecular structures known as chromosomes. Chromosomes consist of two paired helical chains of nucleotides called Deoxyribose Nucleic Acid (DNA), where each nucleotide consists of a deoxyribose sugar, a phosphate group, and a nitrogenous base. The sugars and phosphate groups form the backbone of the DNA to link individual nucleotides in to the continuous strand and the two strands are held together by hydrogen bonds that form between the nitrogenous bases. In DNA there are four bases the two pyrimidines cytosine (C) and thymine (T) and the purines adenine (A), guanine (G). In DNA, A pairs with T and G pairs with C by two and three hydrogen bonds, respectively. Although DNA had been known well before the time of Watson and Crick, their publication in 1953 (Watson and Crick, 1953) describing the structure of double-stranded DNA as a helix provided the major insight into how DNA can be replicated (Alberts et al., 2002a).

Scientists at that time were interested in the role of ribonucleic acid (RNA) in the cell. Soon after the publication by Watson and Crick, it was found that other types of nucleic acids can form structures and through base pairing within the molecule can build complex structures (Felsenfeld et al., 1957). RNA is different from DNA, in that RNA contains a ribose sugar instead of deoxyribose sugar (differentiated by a hydroxyl group instead of hydrogen at the 3rd carbon), and uracil (U) instead of thymine. In general, RNA is usually single-stranded and is best known as the molecule that transmits the information within DNA for construction of the cell (Alberts et al., 2002b). There are number of classes of RNA that are produced from ribosomal RNAs to non-coding RNAs, but the vast majority of work has concentrated on the functions of messenger RNA (mRNA), which encode proteins that catalyse cellular reactions. In eukaryotes, the production of mRNA occurs primarily through the action of RNA polymerase making an RNA “copy” of a sequence of DNA known as transcription. Transcription of RNA starts and ends at particular sites in DNA to produce the pre-spliced RNA, which is then processed by capping, splicing to remove introns and then polyadenylation. The mRNAs are then exported from the nucleus into the cytosol where they are translated either by cytosolic ribosomes or sent to the rough endoplasmic reticulum for translation of secreted proteins. In 1961, Sydney Brenner (Brenner et al., 1961) called mRNAs the most

important part of the genome, specifically because they encode the proteins that makes cells function and cause them to be different. Scientists have long looked at the types of mRNAs present in the cell to see which proteins can be made. Certainly, it is not as simple now with continual revelations as to how RNAs can regulate cellular processes and protein quantity can be affected by post-transcriptional processes. The key point remains in that if an mRNA is not present in the cell, then the protein cannot be made.

1.2 Gene expression and its regulation

1.2.1 The transcriptional machinery

Gene expression dictates the proteins that are made in a cell and ultimately the function of a cell. Furthermore, gene expression changes in cells as they are affected by stresses or are afflicted by diseases (Barry et al., 2005). The transcriptome is described by the complete set of transcripts in a cell. Therefore, it has been of great interest, and still is, to determine the transcriptome in order to be understand the functional elements of the genome as well as to understand how these change in the event of disease (Berretta and Moscato, 2010).

In the nucleus of eukaryotic organisms, the transcription process starts from the transcription start site (TSS) in a gene. This site is preceded by promoter sequence which can attract an RNA polymerase (Pol). There are three types of RNA polymerases I, II and III and each type transcribes a specific class of RNA. RNA Pol I transcribes ribosomal RNAs (rRNAs), which are exported from the nucleus and form structural components of ribosomes. RNA Pol II transcribes genes responsible for producing the mRNA and non-coding RNAs. RNA Pol III transcribes genes encoding transfer RNA (tRNA), which carry amino acids to synthesising proteins, 5S rRNA and some small RNA (Ishiguro et al., 2002). Due to the fact that RNA Pol II is responsible for producing the variety of mRNAs that comprise the transcriptome of a cell, a lot of work has been placed in to determining its function and identifying those components that determine its specificity (Nikolov and Burley, 1997).

The specificity of RNA polymerase relies on a group of proteins called transcription factors (TFs) to start the transcription and controls the types of genes expressed. A TF binds to a *cis*-regulatory sequence (element) on the DNA that, often in combination with other TFs, directs RNA Pol II to bind and transcribe the gene. There are other sequences that can fine tune the amount by which RNA Pol II complexes are recruited to transcribe the gene. These sequences are called enhancers or silencers that can be present within the promoter, transcribed region including introns, or they can be found thousands of nucleotides from the promoter. Because TFs provide the mechanism by which gene expression can be regulated, they have been the focus of intense study in all cellular organisms from bacterial

(Seshasayee et al., 2011) to multicellular eukaryotes (Franco-Zorrilla et al., 2014; Vaquerizas et al., 2009). Transcription factors have generally been considered to possess DNA binding properties, but it must be recognized that many other processes contribute ultimately to the expression of a set of genes including cell signalling, RNA modification, RNA splicing, and chromatin structure. Simply the fact that enhancers and silencers can act from thousands of bases away indicates that chromosome higher order structure is an essential part of gene expression.

1.2.2 Chromatin structure

The regions of chromosomes with transcription activity are associated in general with a particular type of open DNA structure call euchromatin as opposed to the tightly packed DNA structure call heterochromatin. The open structure of euchromatin is important to permit the binding of multisubunit TF complexes and ultimately RNA Polymerase to the promoters of genes destined for transcription. In addition, an open structure to facilitate communication among distantly positioned transcriptional enhancer or repressors sites. Chromatin is in a highly dynamic structural state where sections of a chromosome are opened and closed for gene expression to be induced or repressed. Modifications of chromatin structure, for example, through histone modification by acetylation, methylation, phosphorylation and ubiquitination, etc., are universal mechanism for eukaryotes to alter levels of gene transcription (Bannister and Kouzarides, 2011; Berger, 2002). Large changes to chromatin structure can come through chromatin remodeling complexes, such as the ATP-dependent and histone-modifying complexes (Narlikar et al., 2002). It is clear that chromatin structure and its modifications contribute to the initiation of cell differentiation and its maintenance in multicellular organisms. Furthermore, it plays an important role in how cells continue to function across cell division, for example, in maintaining gene transcription patterns, which has implication in proliferative cellular diseases, such as cancer (Jones and Laird, 1999).

1.3. The basic biology of cancer

Cancer is the most common cause of deaths all over the world. In 2012, there were about 14.1 million new cases and this accounted for about 8.2 million deaths (Ferlay et al., 2014). By the year 2025, it is expected that there will be around 20 million new cancer cases each year. The most common cancers were in the order of lung, breast, and colorectal cancer with the majority of death coming from lung, liver and stomach cancers. The major reason for the large number of deaths is environmental with the use of tobacco and alcohol, poor diet, and lack of physical activity. It has been estimated that more than 30 percent of deaths caused by cancer can be prevented, but lack of awareness and lifestyle factors contribute to the rapid death rate because of the said disease. Cancer is an abnormal cellular state. Our body is

composed of billions of cells each with a specialized function, such as to form skin and its protective layer collagen, filter impurities from the blood by renal cells or transport oxygen throughout the body by blood cells. Each cell has a life cycle with thousands of cell dying every day in a programmed way and which are replaced by new cells. New cells come about by cellular division. Cancer is a genetic disease because the changes occur in particular genes, however, in most of the cases, cancer is not because of inheritance, like a genetic disease. Mutations occur in the genes and can build up in somatic cells throughout the lifespan of the affected person. Once enough mutations occur in the genes responsible for controlling cell growth and division, repairing DNA damage or initiate signalling, there starts an uncontrollable increase in cells. This results in the formation of malignant tumours. Further growth leads to invasion of healthy tissues called metastasis (Lobo et al., 2007).

Tumour initiation generally starts with chemical or physical induced breaks in DNA, epigenetic silencing of tumour-suppressor genes, or activation of dominant oncogenes. (Weston and Harris, 2003). The inhibition of gene function by itself is not sufficient to start tumour formation in a somatic cell, but a genetic background of a number of mutations in key genes could allow tumours to form if a critical gene is then mutated. As mentioned above, the process of tumour formation requires subsequent rounds of cell division, most notably at a rate faster than normal somatic cells of the surrounding tissue or normal processes of initiating cell death are inhibited. Malignancy occurs when cell growth becomes uncontrolled and begins to exert effects on the surrounding tissues causing a malignant phenotype. Malignancy is also the state where cell acquire a state of independancy and can migrate to other tissues. Malignancy then turns into tumour progression with the secretion of proteases that allow release from the area of formation and to expand to nearby or even far away tissues.

1.3.1 Genes involved in carcinogenesis: Oncogenes and tumour-suppressor genes

An oncogene is a gene that is activated in cancer and has a dominant effect. These genes have regular functions within a normal cellular context, but become deregulated in cancer cells. Examples of oncogenes are *ErbB2* (the target of herceptin), *PI3KCA*, *MYC* and the cyclin D1 *CCND1*. Examples of well-known tumour suppressor genes are *TP53*, *PTEN*, *BRCA1* and *BRCA2* (Lee and Muller, 2010). The common nature of these genes is that they are part of normal cellular signal pathways that initiate responses to changes in cell or extracellular environmental states. When they become deregulated, entire down-stream pathways can become deregulated causing altered cell processes. The PI3K signaling pathway in breast cancer is a good example, because it responds to extracellular growth factors and can initiate cell division or changes in cell metabolism. It works through other signalling pathways, such as the AKT serine kinase family of proteins, therefore activation of fundamental growth pathways has multiple consequences on

cell processes. The suppressor of the PI3K signalling pathway is PTEN. Thus a common mechanism of breast cancer development is constitutive activation of PI3K signalling with a mutation inactivating PTEN. The ultimate tumour-suppressor is p53 where it responds to a number of cellular stress signals. p53 protein is kept low in normal cells, but it is activated by cellular stress through post-translational modification (Brown et al., 2009). Interestingly, part of the post-translational modification is through deubiquitylation, which keeps it present in the nucleus where it can induce transcription (Li et al., 2002). Activation of p53 leads to induction of a number of anti-growth processes, such as the induction of cell cycle checkpoint, apoptosis and autophagy. The activation of p53 can directly induce the expression of hundreds of different genes (Menendez et al., 2009). Therefore, it is clear that inactivation of p53 function will greatly disrupt a cell's ability to control proliferation and tumour development.

There are many types of tumour suppressors in cells in addition to the major ones discussed above. In general, any gene that exhibits no or reduced expression in cancer development, or is induced in response to apoptotic signals can be considered as being involved in the tumour suppression response. The protein HID1, which is the basis of my work, has been shown to have reduced expression in a large proportion of certain cancer types, such as liver and lung cancer (Harada et al., 2001). This work seeks to obtain molecular evidence using *S. pombe* mutants lacking the ortholog of HID1 that this protein may be able to act as tumour suppressor gene.

1.4 Gene expression in cancer

Regulation of gene expression is important for eukaryotic cells, because it is used to create different types of cells (Chen et al., 2013) as well as lead to cell differentiation (Thorrez et al., 2011). The regulation of gene expression is based on the concept that genes can be divided into ubiquitously expressed genes and genes that alter expression depending on a signal input (Ramsköld et al., 2009; Zhu et al., 2008). The major discrepancy between normal and cancer cells in terms of genes is in the up-regulation or down-regulation of particular genes (oncogenes and proto-oncogenes) that leads to cancer development and metastasis. For more than a decade, gene expression profiling has been used to try to understand on a genetic basis what makes a cancer cell behave differently from a normal cell (Rapin et al., 2014). There are many studies of gene expression from different types of cancer. These studies have demonstrated that gene expression can be used to differentiate between cancerous and normal tissues and that gene expression can be used to distinguish cancer subtypes, which may lead to specialized treatments. Gene expression has also been extensively used to try to identify markers of cancerous tissues. For example, there are genetic markers for several types of cancer, such as BRCA1 and BRCA2 for breast cancer and CDH1 for gastric cancer (Xu et al., 2010). It is clear that the onset and proliferation of cancer has severe

effects on gene expression. Many of the observed changes in gene expression share common patterns with other diseases, such as chronic inflammatory and metabolic diseases (Hirsch et al., 2010). This suggests that there are fundamental mechanisms by which cells change their transcriptional programs. These basic mechanisms relate to the response to signalling mechanism whereby TFs associate to or disassociate from regions of transcription. These processes are governed by a complex set of protein modifications that occur from the initiation of signalling down to the regulation of transcription. I will be focusing on one type of protein modification, which is ubiquitination. This is because it has been implicated in the function of HID1 in humans.

1.4.1 Regulation of transcription by Ubiquitin

Ubiquitin (Ub) is a 76-amino acid protein and it found in all eukaryotic organisms (Zhang, 2003). A primary function of ubiquitin is as a tag to mark proteins for degradation by the 26S proteasome (Kornitzer and Ciechanover, 2000). Proteins can be modified by attachment of one ubiquitin on a single site (monoubiquitination), multiple sites (multiple monoubiquitination) or to substrate-linked ubiquitin to form a polyubiquitin chain (polyubiquitination) (Ikeda and Dikic, 2008). The modifications of the proteins by ubiquitin are controlled by a three-step process, activation of ubiquitin by E1 ligases, conjugation of ubiquitin to E3 ligases by E2 ligases, and E3 ligase enzymes for modification of target proteins. In eukaryotic cells E3 ligases are present in large families to facilitate the modification of a wide variety of protein targets (Zhang, 2003). The action of ubiquitin ligases is balanced by the activity of deubiquitinating enzymes, which remove the ubiquitin chain or monomers. These enzymes also exist in large families (Komander, 2010). The balance of ubiquitination/deubiquitination plays a role in the regulation of numerous cellular functions, including protein degradation, DNA repair, gene transcription, chromatin dynamics and cell cycle progression (Hochstrasser, 2004; Schnell and Hicke, 2003). Their capacities to be used for a range of diverse modifications gives them in many respects functional properties similar to kinase/phosphatase pairs for their ability to modify proteins and regulate their function (Yao and Ndoja, 2012). The deubiquitinases of *S. pombe* have been the subject of an experimental survey into the functions of the various family members (Kouranti et al., 2010). What follows below is a discussion about the role of ubiquitination in the regulation of transcriptional processes.

The organization of DNA in eukaryotic cells as chromatin requires it to be reorganised depending on the transcriptional state required, either active or inactive. Opening up chromatin allows for the transcription machinery to access to target DNA sequences to facilitate transcription (Ouni et al., 2011). The precise mechanisms by which chromatin reorganization is initiated is not well known, but it clear that

posttranslational modifications of histones, leading to the recruitment of chromatin-remodeling factors is necessary (Li et al., 2007; Smith and Shilatifard, 2010) and ubiquitination appears to be involved in modification of chromatin structure. The first ubiquitinated protein described was histone H2A (Muratani and Tansey, 2003). H2A is monoubiquitylated, which appears to have a repressive role in chromatin restructuring (Wang et al., 2004; Zhou et al., 2008). In contrast, H2B, which interacts with H2A, is also monoubiquitylated and it appears to help maintain chromatin in an open state (Naresh et al., 2003). It also appears to activate the methylation of histone H3 to stimulating RNA polymerase II (Ouni et al., 2011). There are reports that histone ubiquitination may be one of the first markers by which transcriptional complexes can recognize chromatin in a transcriptionally active state (Muratani and Tansey, 2003).

The target for any transcriptional process is to control the activity and recruitment of RNA polymerase II. RNA Pol II is a target for regulation by ubiquitination and thus ubiquitination has a general role in transcriptional control (Muratani and Tansey, 2003). Ubiquitylation of RNA Pol II was first observed in response to UV-related DNA damage (Ratner et al., 1998). Transcriptional arrest or DNA damage induces ubiquitination and degradation of the largest subunit of RNA Pol II, Rpb1 (Svejstrup, 2007). The function of Rpb1 ubiquitination is really a last-result mechanism to resolve the situation of stalled RNA Pol II during transcription whereby it is tagged and degraded (Somesh et al., 2005). However, elongation pausing, Ub-tagging and degradation may also be independent of DNA damage (Daulny and Tansey, 2009).

1.4.2 Ubiquitination and Transcription Factor activity

Evidence from the work of many laboratories has shown that the ubiquitin system functions in the regulation of transcription either by limiting TF abundance or by facilitating transcriptional activator interactions. For example, the ubiquitin-proteasome pathway is one of the ways that are used to control the amount of certain transcription factors in the cell. Another is the turnover of activators that is coupled to an increase the transcriptional activity. This occurs to beta-catenin in the Wnt signalling pathway. In the resting state, beta-catenin is phosphorylated and subsequently targeted for degradation by ubiquitination. Activation by binding of Wnt to its receptor inactivates glycogen synthase-3 thereby preventing degradation due to the lack of phosphorylation and subsequent ubiquitination. This allow beta-catenin to activate transcription of its target genes. This system maintains a basal level of beta-catenin that can increase in response to Wnt signalling (Muratani and Tansey, 2003). It appears that p53 and the human HIF-1alpha (hypoxia induced factor 1-alpha) are also regulated in a similar manner. There are also a number of examples of a similar system in yeast, such as for the regulation of the transcription factors Gcn4 and Gal4 (Lipford et al., 2005; Muratani et al., 2005). When the F-box protein Dsg1 is deleted in

yeast, the mutant cannot use galactose as carbon source, because target genes cannot be transcribed. Evidence suggests that Dsg1p participates in the ubiquitination and degradation of the pool of Gal4 that is required to insure that transcription initiation complexes turn into completely active elongation complexes that produce mRNAs that can be properly processed (Muratani et al., 2005). These examples demonstrate that ubiquitination processes are directly, and intricately, involved in transcriptional regulation.

1.5 Introduction to the Golgi apparatus

The focus of the Golgi apparatus in this work relates to the known subcellular location of the HID1 proteins of humans, worms and fungi. As a prelude to the study, it was necessary to determine if processes occurring in the Golgi apparatus could affect gene expression or *vice versa*. Certainly, if a genetic mutation alters Golgi function directly then this might have knock-on effects on gene expression. The purpose of the next few sections is to show that the Golgi apparatus contains signaling systems that directly feed-back to transcriptional regulation in the nucleus. I first give a brief introduction and then describe in more detail about two signaling systems present in the Golgi involving GOLPH3 and SREBP, respectively. These are only examples as there are other Golgi located signaling systems.

The Golgi apparatus is part of the cell endomembrane system. In electron micrographs of cells, it looks like a set of flattened discs, but the size, shape and number of discs is dependent on the organism. The Golgi is divided in sections called the *cis*-, *medial*- and *trans*-cisternae and the *trans*-Golgi network. The primary role of the Golgi apparatus is the processing and trafficking of lipid molecules and glycosylated proteins (Scott and Chin, 2010). The *cis*-Golgi receives cargo from the endoplasmic reticulum via coated vesicles. The proteins are processed as they traverse the Golgi and are ultimately packaged into vesicle for sorting at the *trans*-face of the Golgi. The Golgi apparatus processes secretory cargoes and directs their delivery to various places both inside and outside of the cell. The Golgi apparatus also serves as a focal point for microtubule association and a store of calcium (Wilson et al., 2011). However, what is particularly intriguing and an essential foundation to this work are the potential signaling properties inherent to Golgi function. There is evidence to suggest that there are many signaling factors localized at the Golgi, which have function to sense and transmit stress signals (Hicks and Machamer, 2005). It appears that aberrant expression of vesicular trafficking components is involved in tumorigenesis (Mosesson et al., 2008; Polo et al., 2004). The expression of some of the Rab GTPases and their effectors, which work to regulate intracellular transport, are frequently deregulated in human cancers (Agarwal et al., 2009; Ho et al., 2012).

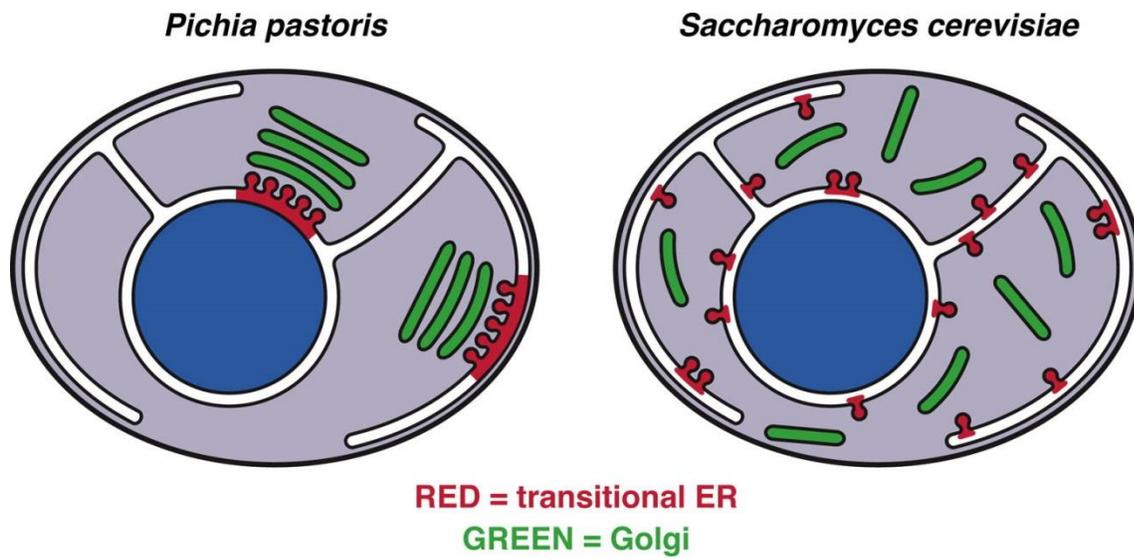


Figure 1.1. Diagram of the Golgi apparatus and ERES in *P. pastoris* and *S. cerevisiae*. In *P. pastoris* Golgi cisternae are organized into stacks and it has 2-5 discrete spot of ERES. However, in *S. cerevisiae* lost stacked Golgi and has numerous small spots of ERES (Papanikou and Glick, 2009).

1.5.1 Golgi apparatus in yeast

In eukaryotic cells, proteins destined for the plasma membrane or to be secreted to the extra-cellular space are synthesized within the endoplasmic reticulum (ER) and transported to the Golgi apparatus. These secretory cargos are modified and processed by Golgi enzymes in an ordered manner and then those proteins are sorted into carriers for transport to their ultimate destination at the trans-Golgi network (Papanikou and Glick, 2009; Suda and Nakano, 2012). Research suggests that, in general, unicellular fungi have stacked Golgi cisternae (Mowbrey and Dacks, 2009), however, some yeasts, such as *S. cerevisiae* have developed a dispersed, unstacked Golgi (Figure 1.4). Even in *S. pombe* the stacking is comparatively weak (Papanikou and Glick, 2009; Suda and Nakano, 2012). In addition, *S. pombe* appears to have multiple Golgi organelles with motile properties similar to higher eukaryotes, particularly plants.

Proteins destined for transport are exported from ER in the COPII-coated vesicles that form in specialized domains called ER Exit Sites (ERES) or transitional-ER (tER). There is a view that the ERES are a site of Golgi generation mediated by COPII formation of the pre-Golgi complex (Suda and Nakano, 2012). There are studies using *Pichia pastoris* suggesting that Golgi structure depends on the organization of ERES (Connerly et al., 2005). However; the ERES organization is different among yeast species. For example, the organization of ERES in *S. cerevisiae* looks like numerous small spots, whereas in *P. pastoris* ERES comprises only 2-5 discrete spots (Rossanese et al., 1999). The ERES in *S. pombe* has not been seen by electron microscopy, but it has been seen by fluorescence microscopy and it seems more like that in *S. cerevisiae* than in *P. pastoris* (Papanikou and Glick, 2009). The Golgi apparatus in budding yeasts has some similarity to the Golgi in mammals and plants, such the system of Golgi cisternae that are fenestrated and have tubular extensions (Papanikou and Glick, 2009). The Golgi reassembly stacking proteins factors of mammalian Golgi that play an important role in the formation of Golgi stacks have been found as well in the fission and budding yeasts, but they do not appear to be conserved in plants, and which also have clear stacks of Golgi cisternae (Suda and Nakano, 2012). Although there are differences in the organization of the Golgi apparatus among the yeast species and between unicellular and multicellular eukaryotes, the mechanisms of protein trafficking are quite similar (Papanikou and Glick, 2009; Suda and Nakano, 2012). Therefore, the yeast can serve as a good model for studying Golgi function in mammalian systems.

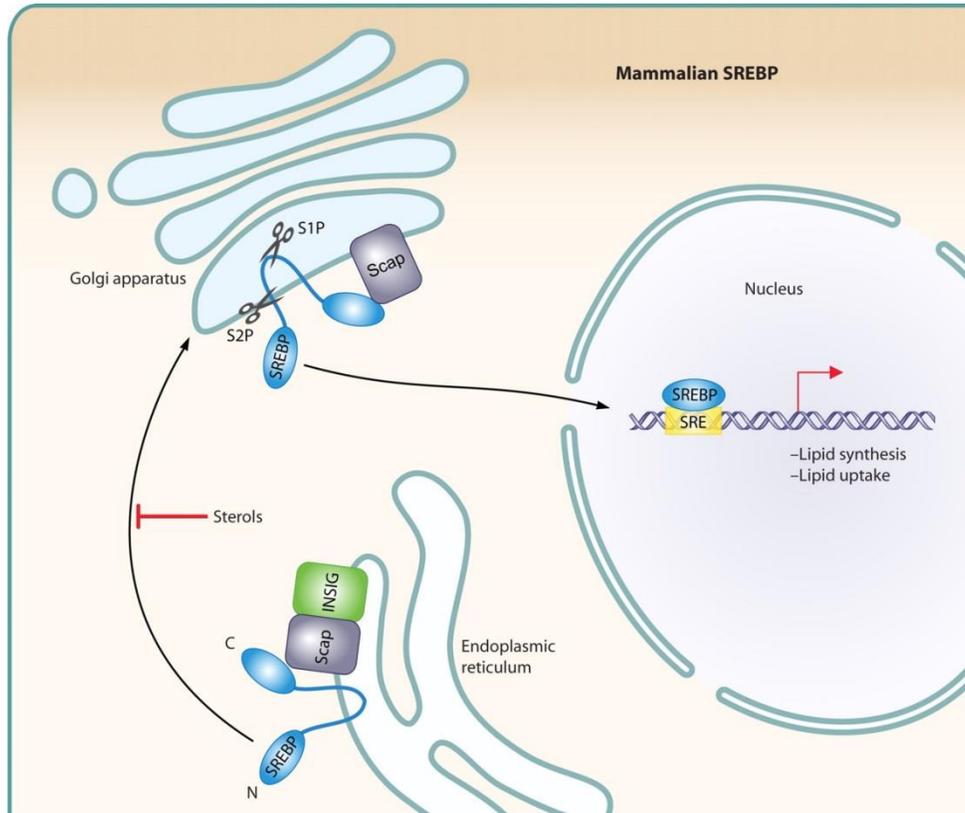


Figure 1.2. The scheme of the SREBP signalling pathway in mammals. SREBP is located in the ER. Upon stimulation, the protein is transported to the Golgi where ubiquitin-mediated cleavage occurs to release the TF so that it can bind to the SRE element in lipid/sterol biosynthesis genes. Image taken from Bien and Espenshade (2010).

1.5.2 Signaling through the Sterol receptor element binding protein (SREBP) pathway

The SREBP pathway is one of the best characterised systems by which processes occurring at the Golgi apparatus directly influence gene expression. In the eukaryotic organisms, the sterol lipids like cholesterol are important for maintenance of the membrane structure of the cell (Bien and Espenshade, 2010). For long periods of time, scientists had been trying to understand the mechanisms that kept an appropriate level of cholesterol in cells, because high cholesterol levels have been associated with metabolic diseases. High cholesterol levels can form solid crystals within cells being lethal and in the blood it is deposited in arteries leading to their blockage (Brown and Goldstein, 1997).

The SREBP pathway plays important role in the regulating of lipid homeostasis by controlling the synthesis of fatty acids, triglycerides and cholesterol (Steinberg, 2002). Sterol biosynthesis is controlled primarily at the level of transcription through the classic mechanism of supply and demand (Espenshade and Hughes, 2007). When sterol supply is not sufficient the induction of genes leads to the production of proteins for the synthesis of cholesterol and when it is sufficient transcription is arrested. The up-regulation of cholesterol biosynthesis genes is controlled by the SREBPs of which there are two genes *SREBP1* and *SREBP2* in humans. The two genes encode three SREBP isoforms which are SREBP-1a, SREBP-1c/ADD1 and SREBP-2 (Brown and Goldstein, 1997; Tontonoz et al., 1993). SREBP-1a can affect the expression of all SREBP-responsive genes, whereas SREBP-1c/ADD1 and SREBP-2 regulate genes involved in Fatty acid synthesis and cholesterol synthesis/homeostasis, respectively (Horton et al., 2002). In the ER membrane of mammalian cells, SREBP binds to the Scap protein, which mediates sterol-dependent regulation of SREBP activity (Hua et al., 1996; Rawson et al., 1999). If the membrane has enough cholesterol, the Scap binds to cholesterol leading to a conformational change (Horton et al., 2002; Radhakrishnan et al., 2004) which allows it to bind with SREBP to the ER-resident protein INSIG (Peng et al., 1997). The SREBP-Scap-INSIG complex remains on the ER membrane in an inactive mode (Espenshade, 2006; Rawson, 2003).

If cellular cholesterol levels drop, the binding of SREBP-Scap-INSIG complex is disrupted and the SREBP-Scap is transported to the Golgi by COPII vesicle-mediated transport. In the Golgi, there are two proteases required to process SREBP. The proteases Site-1 and Site-2 cleave SREBP in a sequential manner to release the N-terminal portion of the transcription factor from the membrane, and the SREBP can be transported to the nucleus to activate gene expression of cholesterol biosynthetic genes (Figure 1.1). Scap is then recycled from the Golgi back to the ER (Gong et al., 2015; Shao and Espenshade, 2014).

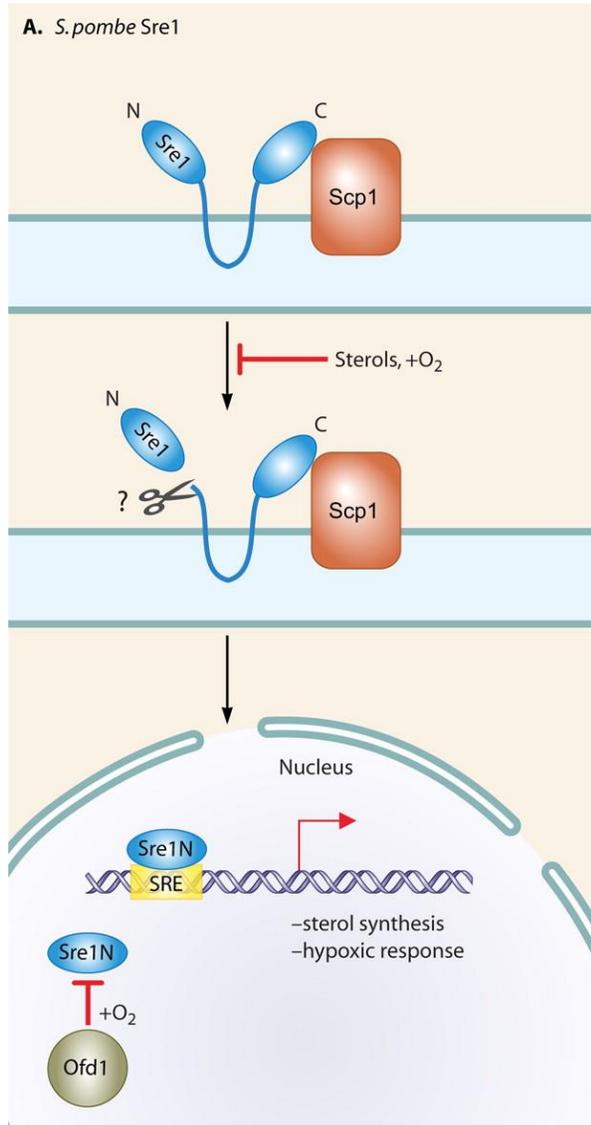


Figure 1.3. The Sre1 signalling pathway of fission yeast. Like in mammals systems Sre1-Scp1 are transported to the Golgi when ergosterol levels are low. The TF Sre1 is cleaved and then migrates to the nucleus where is can bind to SRE sites to induce fatty acid synthesis genes. The nuclear form is also responsive to oxidative stress. Image taken from Bien and Espenshade (2010)

The first *S. pombe* orthologue of SREBP was identified in 2005 and it was called Sre1 (Hughes et al., 2005). In *S. pombe*, not only does Sre1 regulate the proteins involved in sterol synthesis, it is also the major regulator of hypoxia related gene expression (Todd et al., 2006). Sre1 and Scp1 are present in the ER, and under low ergosterol or oxygen levels, Sre1 binds to Scp1 and is transported to the Golgi, where Sre1 is proteolytically cleaved to release the transcriptionally active Sre1N domain. Until recently, the proteolytic mechanisms for Sre1 cleavage remained unknown, because the *S. pombe* orthologues of Site-1 and Site-2 proteases were not apparent. Through a study to identify mutants that could not produce the Sre1N fragment, Stewart et al. (2011) identified four members of the Dsc E3 ubiquitin ligase complex. The Dsc E3 ligase complex is also required for the cleavage of *S. pombe* Sre2 (Stewart et al., 2011). Therefore, the Sre signaling system is conserved between humans and *S. pombe* with the key difference being the proteolytic cleavage of the TF domain (Figure 1.2). The Dsc E3 ligases form a stable structure and appear to be involved in proper targeting of proteins through the vacuole as part of the ESCRT pathway in *S. pombe* (Stewart et al., 2011; Takegawa et al., 2003).

As mentioned earlier, Sre1 is a key regulator of gene expression under hypoxic conditions. Under conditions of reoxygenation, its activity must be inhibited. This inhibition comes through the protein Ofd1, which is a prolyl 4-hydroxylase-like 2-OG-Fe(II) dioxygenase. The N-terminal domain of Ofd1 is an oxygen sensor and the C-terminus binds to Sre1 to target it for degradation (Hughes and Espenshade, 2008). It is known that Sre1N is degraded by the proteasome, but it is not clear if ubiquitination of the protein is required. In contrast, Nro1 positively regulates Sre1N activity by inhibiting Ofd1. In the presence of oxygen, the binding of Nro1 and Ofd1 is disrupted which leads to degradation of Sre1N (Lee et al., 2009).

1.5.3 GOLPH3

Golgi Phosphoprotein 3 (GOLPH3), which is also referred as GPP34/GMx33/MIDA is a peripheral membrane protein located at the *trans*-Golgi network (TGN). GOLPH3 protein is conserved among eukaryotes (Bell et al., 2001; Wu et al., 2000). GOLPH3 protein is important for protein trafficking and maintaining Golgi structure. It appears mobilized by stress and potentially represents a new type of oncoprotein that is involved in vesicular trafficking and cell signal transduction (Li et al., 2014). In humans there is a paralogue called GOLPH3L, and the interplay between GOLPH3 and GOLPH3L is important for maintaining Golgi structure in certain tissues where GOLPH3L is expressed (Ng et al., 2013). There is only one GOLPH3 protein in single cell eukaryotes. There are a number of studies that indicate that *GOLPH3* is an oncogene involved in several types of cancer, such as ovarian, breast, lung, pancreatic,

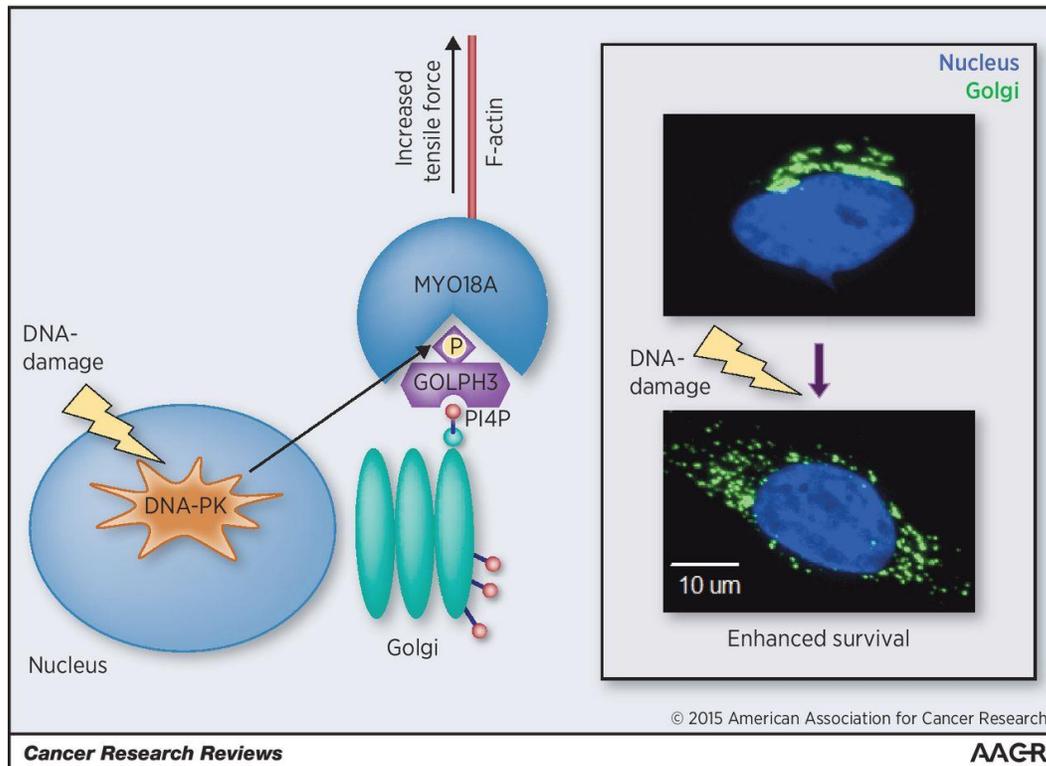


Figure 1.4. DNA damage leading to Golgi dispersal. UV radiation or other DNA damaging agents leads to double-stranded DNA breaks that activated DNA-PK. DNA-PK phosphorylates GOLPH3 which increases its affinity for MYO18A. It is thought that this increased affinity induced Golgi budding to the point that the Golgi apparatus disperses and shown by the fluorescent protein imaging of the Golgi in damaged cells. The image is taken from Buschman et al., (2015).

colon, prostate and melanomas. Topical over-expression of the gene in cancer cell lines increased proliferation (Zeng et al., 2012).

Phosphatidylinositol-4-phosphate is a lipid that is enriched in the *trans*-Golgi. GOLPH3 has been shown to interact tightly with PI4P, which dictates GOLPH3 localisation to this area (Dippold et al., 2009). Bound to PI4P, GOLPH3 recruits that complex MYO18A and F-actin to the *trans*-Golgi in order to facilitate vesicle budding and transport to the plasma membrane (Buschman et al., 2015). This appears to be the mechanism by which Golgi originating vesicles tie onto the actin network and disruption of any of these interactions stops vesicle budding. GOLPH3L also binds to PI4P but not to MYO18A and so for that reason there exists a competition between GOLPH3 and GOLPH3L and they have opposite effects on Golgi morphology. It is suggested that GOLPH3L serves to reduce secretory traffic, but this will occur only in cells where GOLPH3L is expressed.

1.5.4 GOLPH3 function and cancer

An interesting study by Farber-Katz and colleagues (Farber-Katz et al., 2014) revealed that DNA damage exerted by ionizing radiation affects the structure of the Golgi ribbon in mammals and leads to the Golgi dispersal throughout the cytoplasm. Therefore, it appeared that processes affecting DNA in the nucleus could have an effect on the Golgi. The mechanism that has been proposed to explain this is that in response to DNA damage, GOLPH3 is phosphorylated by DNA Protein Kinase (DNA-PK). Phosphorylation of GOLPH3 increases its interaction with MYO18A thereby increasing vesiculation and Golgi rupture and impaired protein movement from the Golgi to the plasma membrane. This process could be reversed or increased by overexpression or repression, respectively, of GOLPH3. The schematic showing the relationship between DNA damage and GOLPH3 function is shown in Figure 1.3. DNA-PK is a protein kinase specific to multicellular eukaryotes. Upon double-stranded DNA breaks it is recruited by the Ku complex DNA-PK binds and initiates DNA repair sequences. Phosphorylation of the complex binding the double-stranded breaks appears to stabilise the ends to keep them from improperly rejoining (Ciccio and Elledge, 2010). It is not quite clear how DNA-PK can phosphorylate GOLPH3. It has been suggested that GOLPH3 can cycle back and forth between the Golgi (Snyder et al., 2006), and so it could travel to the nucleus to be phosphorylated by DNA-PK. It is plausible that signaling mechanisms could increase the amount of PI4P in the *trans*-Golgi by relocation of the SAC1 phosphatase to the ER (Blagoveshchenskaya et al., 2008; Buschman et al., 2015). The increase in PI4P would then attract phosphorylated GOLPH3 back to the Golgi.

It is well known that disruption of Golgi processes can result in severe physiological problems in animals (Bexiga and Simpson, 2013). One example is the skeletal dysplasia that is the result of mutation in the DYMECLIN protein (Ghouzzi et al., 2003) to which HID1 has significant sequence similarity. Does disruption of Golgi function influence cell proliferative processes, such as tumorigenesis and cancer progression? It has been found in previous studies that many kinds of human cancers show overexpression of GOLPH3, and there is a good correlation between overexpression of GOLPH3 and poor prognosis for many cancers (Farber-Katz et al., 2014; Wang et al., 2014b). Furthermore, overexpression of GOLPH3 can prevent the apoptosis induced by DNA damaging therapeutic agents (Sechi et al., 2015). It has been shown that expression of GOLPH3 is a useful marker in the diagnosis and prognosis of certain cancers, such as mouth and breast cancer (Hua et al., 2012). It has also been suggested that GOLPH3 can be a useful marker to evaluate responsiveness to particular therapeutic treatments (Buschman et al., 2015).

1.6 Human HID1 and its role in cancer development

With GOLPH3, disruption of the Golgi occurs when the protein is phosphorylated and this can be reversed by topical expression of GOLPH3 allowing the cell to grow. Therefore, GOLPH3 represents a good example on how maintaining Golgi function may contribute to the development of cancer. It is well known that cancer development involves the creation of new blood vessels in order to provide nutrients to the expanding tumour (Nishida et al., 2006). Rapid tumour proliferation through cell replication then becomes a major sink for the consumption of nutrients and oxygen. Cells actively replicating within this environment would require synthesis and proper localisation to the plasma membrane of the proteins involved in the nutrient uptake. Cell progression in cancer requires an intact functional Golgi, perhaps even a hyperactive Golgi that will enable it to interact with its surrounding environment. It follows that a decrease in Golgi function may inhibit tumour proliferation.

The loss of heterozygosity in the 17q21 region of the long arm of Chromosome 17 has been implicated in the development of a number of cancers. This region holds the *BRCA1* gene involved in breast cancer development (King, 2014). Another locus near to *BRCA1* is the 17q25 region, which had been associated with the onset of esophageal and ovarian cancers (Kalikin et al., 1997; Risk et al., 1999). Harada et al, (2001) identified and cloned a previously unreported gene from this region 17q25.1 that they called Down-regulated in multiple cancers (*DMC1*). They went on to show that the expression of this gene is repressed or missing in a large proportion of different types of cancers, such as gastric, uterine, lung and liver. They concluded that *DMC1* represented a tumour suppressor gene, that when missing promoted cancer proliferation.

In 2003, *DMC1* was found to be orthologous to the *Caenorhabditis. elegans* HID-1 protein which when mutated leads to dauer formation in worms under conditions favourable for growth (Ailion and Thomas, 2003). In *C. elegans*, this protein is required for normal proliferative growth. Within recent years, evidence has accumulated that HID-1 of *C. elegans* is involved in neuropeptide secretion in neuronal cells, where it is mainly expressed (Mesa et al., 2011; Yu et al., 2011). HID-1 is a peripheral membrane protein of the medial- and trans-Golgi apparatus (Wang et al., 2011) and it has been found in the *trans*-Golgi network, in dense core vesicles and synaptic regions of nerve cells in *C. elegans* (Mesa et al., 2011). The orthologue of HID-1 in *S. pombe* is Ftp105, which has been shown to be Golgi localised and responsible for the localisation of the deubiquitinase Ubp5 to the Golgi (Kouranti et al., 2010). Kouranti et al, hypothesize that Ftp105 is removing Ubp5 away from the nucleus where it cannot function in the activation of stress-response pathways.

1.7 Introduction to non-coding RNAs

RNAs are split into two distinct classes based on their potential to encode proteins (Barciszewski, 2009). One category is messenger RNAs (mRNAs), which can be translated into proteins, and the other is non-protein coding RNAs (ncRNA) that do not encode for complete functional proteins (Eddy, 2001; Mattick, 2009). Since the discovery of mRNA and its initial characterization (Brenner et al., 1961), it has been the focus of most study, because mRNAs have long been thought to represent the active part of the genome. They encode the proteins that catalyse the reactions and that are the structural and signaling elements in the cell. However, 97–98% of the human genome is comprised of ncRNA (Griffiths-Jones, 2007). Although ncRNAs do not code for proteins, they are divided into numerous classes each of which has important and unique functions. In recent years, the number of ncRNAs known to be expressed in cells has grown enormously, particularly through high-throughput sequencing technologies that are not biased for one particular RNA type, such as poly A+ mRNA.

1.7.1 Classes of ncRNA

1.7.1.1 Ribosomal and transfer RNAs

The most abundant ncRNAs (at least in terms of mass) are the transfer RNAs (tRNA) and ribosomal RNAs (rRNA). Both tRNA and rRNA are essential in the translation of mRNA to proteins where rRNAs are structural components of ribosomes and tRNAs deliver amino acids to the growing protein chain during translation. The RNA component of ribosome called Ribosomal ribonucleic acid (rRNA). The ribosome is a large macromolecular assembly of RNA and protein in all cellular organisms (Pruesse et al., 2007). rRNA genes are easy to study and manipulate them in biological importance therefore these genes

were among the first genes that have been studied in detail (Sollner-Webb and Mougey, 1991). Because rRNAs are ubiquitous to cellular organisms, they are one of the most heavily-used models that scientists have used for studying molecular evolution of organisms. Ribosomes in both prokaryotes and eukaryotes consist of two subunits a small and a large. The small subunit is the site of codon-anticodon interaction between the mRNA and tRNA and the large subunit catalyzes the peptide bond formation (Barciszewska et al., 2001). In prokaryotes, the complete 70S ribosome consists of a 30S subunit that contains the 16S rRNA and 21 different ribosomal proteins. The large 50S subunit contains the 5S and 23S rRNAs and 34 different proteins. The 80S ribosome of eukaryotes also has two unequal subunits with four rRNA species and more than 70 proteins. The small 40S ribosome subunit contains the 18S rRNA and the large 60S ribosome subunit contains the 28S, the 5.8S and the 5S rRNAs (Torres-Machorro et al., 2010). Through the ribosome, the ribosomal RNAs provide the structure to allow association of the growing peptide chain with the tRNA and provide the peptidyl transferase activity. *S. pombe* as a eukaryotic organism has the 80S ribosome and corresponding rRNAs.

The tRNAs bring to the ribosome the necessary amino acids corresponding to the appropriate mRNA codon. The processes of transcribing, modifying, exporting and recruiting tRNAs to ribosomes is huge. There are more than 500 tRNA genes in the human genome and the number of tRNAs transcribed in the life-time of a cell is about 50-times more than all the mRNAs. In *S. pombe*, it appears that there are more than 180 potential tRNA genes including two pseudo genes (Chan and Lowe, 2009; Wood et al., 2002). tRNAs represent the largest class of known functional genes (Aziz et al., 2010). The processing of tRNAs have a number of steps including RNA splicing, cleavage of the 5' and 3' ends and many possible internal modifications (Phizicky and Hopper, 2010). Some tRNAs contain introns which are split out in order to form the functional tRNA molecule (Tocchini-Valentini et al., 2009). For example, in eukaryotes and archaea they removed by the tRNA splicing endonuclease, but in the bacteria these self-splice. In eukaryotes, a special enzyme RNase P removes the 5' sequence (Frank and Pace, 1998) whereas tRNase Z enzyme removes the 3' end (Ceballos and Vioque, 2007). In some organisms, 5' end-splicing does not occur and transcription starts at the site of the mature tRNA (Randau et al., 2008). As mentioned above, Pols I and III transcribe rRNA and tRNA genes, respectively. The transcription coming from both polymerases and the subsequent processing of RNAs must be tightly regulated in order avoid an imbalance in the components of protein synthesis (Phizicky and Hopper, 2010).

1.7.1.2 Regulatory ncRNAs

ncRNAs have become a hot topic within the past decade as a mechanism to explain the regulation of gene expression. This interest was increased, in general, by transcriptomic experiment showing that large parts

of the non-protein coding genome is expressed in cellular organisms (Barrandon et al., 2008). It was about 20 years after the discovery of mRNA that large RNA molecules were discovered that were 5' capped, but not polyadenylated, and they were not translated. These modified heteronuclear RNAs were almost 3-times more abundant than polyadenylated mRNAs (Salditt-Georgieff et al., 1981). In prokaryotes, ncRNAs are of interest, because they have specialised housekeeping functions and they participate in the response to different stress situations, and they may contribute to microorganism pathogenicity (Herbig and Nieselt, 2011). In eukaryotic cells, which are of main interest in this project, there are a number of classes of potentially regulatory RNAs. One type is an RNA regulatory element, which affects gene expression in both prokaryotes and eukaryotes, called riboswitches. This type was recently discovered in 2002 (Winkler et al., 2002). Riboswitches are parts of the mRNAs themselves that bind to and sense a specific small molecule and regulate the expression of certain genes. They are widespread in prokaryotes, but have been found in some plants and fungi (Sudarsan et al., 2003).

The various classes of regulatory ncRNAs participate in essential cellular processes especially in the regulation of gene expression (Barrandon et al., 2008; Eddy, 2001). They can exert their effects through a combination of complementary base pairing, forming complexes with proteins, and through their own enzymatic activities. For example, in 1982 it was discovered that RNA molecules can have enzymatic activities (Kruger et al., 1982) and that these activities may be regulated by other RNA molecules (Guerrier-Takada et al., 1983). Regulatory ncRNAs have been divided into two classes based on length. RNAs greater than 200 bp are known as long ncRNAs (lncRNAs) and those less than 200 bp are considered as small ncRNAs (sncRNA) (Brosnan and Voinnet, 2009).

1.7.1.2.1 lncRNAs

The size of lncRNAs ranges from 200 nucleotides to over 100 kilobases. Like mRNA, lncRNAs are transcribed by RNA polymerase II, and lncRNAs make up a large portion of the mammalian transcriptome (Mercer et al., 2009). lncRNAs frequently contain introns that are properly spliced out, but they can also generally avoid the subsequent standard mRNA processing steps. This enables them to remain in the nucleus close to sites of transcription (Motamedi et al., 2004). There is a kind of ncRNA that was discovered recently that are called ultraconserved genes (UCGs), which are transcribed from regions of the human genome called ultraconserved regions (UCRs) (Bejerano et al., 2004). The expression of these UCGs has a distinct tissue-specific pattern (Calin et al., 2007). It is thought that UCRs tie together the evolution of vertebrates, especially mammals, because UCRs are more conserved than protein-coding genes (Barrandon et al., 2008; Bejerano et al., 2004). Alignments of UCRs have shown that they are extremely similar among humans, rats, and mice despite 300 million years of negative selection (Bejerano et al., 2004; Mercer et al., 2009). UCRs are important because they contain evolutionarily retained microRNAs

(Berezikov et al., 2005). The function of the majority of lncRNA is unclear so far, but there is strong evidence indicating role in the regulation of the expression of the protein coding genes (Guttman and Rinn, 2012). The examples that have been uncovered have shown a diverse range of biological function, such as regulating gene transcription, RNA processing, chromatin modification and post-transcriptional gene regulation such as transportation, translation, splicing, editing, and degradation (Mercer et al., 2009).

There are a number of good examples of functional lncRNAs that have been well-characterised (Mattick, 2009). XIST is a lncRNA that is involved in the silencing of the X chromosome. It functions by remaining within the nucleus near the site of action (Guttman and Rinn, 2012). XIST interacts with the polycomb protein complex to condense chromatin structure thereby silencing the X chromosome, called chromosomal imprinting (Plath et al., 2002). The inactivation of a chromosomal region by XIST is accompanied by a large number of chromatin modifications, such as methylation of histones that distinguishes the inactive X from the active X chromosome. However, not all genes are silenced on the inactive X chromosome and these appear to have different histone modifications than XIST inactive sites. For example, it appears that histones in these sites are more acetylated than inactive sites (Plath et al., 2002). Disruption of polycomb group proteins that affect the association of the group with histone deacetylases appears to disrupt XIST-mediated inactivation, due to retention of acetylated histones (van der Vlag and Otte, 1999). Like XIST, HOX antisense intergenic RNA (*HOTAIR*) is a long intergenic ncRNAs that can silence certain chromosomal regions. *HOTAIR* is important in defining certain aspects of body differentiation in drosophila (Mattick, 2009). Where XIST regulates loci that are closely placed to it, *HOTAIR* silences the *HOXD* locus, which is spatially distantly removed from *HOTAIR*. In humans, ncRNA expression from the various *HOX* loci appear to have silencing properties that range to other chromosomes, and it too appears to work through interaction and recruitment of Polycomb group proteins to the silenced site (Rinn et al., 2007). How altering the function of these ncRNAs plays a role in disease is discussed below.

1.7.1.2.2 sncRNAs

1.7.1.2.2.1 miRNAs, siRNAs and piRNAs

MicroRNAs are probably the most intensely studied class of small RNAs in eukaryotes (Amaral and Mattick, 2008). The recent years has seen the identification of hundreds of miRNAs in animals and plants, with at least thousand having been found in humans. These small sequences play a central role in controlling gene expression (Stefani and Slack, 2008). The lengths of miRNAs are from 19 to 26 nucleotides and they were initially characterized in *C. elegans* relating to the regulation of gene expression in the timing of developmental events (He and Hannon, 2004). Since the time of their

discovery, miRNAs have been found in most a range of eukaryotic organisms including the viridiplantae, chromoalveolata, mycetozoa and even viruses, as well as the metazoa, which include the animals. According to the latest release of miRBase, there are 2588 mature miRNAs from 1881 precursor sequences. miRNAs have been shown regulate the expression o genes by affecting the translation of the mRNA. Base pairing between the miRNA and a target mRNA prevents movement of the ribosome, where it then dissociates from the mRNA. Thus, miRNAs generally act to decrease gene expression under conditions where the host sequence is expressed (Gottesman, 2002). miRNAs are synthesised by a particular complex called DICER that recognises (with the help of other components) a short a hairpin structure in an ncRNA retained within the cell. Once the RNA molecule is processed, it is release and can pair with its target mRNA (Aalto and Pasquinelli, 2012). Altered function of miRNAs has been implicated in a number of disease situations including cancer and this is discussed below.

Short interfering RNAs (siRNAs) are about 20-25 nucleotides and act much like miRNAs. siRNAs are derived from double-stranded RNA precursors that, like miRNAs, are cleaved by Dicer into short double-stranded RNA fragments (Farazi et al., 2008). They have a strong affinity with their target RNAs due to precise base pairing, unlike miRNAs which can have a range of target substrates due to certain mismatches being tolerated between miRNA and target mRNA. These associations are strong such that siRNA-modified transcript targets are usually degraded. The first description of this phenomenon was in plants, and the original idea was thought that it was specialized protection against RNA viruses (Hamilton and Baulcombe, 1999). It was also discovered that a large number of endogenous regulatory siRNAs are produced in animal and plant cells. siRNAs also can act as transcriptional repressors by modifying chromatin structure. In fact, *S. pombe* has an siRNA silencing mechanism and it has the genes Dicer and Argonaute. It has been most extensively studied in relation to heterochromatin silencing at centromere regions where siRNA molecules attract the silencing complexes starting with RITS. The addition of the complex containing RNA-directed RNA polymerase permits a feedback loop to form where more siRNA is generated from transcripts produced at the centromere (Djupedal et al., 2009).

A brief discussion of piRNAs is made, because they are the most abundant small RNAs in animals. piRNAs are a system that has evolved to fight against the negative properties caused by transposable elements (TE) jumping within the genome. The movement of transposable elements can lead to inactivation of important genes or even activation in some cases (Feschotte, 2008). It is estimated that almost 50 % of the human genome is made of TEs (Belancio et al., 2009), but *S. pombe* has only about 200 TEs only 13 of which are full-length (Bowen et al., 2003). It appears that in *S. pombe* that TE replication has come more through homologous recombination than trough transposition. piRNAs are

slightly longer than miRNAs being 24-31 nucleotides in length. There are nearly 20 million potential piRNAs in the human genome coming from about 6000 genomic locations. piRNAs are active in testes and bind to the Piwi class of Argonaute proteins for chromatin silencing of TE regions (Ha et al., 2014). Interestingly, in *Drosophila*, piRNA synthesis depends on a protein called Qin which has a terminal E3 ubiquitin ligase, but the function of this domain is unknown (Zhang et al., 2011).

1.7.1.2.3 snoRNAs

SnoRNAs are larger than miRNAs being from 60 to 300 nucleotides. More than a thousand different snoRNAs have been described in various organisms (Xie et al., 2007), including many hundreds in humans (Lestrade and Weber, 2006). There are generally two types of snoRNAs, those that are called C/D box snoRNAs and those called H/ACA box snoRNAs (Tollervey and Kiss, 1997). Most snoRNAs are processed from introns of precursor transcripts (Bachellerie et al., 2002). They guide chemical modification of and are involved in the processing of rRNAs, tRNAs and snRNAs and help in their correct folding (Fayet-Lebaron et al., 2009; Samarsky et al., 1999). Some snoRNAs appear to be involved in alternative transcript splicing (Kishore and Stamm, 2006) and altered snoRNA function can have negative effects on development as has been shown by removal of the Snord116 in mice (Ding et al., 2008). Moreover, snoRNAs were discovered whose altered functions are associated with particular disease states, such as Prader-Willi syndrome (Sahoo et al., 2008). There have been web tools developed that can help to identify snoRNAs in genomic sequences, such as RNAsnoop (Tafer et al., 2010) and snoSeeker (Yang et al., 2006). *S. pombe* has snoRNAs also and 20 have been found of which 13 were characterised and shown to guide modifications of the 18S and 25S RNAs (Li et al., 2005). Even though snoRNAs are highly expressed, it is not easy to identify the potential targets of action (Aspegren et al., 2004).

1.7.2 General features of ncRNA function

In order for mRNAs to silence gene expression, the microRNA binds to its target mRNA sequence and attracts the RNA-induced silencing complex RISC to bind to the mRNA. Usually miRNAs bind to the 3' UTRs of mRNAs, but can also bind to exon regions (Bartel, 2009). Because of the imprecise nature of the pairing between the miRNA and mRNA target, it is difficult to know exactly which nucleotides are being bound by the mRNA. As discussed briefly above, it was first believed that miRNAs simply prevented the complete translation of an mRNA by stopping the ribosome. This would appear to be a wasteful process if the miRNA was binding to the 3' UTR. However, it could also serve as a way to cause a positive feedback loop of inhibition if miRNAs were made from stalled messages through their degradation as explained for siRNAs. It is now believed that mRNA degradation following association with silencing complexes is a common mechanism of miRNA function and that is probably works in combination with

translational silencing (Huntzinger and Izaurralde, 2011). It is likely that the predominant mechanism of silencing depends on the cell type and developmental state (Aalto and Pasquinelli, 2012).

The majority of ncRNAs display some type of differential expression depending on cell and tissue type, and this differential expression may be involved in physiological or developmental processes *in vivo*. Over the past decade, the analysis of transcriptomic data has shown that expression of numerous intronic and intergenic ncRNAs is cell- or tissue-specific (Nakaya et al., 2007) and it can respond to environmental signals (Louro et al., 2007). It has been found that the individual expression of ncRNA respond to specific signaling pathways, and also many of them in adult cells are imprinted and differentially expressed during development (Kohtz and Fishell, 2004). There are common features between specific subcellular localization and ncRNAs (Amaral and Mattick, 2008). All of these signs lead to the conclusion that ncRNAs have biological functions. A good example is the expression of ncRNAs in embryonic stem cells. Many of them are involved in the organization of embryoid body differentiation and they are work to induce change in the structure of chromatin (Efroni et al., 2008) that are associated with specific differentiation events (Dinger et al., 2008). In general, expression levels of the ncRNA are lower than mRNA. In the yeast *S. cerevisiae*, many of ncRNAs are transcribed and they are detectable by conventional expression methods, such as RT-PCR or transcriptome profiling, but it is difficult to assign function, because they are rapidly degraded (Wyers et al., 2005). However, low abundance or rapid turnover of an miRNA would be required for it to act in a signaling role (Amaral and Mattick, 2008).

1.7.3 ncRNAs and cancer

Many studies indicate that miRNAs not only regulate various developmental and physiologic processes, but they also play significant roles in cancer and other diseases (Barbarotto et al., 2008). For example, approximately 50% of annotated human miRNAs are mapped to fragile regions of chromosomes. Those fragile loci have been linked with different types of human cancers (Calin et al., 2004). Fragile regions lead to genetic instability that can disrupt miRNA-target gene interactions and this regulation has been associated with an expanding number of cancer types (Selcuklu et al., 2009). Inhibition of miRNA function can lead tumorigenicity and cell transformation (Kumar et al., 2007). Moreover, miRNA have been shown to act as tumor suppressors by blocking the expression of malignant genes, but they also have the potential to be oncogenes if their function is altered and they are no longer able to repress the expression of genes promoting cell growth or division (Hanahan and Weinberg, 2000; Negrini et al., 2009). The miRNA mir-84 is a good example (Johnson et al., 2005). When the expression of mir-84 is reduced, RAS expression increases leading to cell proliferation and reduced mir-84 expression is an indicator of poor prognosis for lung cancers. In general, miRNAs are good markers for the early

discovery, diagnosis, and prognosis for cancer development and response to treatments, because they play a big role oncogenesis. For instance, miRNA-21 was the first miRNA shown to be plentiful enough in the sera of diffuse large B-cell lymphoma patients to be used as a marker for relapse after surgery (Lawrie et al., 2008). The expression of miR-21 was evaluated in serum and in solid tumours of this cancer type. In When the miR-21 is highly expressed it appears that cells will exhibit an increased response to the anti-cancer drug 5-fluorouracil, but while it is inhibited cell sensitivity to gemcitabine-based treatment will be increased (Rossi et al., 2007). In recent years, more and more studies have shown miRNA expression to be useful in the diagnosis of cancer, for example the use of miR-29a and miR-92a to diagnosis colorectal tumors (Huang et al., 2010). In addition, miR-92 also can be distinguished among inflammatory bowel disease, gastric cancer, colorectal cancer, inflammatory bowel disease, and healthy colons (Ng et al., 2009). The benefit of using miRNAs for diagnosis is that they can be detected by simple analysis of the blood, which is less invasive than tumour biopsy (Lee and Calin, 2011). There are two ways to treat cancer based on gene therapy by RNA. The first way is by using RNA as a remedy against the mRNA of the genes responsible for causing cancer. The second way is by targeting ncRNA that are directly involved in the disease. It is possible that manipulation of miRNA levels in tumour cells can be used to treat cancer. However, until now there are not any published of the toxicological or clinical studies regarding the treatment by using miRNA, but some are in clinical trials (Herbig and Nieselt, 2011; Lee and Calin, 2011). Since *S. pombe* does not have a classic miRNA system for regulating gene expression, it is not appropriate to go into large detail on the potential application of miRNA research for treating cancer with respect to this work. What may be more relevant to *S. pombe* as a model in studying gene regulation in respect to cancer, is the increasing evidence that deregulation of snoRNA function may be involved in some cancers. For example, C/D box snoRNAs have shown to be up-regulated in human breast and prostate cancers and snoRNA production can alter p53 function (Su et al., 2014a). It is likely that altered snoRNA function will become an increasingly important topic in cancer research.

1.7.4 Modeling small RNA structures and determining expression

NcRNAs do not have the same features as protein coding RNA that are capped and polyadenylated, therefore it is not obvious that a ncRNA may be functional in a cell. Some types of RNAs have well-defined structures that make it clear as to their function, such as tRNAs (Chan and Lowe, 2009) or rRNAs (Barciszewska et al., 2001). Because ncRNAs, particularly sncRNAs, have a short life-time in a cell and it is often difficult to find an RNA target for them, it is useful to have other means to identifying if an expressed RNA may be functional *in vivo*. The different classes of sncRNAs have been found to have particular secondary structures that can be recognized by processing enzymes, such as Argonaute and Dicer mentioned above. Therefore, it is possible to suggest that a particular sncRNA may be functional if

it has a secondary structure that may allow it to be retained in the cell. It is possible now to use de novo computational tools to determine if a short sequence of RNA has sufficient base-pairing capabilities for it to have a structure need to fulfill a function (Gorodkin et al., 2010). If an ncRNA has a function *in vivo*, it is likely that this function has been conserved in closely related species. Therefore, computational tools generally determine a structure for a potential ncRNA from a consensus sequence made from orthologous sequences from more than one species. However, this is not necessary easy to identify orthologous sncRNA sequences, because the sequences may have diverged but the structure has been conserved (Menzel et al., 2009). However, it is possible to identify orthologous sequences, such as introns in the same genes from closely related species. Currently available programs for RNA structure prediction are QRNA, EvoFold, CMfindertools, FOLDALIGN, Dynalign, RNAz and RNAalifold (Gorodkin et al., 2010). Once an RNA sequence has been shown to have a potential structure, it is necessary to confirm that sequence is expressed *in vivo* (Kavanaugh and Dietrich, 2009). There are many ways to show the expression of a predicted ncRNA, such as Northern blotting or RT-PCR (Kavanaugh and Dietrich, 2009). RT-PCR does have its limits in that primers are designed for specific sequences and sncRNA are frequently further processed, such that the predicted sequence may not be intact *in vivo*. However, the full-length RNA sequence can be determined by techniques, such as Rapid Amplification of ncDNA Ends (RACE) (Kavanaugh and Dietrich, 2009). DNA tiling microarrays can be effective in detecting sncRNAs, since the entire genome is usually covered on a chip (Hüttenhofer and Vogel, 2006). Now, NGS has become the primary technique for confirming the expression of short RNA sequences *in vivo*. The ability of RNA-seq to find expressed ncRNAs has already been demonstrated for both animals (Ruby et al., 2006) and plants (Kasschau et al., 2007). It is clear that transcriptomic techniques can only give hypotheses about the potential function of a ncRNA and further experiments are always needed to confirm both expression and function.

1.8 RNAseq as a gene expression tool

Gene expression is the process of converting the information stored within a cell's genome into the nucleic acid and protein components that conduct the biochemical activity of the cell. Gene expression can be considered to be two levels, the level to produce RNA from the gene, called gene expression, and the subsequent production of proteins, called protein expression. Protein expression also requires the complex process of converting the information in RNA to proteins. As discussed at the beginning of this introduction, molecular biology has greatly evolved from determining the expression of one or a few genes using Northern blotting techniques to studying the expression of all genes in one experiment using a transcriptomic technique. The desire is to now understand how genes come together to create a cell with a particular function, and how the processes of gene expression are regulated (Mutz et al., 2013). Taking

over from microarrays, RNAseq using Next Generation Sequencing technologies has become the standard for conducting transcriptomic studies. Microarrays or what is called reverse Northern blotting, was limited in the types of analyses that could be done, because of the limits of elements that can be placed on the solid support. In contrast, RNA-sequencing (RNA-Seq) can be used to address a number of biological questions, as well as measure transcripts, in one or a few experiments. It can be used to discover a novel fusion genes in cancer (Berger et al., 2010; Maher et al., 2009) and to quantify levels of alternative splicing in tissues (Wang et al., 2008) and disease (Garber et al., 2011). Moreover, it can be used to discover new genes and to measure their transcripts in single assay (Cloonan et al., 2008; Mortazavi et al., 2008) and it can measure allele-specific expression (Montgomery et al., 2010).

However, RNAseq is still a relatively new technique and there are still problems to overcome, mainly regarding the bioinformatic tools needed to explore the data. One of the most important challenges facing the researchers in the study of RNA-Seq experiments is the complexity of the analysis of RNA-Seq data, mainly because it lacks user-friendly software tools (Tjaden, 2015). An RNA-Seq experiment is subject to three main computational processes, alignment of reads onto a genome, combination of reads into transcripts, and finally quantification of the reads in each transcriptional unit. Each step gives information that is used in each subsequent step. The objective of a transcriptomic experiment is to measure those genes that change in a biological sample relative to a control sample, usually with replicates of each for statistical calculations. Therefore, computer programs that have been developed allow for simultaneous quantification of reads through files of common transcripts (Garber et al., 2011). There are a number of programs that have been developed to map reads to reference genomes, such as AceView, Novoalign, BWA, Tophat and Bowtie (Tophat uses Bowtie as its initial source of alignment), and several programs for calculating differential gene expression, such as EdgeR, DeSeq², and Limma (Wang et al., 2014a). Two particular tools have become very popular as a pipeline for analyzing RNA-seq data, the mapping program Tophat (Trapnell et al., 2009) and the read quantification program Cufflinks (Trapnell et al., 2010).

NGS sequencers, such as Illumina, Ion torrents (Life Technologies) or SOLiD (Applied Biosystems) output their sequences in the form of 'fastq' files that contain the sequence with a numerical identifier. Each fastq file can contain millions of RNAseq reads. The program Tophat take the reads from the fastq file for each sequencing run and independently aligns the reads to a reference genome. The number of reads mapped depends on a number of factors, such as the presence of multiple alignment sites and the number of mismatches tolerated for a sequence to be aligned. The output from Tophat is an alignment file for each sample run, such at this time each replicate has remained independent. A program, such as the

Integrative Genomics Viewer (Robinson et al., 2011), can be used to directly look at the reads that have been mapped to a locus. At this point there is no information as to what reads make a transcriptional unit, such as a gene or exon. Cufflinks is the first in a series of modules that groups the reads into genes and then quantifies them. Cufflinks uses the read file from Tophat and a genome annotation file to assign reads to loci and ultimately genes creating a transcriptome file for each sample. Because expression can only be quantified between the same genes in different samples, it is necessary to make a transcript file that has the common elements among the various individual transcriptome files. The Cuffmerge program does this to create one transcript file. The next step in the pipeline is the quantification of each gene based on the number of reads, which is the total sum of reads for a gene. Since longer genes will have more reads, a common way to express a value is based on the number of total Reads Per Kilobase (of sequence) Mapped, RPKM. This really isn't necessary for relative quantification between samples because the gene sizes will be the same, so some quantification programs do not normalize. The Tophat/Cufflinks pipeline uses two quantification methods Cuffnorm and Cuffdiff. Each of these requires assignment of read values to each individual gene by Cuffquant. Cuffquant calculates RPKM values for each gene for each sample alone using the file of common transcripts from Cuffmerge. Cuffnorm takes the quantified genes for each individual sample and does a sample-wise normalization of the gene data, similar to what Affymetrix does in order to compare different hybridization experiments. The data from Cuffnorm can be used for multivariate statistical analysis to compare sample properties. Cuffnorm does not combine the replicates. Cuffdiff is the program that determines DE expression. Using the RPKM data from Cuffquant, it calculates those genes that are statistically significantly different between biological samples. It combines replicates and outputs a pair-wise comparison between each two biological samples. From this, the experimenter has a list of total genes and those that are changing between samples. Cuffdiff has been shown to be a rigorous tool for performing comparisons in many high-resolution transcriptome studies (Graveley et al., 2011; Lister et al., 2011; Mizuno et al., 2010; Twine et al., 2011).

1.9 *S. pombe* as a model organism

The fission yeast, *S. pombe*, more commonly known as fission yeast, is unicellular eukaryote that was discovered in 1890 and began to be used as a model organism in the 1940s (Forsburg, 2005). The genome of *S. pombe* has been sequenced and annotated in 2002. It has 3 chromosomes with a total size of 12.6 Mbp making it a eukaryote with one of the smallest genomes known. The current statistics for the *S. pombe* is that it has 5048 known and predicted genes. In addition, it has 57 predicted snoRNAs, 1535 non-coding RNAs (ncRNA) and 5 small nucleolar RNAs (snRNA), suggesting that it has some of the

same epigenetic regulatory mechanisms as other eukaryotes (www.pombase.org). However, it does not have the typical microRNA processing and regulatory system seen in mammalian systems.

S. pombe is a popular model system that has been used for studies of cell growth and division, biochemistry and molecular genetics, because it is easy and inexpensive to grow in the laboratory and is genetically modifiable (Yanagida, 2002). Also, *S. pombe* is a good model for human chromosome studies, because it shares many features of human chromosomes such as heterochromatin, replication origins, small non-coding RNA regulation and complex centromeres, which are missing or different in budding yeast and other unicellular organisms (Wood et al., 2002b). In addition, some of the cell signalling pathways in *S. pombe* are similar to human such as MAPK signalling pathways and many human genes implicated in cancer in humans have homologues in *S. pombe* (Bailis and Forsburg, 2007).

1.9.1 The cell cycle in *S. pombe*

In all eukaryotes, the cell cycle is a central process that regulates growth and division (KV Venkatesh, 2013). Cell division consists of four main phases, which are genome duplication (S-phase), nuclear division (mitosis or M-phase) that are separated by two gap phases G1 and G2 that come between M-phase and S phase and between S-phase and M-phase, respectively (Bähler, 2005). *S. pombe* has become one of the most popular organisms to study cell cycle and division and it has a long history of research in this respect (Nurse et al., 1976). Cell division in *S. pombe* is characterised by the formation of a septum across the center of the cell cylinder that eventually allows the two daughter cells to separate into cells of equal size. The generation time of *S. pombe* that includes growth and cell division is between 2-4 hours depending on culture conditions (Guthrie and Fink, 2004).

In the cell cycle of *S. pombe*, the main control point for cell size is the G2 phase, whereas in many other organisms, such as *S. cerevisiae*, the main control point regulating when a cell divides is the G1-phase (Oliva et al., 2005). When the cell encounters the proper conditions in G2 with DNA fully replicated and cells matured to the proper size, the cell moves into M-phase that sees formation of the nuclear spindle and eventually separation of the sister chromatids. The septum begins to form and cell division is completed with cytokinesis and cell separation (Figure 1.5). Each of the daughter cells is in G1-phase and can enter into S-phase to synthesise DNA if the conditions are favourable. Once the complement of DNA is synthesised, the cell can enter into the G2-phase to grow. When culture conditions become poor, cryptic size control appears in G1 that leads to G1-phase becoming longer and cell division slows down. Cell division in *S. pombe* is regulated by the interaction of a number of cyclins, the master regulator being the cell cycle-dependent kinase Cdc2 and its interactions with three B-type cyclins Cdc13, Cig1, and Cig2 (Caspari and Hilditch, 2015; KV Venkatesh, 2013). The interaction of Cdc2 with Cdc13 is absolutely

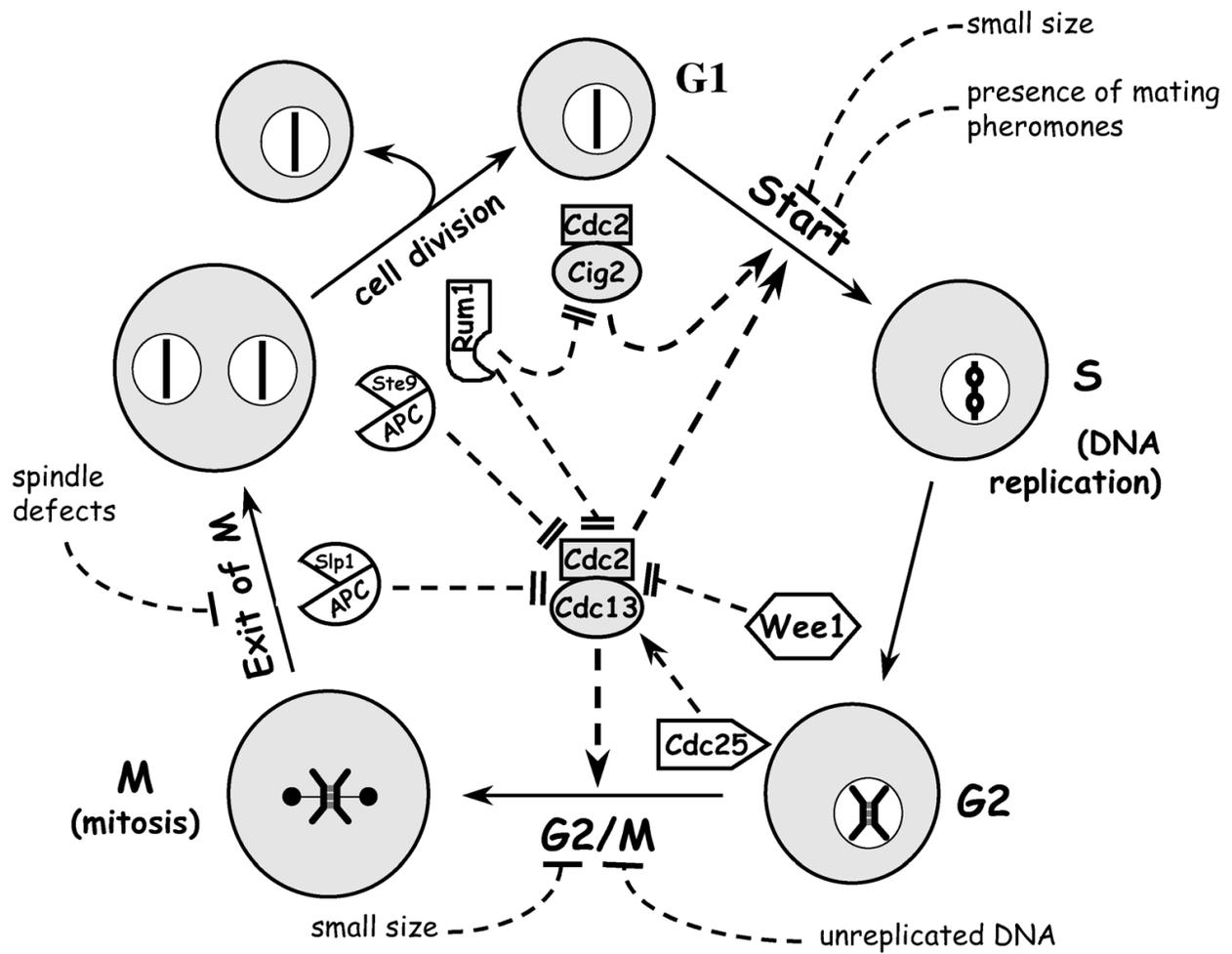


Figure 1.5: The cell cycle of a eukaryotic cell. In this figure show that the cell cycle phase ensured by regulation of by cyclin-dependent kinase (Cdc2/Cig2 and Cdc2/Cdc13). Cdc2/Cdc13 complex is activated by Cdc25 and in contrast it is negatively regulated by Slp1, Ste9, Rum1 and Wee1 (Sveiczer et al., 2004).

essential to initiate mitosis. Cig1-dependent kinase peaks at M-phase, whereas Cig2-dependent kinase peaks at G1/S transition (Sveiczner et al., 2004). The activity of Cig2 with Cdc13-dependent kinase is responsible for S-phase initiation. The action of Cdc2 and its interaction with these cyclins is governed by a number of other factors that check the cell cycle in phase transitions. Cdc2/Cdc13 is negatively regulated by four proteins Slp1, Ste9, Rum1 and Wee1 and is activated by Cdc25 (Figure 1.5). For example, when cells reach a critical size in G2, Cdc25 removes the inhibitory phosphate from tyrosine 15 of Cdc2 that then allows the transition into mitosis.

1.10 Objectives of the work and strategy of the thesis

The main question arising from the description of the function of HID-1 above is if its function as a Golgi protein is compatible with it being a tumour suppressor gene. I have presented reports that Golgi function is tied to nuclear processes relating to gene expression, but this evidence seems to go to the opinion that maintaining Golgi processes is necessary for cancer progression. The aim of the work is to obtain molecular evidence, using *S. pombe* as a model system, to determine if the loss of HID1 could promote tumourigenesis or proliferation or prevent it.

Chapter 3 presents the creation of the mutant lacking the HID1 orthologue *ftp105*, which the group has come to call *Hid3*. This is because *S. pombe* has two other orthologues of *HID1* *SPAC27G11.12* and *SPBP19A11.07c*. Accordingly, this study has become broader than originally planned to try to integrate these other genes. Mutant strains were created where a particular gene had been replaced by a selective marker. The bulk of the work entailed verifying the proper deletions and determining that the expression of the other orthologues was not affected in the individual mutant strains. The mutants were then tested for growth properties and cell morphological changes. The *hid3Δ* mutant was found to grow more slowly than the other mutants and controls. It had normal morphology, but a highly disrupted Golgi.

Chapter 4 presents the bulk of the molecular analysis. Transcriptional profiling was conducted using the Next Generation Sequencing technology RNAseq on replicate samples from wild-type, control strains expressing the marker gene noursethricin-N-acetylase (Nat^R), *hid1Δ* and *hid3Δ* strains. Gene expression was highly changed in the *hid3Δ* mutant compared to the other genotypes with the types of genes induced and repressed suggesting cells under stress. Interestingly, the most highly induced genes were those encoding plasma membrane or secreted proteins that would be sorted by the Golgi apparatus. This suggest that there is some process occurring at the plasma membrane that is being disrupted in *hid3Δ* mutants, which is likely nutrient deprivation leading to feedback up-regulation of the genes.

Chapter 5 has been presented out of order on purpose. My thesis work began with a study of possible expression of introns or sequences within introns with a potentially functional role. This was started because of the unusual retention of the terminal intron of the *psc3+* gene of *S. pombe*. The major objective of this chapter was to give bioinformatic experience for me to be able to retrieve, align and manipulate nucleic acid sequences to form functional hypotheses through structure determination. This was combined with practical experimental work to determine it is possible that this intron was expressed in pombe using RT-PCR. The reason why this chapter has been moved the end is because it presents a mechanism for regulating gene expression regulation by non-coding RNAs. This could represent a next step in investigate the regulation of gene expression in the mutants. As a follow up to the inconclusive RT-PCR study, an attempt was made to look at global intron expression using small RNA libraries of wild-type and *hid3Δ*. There were background reads to nearly all introns of *psc3+* and its paralogue *rec11+*, but introns from some genes were expressed. The main finding was that the small RNAs were a good tool to identify snoRNAs of which there were around 36 unannotated ones found.

Chapter 2

Materials and Methods

2.1 Bioinformatic analysis and expression of *psc3* intronic sequences

2.1.1 Prediction of structural features within *psc3* introns

The overall strategy of predicting structures within the introns of the *psc3* gene was to first identify orthologues of the encoded protein in diverse fungal species, retrieve the cDNA for multiple alignments in order to identify and confirm introns within Psc3, and then conduct structure prediction analyses on intron consensus sequences.

2.1.1.1 Identification of orthologous proteins in fungi

The first step in the identification and construction of the *S. pombe psc3* introns was to create a subset of all homologous proteins within fungi through searches of publicly accessible protein sequence databases. Searches were conducted using BLASTP as part of the BLAST (Lopez, 2003) toolkit available from the European Bioinformatics Institute (www.ebi.ac.uk). BLAST searches employed default settings (matrix BLOSUM62, topcombo 1, normal sensitivity, 1×10^{-10}) and the Psc3 protein as query template. The corresponding cDNA and genomic DNA sequence for each Psc3 orthologue was also retrieved for sequence alignment purposes. In cases where an organism contained multiple Psc3 paralogues, only the paralogue with the highest score was used.

2.1.1.2 Identification of orthologous introns in fungal cohesins

In this study, an orthologous intron was defined as a DNA sequence that was found between two corresponding exons of the orthologous *psc3* gene, as long as this sequence was present in the majority of fungal species retrieved in the BLASTP search. Orthologous introns in the *psc3* genes from the various fungal species were identified by BLASTN using the intronic sequences of *S. pombe psc3*. Orthologous intronic sequences were confirmed by mapping them to the corresponding positions within the primary sequences. Alignment of putative intron sequences with the reverse-translated cDNA sequences was conducted using the multiple alignment tool of MEGA (Tamura et al., 2011). In cases where the *psc3* orthologous gene did not contain any introns or did not possess the corresponding intron, these sequences were excluded from subsequent analyses. All of non-excluded genes were divided into five groups based on the phylogenetic distance.

Table 2.1. List of strains used in experimentation. This does not represent all the mutant strains created as part of the study. All mutant strains created are given in Table 3.1. In the VC strains, the marker has been inserted, but the locus is unknown.

Strain	Genotype	Source
WT	<i>h⁻ ade6-M26 ura4-D18 leu1-32</i>	McFarlane, BP90
<i>hid1Δ iso1</i>	<i>h⁻ hid1Δ::natMX6 ade6-M26 ura4-D18 leu1-32</i>	this study
<i>hid1Δ iso3</i>	<i>h⁻ hid1Δ::natMX6 ade6-M26 ura4-D18 leu1-32</i>	this study
<i>hid2Δ iso1</i>	<i>h⁻ hid2Δ::ura4⁺ ade6-M26 ura4-D18 leu1-32</i>	this study
<i>hid2Δ iso2</i>	<i>h⁻ hid3Δ::kanMX6 ade6-M26 ura4-D18 leu1-32</i>	this study
<i>hid2Δ iso3</i>	<i>h⁻ hid3Δ::kanMX6 ade6-M26 ura4-D18 leu1-32</i>	this study
<i>hid3Δ iso1</i>	<i>h⁻ hid3Δ::natMX6 ade6-M26 ura4-D18 leu1-32</i>	this study
<i>hid3Δ iso2</i>	<i>h⁻ hid3Δ::natMX6 ade6-M26 ura4-D18 leu1-32</i>	this study
VCU1	<i>h⁻ ade6-M26 ura4-D18 leu1-32 ura4⁺</i>	this study
VCN1	<i>h⁻ ade6-M26 ura4-D18 leu1-32 natMX6⁺</i>	this study
VCN2	<i>h⁻ ade6-M26 ura4-D18 leu1-32 natMX6⁺</i>	this study

2.1.1.3 Prediction of structure within introns of *psc3*

The realignment of intron sequences for the purposes of structure prediction was done for each of the five groups individually using ClustalOmega (Sievers et al., 2011). To predict the most stable RNA structure for each intron, sequence alignments were uploaded to RNAalifold program via the web server <http://ma.tbi.univie.ac.at/cgi-bin/RNAalifold.cgi>. The predictions were performed using RIBOSUM scoring and minimum free energy with partition function (Bernhart et al., 2008).

2.1.2 Detection of ncRNA transcripts from *psc3* intronic sequences by PCR

2.1.2.1 Strains and culture and storage conditions of *S. pombe*

The list of strains used in this study is given in Table 2.1. Long-term storage of all *S. pombe* strains was by freezing in YES media containing 25% v/v glycerol. The wild-type strain used in this work was genotype *hade6-M26 ura4-D18 leu1-32* (BP90, McFarlane Collection). Routine growth of *S. pombe* strains, unless specified otherwise, was done using a rich yeast extract-based medium (YES), which contained 5 g/l of yeast extract, 28 g/l of Glucose and 250 mg/l of each supplement Adenine, Uracil, Leucine, Histidine and lysine. Solid YES for cell growth on petri dishes was made with the addition of agar at 14 g/l. and incubated it at 30°C for 3 days.

2.1.2.2. Isolation of total RNA

Wild-type cells were grown in 25 ml of YES at 30°C with shaking at 200 r.p.m until they had reached on O.D.₆₀₀ = 0.2. Cells were harvested by centrifugation for 2 minutes at 2,000 r.p.m in a bench-top centrifuge capable of holding 50 ml conical centrifuge tubes. The supernatant was decanted and the cells frozen in liquid N₂. Total RNA was extracted by the hot phenol/chloroform method according to (Lyne et al., 2003) and described in more detail here. Frozen cells (~50 mg) were thawed on ice for 5 minutes, resuspended in 1 ml of pre-chilled DEPC-treated water (0.1% w/v DEPC) and transferred to 2 ml microfuge tubes. The cells were pelleted by centrifugation for 10 second at 5,000 r.p.m and washed once in DEPC water and collected by centrifugation. A volume of 750 µl of TES (10 mM Tris pH 7.5; 10 mM EDTA pH 8; 0.5% SDS) was added to the tube and the cells were gently resuspended with a pipette, and then 750 µl of pre-chilled acidic phenol-chloroform (Sigma P-1944) was added and the cells mixed by vortex, and incubated at 65°C for 1 hour in a water bath or heating block.

The sample was placed on ice for 1 minute, vortexed for 20 seconds, and then centrifuged at 14,000 r.p.m at 4°C for 15 minutes. A volume of 700 µl of the supernatant was transferred to a 2 ml, yellow phase-lock tube, which held 700 µl of acidic phenol-chloroform. The solution was mixed by inverting the tube and then centrifuged at 14,000 r.p.m at 4°C for 5 minutes. An aliquot of 700 µl of the top phase of the tube was transferred to a 2 ml, yellow phase-lock tube containing 700 µl of chloroform:isoamyl alcohol (24:1)

(Sigma C-0549). The solution was mixed by inverting the tube and the phases separated by centrifugation at 14,000 r.p.m at 4°C for 5 minutes. A volume of 500 µl of the supernatant was transferred to a 2 ml Eppendorf tubes containing 1.5 ml of 100% EtOH (-20°C) and 50 µl of 3 M NaAc, pH 5.2, vortexed for 10 seconds, and the RNA precipitated by incubating the tube at -20°C overnight. The following morning the RNA was collected by centrifugation at 14,000 r.p.m at room temperature for 10 minutes. The supernatant was removed, the RNA pellet washed with 500 µl of pre-chilled 70% EtOH (made with DEPC water) and then centrifuged at 14,000 r.p.m at room temperature for 1 minute to collect the RNA. The RNA sample was dried for 10-15 minutes by inverting the tube on a clean tissue. A volume of 100 µl of DEPC-treated water was added to the tube and it was kept at room temperature for 10 minutes and then resuspended by pipette.

2.1.2.3 Quality and quantity of RNA

The quality and quantity of RNA was determined by the absorbance measurements at 260 nm and 280 nm. The absorbance at 260 nm was used to quantify the amount of RNA according to 1 O.D. corresponding to a concentration of 40 µg/ml. The ratio of O.D.₂₆₀/O.D.₂₈₀ provided the measure of the purity of RNA with a good quality preparation having a ratio between 1.6 and 2.0 (Sambrook et al., 1989). The quality of RNA was also visualized by agarose formaldehyde gel electrophoresis using standard protocols.

2.1.2.4 Reverse Transcription

Potentially contaminating DNA was removed from total RNA samples by adding 1 µl DNase (1 U/µl) to 1 µg of total RNA in a total volume of 10 µl. Samples were incubated at 37°C for 30 minutes followed by the addition of 1µl of Stop solution and a further incubation of 10 minutes incubation at 65°C to inactivate the DNase. First-strand cDNA was produced by random-priming from 1 µg of total RNA using the QuantiTect™ Reverse Transcription kit according to manufacturer's instructions (QIAGEN Ltd.). Each reaction mix consisted of 2 µl of 1x gDNA Wipeout Buffer, 1 µg of DNase-treated total RNA, which was brought to a final volume of 14 µl using RNase-free water. The sample was incubated for 2 minutes at 42°C and then it was transferred immediately to ice. A volume of 1 µl of Quantiscript™ (Qiagen Ltd.) reverse transcriptase, 4 µl of 5x Quantiscript™ RT Buffer and 1 µl of RT Primer Mix were combined and added to the 14 µl of RNA template, mixed and incubated at 42°C for 15 minutes. The sample was incubated at 95°C for 3 minutes to inactivate the reverse transcriptase and the sample was diluted to an estimated concentration of 10 ng/µl for use in PCR reactions. The remainder of the total RNA sample was stored at -20°C until needed.

Table 2.2. Primers for amplifying sequences in *psc3*. Sections of primers given in *italics* represent intron sequences.

Gene Name	Description	Sense Primer (5'-3')	Antisense Primer (5'-3')
<i>psc3</i>	Spanning primer for Intron 1	GGATCCTCAAGAAGAAATTTTAAAC	TCAACTTTTTGGTCAAGCAGA
	Spanning primer for Intron 2	TGCCCTTGCAAATCTTATTAAC	ACGATTGCAACCACAGCA
	Spanning primer for Intron 3	TCTCAAATTCAACTTTCTGTAGAGC	TGAATTTAAAGGATAATCACGAGTAGA
	Spanning primer for Intron 4	TGGTGAAATAGTCAAGCAACAA	CGATATCGATGGACGAAGAC
	Spanning primer for Intron 5	AGCTCTAAAATTTTGCAACTTCA	TCACTGTTTTCCAGAAAAGTTCCG
	Junction primer for Intron 1	AACTTATTTGGTAATGATAAACTCC	AGCAGAAAAGCTTTTCAGTGT
	Junction primer for Intron 2	TATTTTAAAGGTAGAGTCATTCTT	AACCACAGCACTAAAATATTAG
	Junction primer for Intron 3	TGTAGAGCGTGTAAAGTGCT	AGTAGAAGTCTTAAAGTTGTTAATA
	Junction primer for Intron 4	ACTTTGATAGGTAAAGATCCC	GACGAAGACACTAATATAAA
	Junction primer for Intron 5	TGCAACTTCAGGTATACAAAATC	TTCGGGCTGCCTGGTC
	Internal primer for Intron 1	GTAATGATAAAGTCCTTC	CTTTTCAGTTAGTAG
	Internal primer for Intron 2	GTAGAGTCATTCTTTTC	CTAAAATATTAGTAAAG
	Internal primer for Intron 3	GTAAGTGCTTTTGCGA	CTTAAAGTTGTTAATATACA
	Internal primer for Intron 4	GTAAAGATCCCTCAA	CTAATATAAAGGTTAGTATA
	Internal primer for Intron 5	GTATACAAAATCAACTTA	CTGGTCAAGTTAGTTTC

Table 2.3. Primers for amplifying sequences of ncRNA genes.

Primer Name	Sense Primer (5'-3')	Antisense Primer (5'-3')
SPNCRNA.184 (1)	CGGGCGTATCCCTATATTCA	TCTGAATGCTAACCCTCAGTTG
SPNCRNA.184 (2)	ACAAGAAAGCTCTAGTGAACCTATG	CCCGAGCACGGTAGTTTTT
SPNCRNA.185 (1)	TCAAACCTATAAATCAGATTGGTGA	TTGAAGAATGCCTGCCTCAAT
SPNCRNA.185 (2)	TTGAGCAGGCATTCTTCAAT	TCTCAATCAGATATGCAGTTCCA
SPNCRNA.186 (1)	ACGAGCGTTTCGCTATCATT	TGGTGAGGTGCGATTTGTAA
SPNCRNA.186 (2)	CCCTCTCCCTGTAATCGAAA	CTGAAACGAGGGTACGGAAA
SPNCRNA.487 (1)	CGCAAATTGGAAGTGGTT	TCTTCATTGTGAAGGGAAGTGA
SPNCRNA.487 (2)	TCGCAATTTATCACCAAATGA	ACTCAAATGCGGACGAAAGT

2.1.2.5 Design of primers for PCR

Three different types of primer pairs were designed to determine the expression of intronic sequences of the *psc3* gene, an exon pair with the forward and reverse primers spanning the intron, a junction pair covering the exon/intron boundaries of each intron, and only intron primers. Primers for ncRNA genes were made as a control to test for expression of unrelated ncRNAs. Primers were designed using the OligoPerfect™ primer design tool from Invitrogen (tools.lifetechnologies.com) using genomic DNA sequences from Pombase (www.Pombase.org) with the constraint that all would have an annealing temperature of approximately 57°C. All PCR primers were obtained from eurofins Genomics (<http://www.eurofinsgenomics.eu>). The primers used in this study are given in Tables 2.2 and 2.3.

2.1.2.6 Polymerase Chain Reaction

All primers were tested by PCR on appropriate DNA or cDNA templates to ensure that the primers generated the expected size of fragments. The PCR reactions were prepared in a volume of 10 µl comprised of 1x GoTaq Flexi Buffer, 3 mM magnesium chloride, 0.2 mM of each dNTP, 0.05 µM of forward and reverse primers, 0.25 U of GoTaq DNA Polymerase, 10 ng of cDNA or DNA, and brought to the final volume with DEPC-treated water. PCR was conducted on a MJ Research thermocycler with the following cycling parameters: an initial denaturation step of one cycle of 1 minutes at 94°C, 34 cycles of 15 seconds at 94°C, 15 seconds at 55°C, and 15 seconds at 72°C, followed by an extension period of 72°C for 5 minutes.

2.1.2.7 Visualisation of PCR products by UREA-PAGE

Polyacrylamide gels were used in preference to agarose gels for determining the amplification success of fragments less than 100 bp. amplicons. All PAGE was run using the Miniprotean™ gel electrophoresis system from BioRad. Gels were poured at a thickness of 1.5 mm with a final composition of 1x TBE buffer (1 M Tris, pH 7.5, 0.9 M boric acid, 0.01 M EDTA), 15% bis-acrylamide and 8 M urea, 0.06 % v/v ammonium persulfate, and 0.05% TEMED. The gel was pre-run for 15-30 minutes at 200 V using 1x TBE. PCR samples were prepared in 1x Nucleic Acid Sample Loading Buffer (BioRad, #161-0767), loaded into the wells and the gel run at 200 V until just before the blue exited the gel.

2.2 Generation of mutant strains of *S. pombe*

2.2.1 Biological Material

Three different strains of *E. coli* were used in this project. These strains had the pFA6a-KanMX6, pFA6a-NatMX6 and pAW1 plasmids used to amplify the marker genes used for gene replacement mutagenesis of *S. pombe*. For long-term storage, bacterial cultures containing 25% glycerol were kept at -80°C.

Bacteria were grown on LB-agar plates consisting of 10 g/l (w/v) of tryptone, 5 g/l (w/v) of NaCl and 15 g/l (w/v) of agar or in liquid LB media without agar for isolation of nucleic acids. Agar plates and liquid cultures were incubated at 37°C with liquid cultures being agitated at 160 r.p.m in a rotary shaker. All powdered media was dissolved in MilliQ™ H₂O and sterilized by autoclaving. For screening cells by antibiotic resistance, media were allowed to cool to 55°C before addition of the appropriate antibiotic (100 µg/ml of Ampicillin, 50 µg/ml kanamycin or 100 µg/ml nourseothricin). The background genotype of *S. pombe* for mutagenesis and how it was grown is provided in Section 2.1.2.1.

2.2.2 Isolation of plasmid DNA from *E. coli*

Strains of *E. coli* were grown in LB broth containing the appropriate antibiotic at 37°C with rotation at 150 r.p.m until an O.D.₆₀₀ = 0.5 was reached. Plasmid DNA was extracted from a total of 3 ml of the culture using the alkaline lysis method according to (Sambrook et al., 1989). The bacterial pellet was resuspended in 100 µl of ice-cold Solution I (25 mM Tris-HCl, pH8.0, 50 mM glucose, 10 mM of EDTA) and allowed to sit for 5 minutes at room temperature. To the tube was added 200 µl of fresh Solution II (0.2 N NaOH, 1% w/v SDS) and the sample was mixed by gentle agitation of the tube for 10 seconds. The tube was incubated on ice for 5 minutes and then 150 µl of ice-cold Solution III (3 M KAc, pH 4.8), was added to the tube and the sample gently mixed by vortex for 10 seconds with the tube inverted. The sample was centrifuged in an Eppendorf centrifuge at 4°C for 5 minutes at 14,000 r.p.m. The supernatant was transferred to 2.0 ml PhaseLock™ microfuge tube (5 Prime GmbH, Hamburg, Germany) and an equal volume of phenol / chloroform was added to the tube. The sample was mixed by vortex and then centrifuged for 2 minutes at 14,000 r.p.m in a microcentrifuge. The supernatant was transferred to fresh tube and two volumes of 100% ethanol (-20°C) was added. The sample was mixed by vortex and allowed to incubate at room temperature for 2 minutes. The tube was centrifuged for 10 minutes at room temperature to pellet the DNA and the supernatant was removed by pipette and discarded. A volume of 500 µl of 70% (v/v) ethanol in H₂O (kept at -20°C) was added to the pellet, the sample gently mixed by inverting the tube several times, and the DNA pelleted by centrifugation for 5 minutes at room temperature. The supernatant was decanted and the sample was dried by standing the tube in an inverted position on a paper tissue. The DNA was resuspended in 50 µl of TE buffer, pH 8.0. To each sample was added 1 µl of 10 mg/ml RNase A (Novagen., UK), and the tube was incubated at 37°C for 15 minutes. The concentration and quality of DNA was estimated by absorbance measurements at 260 nm and 280 nm using a spectrophotometer, and an aliquot of the DNA was routinely visualised for quality by agarose gel electrophoresis and stained with SafeView™ (NBS Biologicals, Cambridgeshire, UK).

Table 2.4. PCR primers used to amplify transformation cassettes for gene replacement.

Primer	Sequence
<i>hid1</i> with marker gene (forward)	5`-TCTTTACAAATTCGCAATCATAGATTTTTCATCGTTTTTTTTTTCAGT TCACTCTCTGTGTCACCTTTGATTGTATGCATACTTTTCACCTTCACATTCG- CGGATCCCCGGGTTAATTAA-3`
<i>hid1</i> with marker gene (reverse)	5`-TCTTTGAACATACAAAACCTAGGATAGTTCAAATTTCTCCATTTCAATT CTTAAAACGGAGATATGCCGCTCAAGCATAATCCTCAAGGAGATTTCAAGT- GAATTCGAGCTCGTTTAAAC-3`
<i>hid2</i> with marker gene (forward)	5`-TTGCAGAAGTTGATTGCTCTTATTTCTAAAGCGTGTGGAATTAAG CATTTCCTTCTCAAATCTAAACTTTCATATTTAGCATTATATTGTAATCA- CGGATCCCCGGGTTAATTAA-3`
<i>hid2</i> with marker gene (reverse)	5`-ATAATTATTTAATGAGATATATTAATTTATAAAAATTGCTGCACTTGT AATGCTTATGACTTTTTTTGCGGTGTGAAAATTCATCAGTATTAAAAACG- GAATTCGAGCTCGTTTAAAC-3`
<i>hid3</i> with marker gene (forward)	5`-AAGTATACACATGGACGGTCTACTAGTACCTAACGATTATCGATGAAG CTGCAGCGTTTGTCAAGAGCACTTTACGGTTTTTCGCATCTAATTAGAGAAT- CGGATCCCCGGGTTAATTAA-3`
<i>hid3</i> with marker gene (reverse)	5`-ATAAATTTCAAAGTAAGATACCATAACATGCATGCAGGGTTTTCTTCA AATATTTTTTAGGACAAAAGTATACTTTAATTTACGCATTGAATTTAAGAA- GAATTCGAGCTCGTTTAAAC-3`

2.2.3 Design of primers for amplification of transformation templates

The primer sequences used to generate the fragments for gene replacement by homologous recombination of *hid1*⁺, *hid2*⁺ and *hid3*⁺ were obtained using the primer design strategy specified in (1998) and which is publically available as a design tool at http://128.40.79.33/cgi-bin/PPPP/pppp_deletion.pl. Each primer was 120 bases in length and contained a 20 base segment corresponding to the cassette sequence flanking the marker genes linked to 100 bases of a gene specific sequence (Table 2.4). The marker genes used in this project were the auxotrophic gene *ura4*⁺, the aminoglycoside phosphotransferase (*Kan*^R), and the nourseothricin acetyltransferase (*Nat*^R) gene. The plasmid templates from which the marker genes were amplified were pAW1 containing the *ura4* gene (Watson et al., 2008) , the pFA6A-KanMX and pFA6A-NatMX (Hentges et al., 2005).

2.2.4 Amplification and preparation of templates for gene replacement by homologous recombination.

Isolated plasmid containing the appropriate marker gene was linearized with an appropriate restriction enzyme to improve the PCR efficiency. pAW1, pFA6a-kanMX6 and pFA6a-natMX6 were cut with *Hind* III, *Nde*I and *Eco*RI, respectively. Each restriction digest consisted of 5 µg of DNA, 1x restriction buffer, and 3 Units/µg of DNA restriction enzyme in a total volume of 40 µl. Digests were conducted at 37°C for 1-2 hours and the efficiency of cutting was determined by agarose gel electrophoresis.

The PCR reactions to amplify the transformation sequences were prepared in following manner for a reaction mix of 400 µl: 50 µl of each primer (Stock 5 µM), 200 µl of MyTaqTM HS Red Mix (BIOLINE), and 50 µl of plasmid DNA (10 ng/µl) mixed with 100 µl MilliQTM purified water. This was mixed was divided into 8 PCR tubes and the reactions performed using an MJ Research thermocycler with the following cycling conditions: an initial heating of one cycle at 95°C for 1 minutes, 34 cycles of 15 seconds at 95°C, 15 seconds at 52.2°C the appropriate annealing temperature and 45 seconds at 72°C. The PCR reaction was terminated with elongation at 72°C for 5 minutes. The PCR products were purified according to the method of (1998). An equal volume of phenol/chloroform was added to a phase lock tube containing the pooled PCR reaction mix. The sample was mix gently by inverting the tube repeatedly and the sample centrifuged at 14,000 r.p.m at room temperature for 5 minutes. The aqueous layer was transferred to clean microfuge tube and 0.1 volumes of 0.1 M NaCl and then three volumes of 100% ethanol (-20°C) were added to the tube. The tubes were placed at -80°C for two hours in order to facilitate the precipitation of DNA. The sample was placed at room temperature for 5 minutes and the DNA collected by centrifugation at 14,000 r.p.m at 2°C for 15 minutes. The supernatant was removed and the

DNA pellet wash with 500 µl of 70% EtOH (kept at -20°C), mixed gently to minimize perturbation of the pellet and then centrifuged at 14,000 r.p.m at 2°C for 5 minutes. The supernatant was removed and the sample was allowed to dry for 5 minutes by inversion on a clean tissue. A volume of 13 µl of 1 x TE buffer was added to the sample and the DNA was used directly for transformation of *S. pombe* cells.

2.2.5 Transformation of *S. pombe* cells

S. pombe cells were transformed with the PCR fragments using the standard LiAc/TE transformation protocol according to (Bähler et al., 1998; Keeney and Boeke, 1994). Cells transformed with fragments carrying the Nat^R marker were plated onto two YES plates containing 100 µg/ml clonNATTM (Werner BioAgents, Jena Germany) and the plates were incubate at 30°C for 3 days. The colonies were re-streaked onto fresh YES plates containing clonNAT, allowed to grow for 3 days. Cells transformed with fragments carrying the Kan^R marker were plated onto two YES plates and incubated at 30°C for 18 hours. When the lawn of cells was visible the plates were replicated onto fresh YES plates containing 100 µg/ml G418 (Sigma-Aldrich Ltd) and incubated at 30°C for 2-3 day. The largest colonies were re-streaked onto fresh YES plates containing G418. Cells transformed with fragments carrying the ura4⁺ marker were plated onto two EMM plates without uracil and incubated at 30°C for 3 days. The colonies were re-streaked onto fresh EMM plates without uracil.

2.2.6 Verification of gene replacements

The success of mutant strain creation by gene replacement was verified both by genotyping the mutant strains and demonstrating the lack of corresponding transcript.

2.2.6.1 Isolation of genomic DNA from *S. pombe*

All strains were grown in liquid cultures of 5 ml of YES to late log-phase, and 3 ml of cells were harvested by consecutive centrifugation in a microcentrifuge spun at 6,000 r.p.m for 5 minutes. The supernatant was removed and the pellet was washed in 1 ml of sterile distilled H₂O containing 0.1% (w/v) Sodium azide. DNA was isolated from cell using the WizardTM Genomic DNA isolation kit according to manufacturer's instructions (Promega). In brief, the solution was centrifuged at 6,000 r.p.m for 2 minutes to pellet the cells and the supernatant was removed. The cells were resuspended thoroughly in 300 µl of a buffer containing 50 mM Na₂HPO₄, 11.5 g/l citric acid, 40 mM EDTA, 1 M sorbitol, and 0.1 mg/ml Zymolyase[®] 20T (Amsbio) and the tubes incubated at 37°C for 30 minutes to digest the cell wall. After cooling the tubes to room temperature (~ 2 minutes), the cells were again pelleted by centrifugation in microfuge at 13,000 r.p.m for 2 minutes, and the supernatant discarded. Next, 300 µl of Nuclei Lysis Solution was added to the tubes, the pellets were resuspended by gentle mixing and 100 µl of Protein

Table 2.5. Primer pairs for verification of *hid+* replacement with marker gene.

Primer Name	Sense Primer (5'-3')	Antisense Primer (5'-3')
Internal cDNA primers		
<i>hid1</i>	TTTAGGAACGCGGTTTACG	TAACGCGCATTTTGTAGTGG
<i>hid2</i>	AAAGCCAACAAAACGCAAGT	TAACGCGCATTTTGTAGTGG
<i>hid3</i>	CGCATTTGCCTTTTATCGT	TTATTCTATTATTTCGCCTCCAGCAAGTT
External <i>hid</i> gene primers		
<i>hid1</i>	CCTATGCTGCATCTTTGCTT	GCGAAGCATTATTTGTTTGG
<i>hid2</i>	ATCAATAACCGCTTTTGCCA	AATGCTCACCGCAAGAACT
<i>hid3</i>	TTTTCCTTACCCTGCCTCCT	AAGCTAGTGACAAGGTTGGAGC
Marker gene primers		
pAW1 (Ura4 ⁺)	CAGCTAGAGCTGAGGGGATG	AACATCCAAGCCGATACCAG
pFA6a-kanMX6	TTATGCCTCTTCGACCATC	ATTCGGACTCGTCCAACATC
pFA6a-NatMx6	ACTGGATGGGTCCTTCACC	CAGGGCATGCTCATGTAGAG

Precipitation Solution was added. The tubes were vortexed vigorously for 20 seconds and the tubes incubated on ice for 5 minutes. After centrifugation for 2 minutes and the supernatant containing the DNA was transferred to a clean 1.5 ml microcentrifuge tube containing 300 μ l of isopropanol at ambient temperature and gently mixed by inverting the tubes until the DNA precipitate became visible. The DNA was collected by centrifugation, the supernatant was removed, and the pellet was washed with 300 μ l of room temperature 70% ethanol by inverting the tubes several times. The DNA pellet was recollected by centrifugation for 2 minutes, the ethanol was discarded, and the pellet was allowed to dry by leaving the tube inverted on clean absorbent paper for at least 10 minutes. The DNA was dissolved in 50 μ l of DNA Rehydration Solution to which was added 1.5 μ l of a 1 mg/ml RNase Solution and the tubes incubated at 37°C for 15 minutes. Finally, the DNA was left to rehydrate overnight at 4°C.

2.2.6.2 Mutant genotyping and expression analysis by PCR

Four sets of primers were created to verify replacement of the endogenous *hid* gene by the selectable marker gene: a primer pair flanking the complete insertion site, two pairs of primers consisting of genome-specific and marker specific oligonucleotide to amplify the junction regions at both ends, internal cDNA primers to demonstrate removal of the *hid* gene, and a primer pair for marker gene insertion. All primers were designed online using the Primer3 program (http://biotools.umassmed.edu/bioapps/primer3_www.cgi) and the *S. pombe* gene sequences from Pombase as templates. The list of primers is given in Table 2.5. PCR was conducted using 10 ng of genomic DNA as described in Section 2.1.2.6. Total mRNA isolation and RT-PCR were conducted as described in Sections 2.1.2.2 to 2.1.2.4. The success of PCR reactions was determined visually by agarose gel electrophoresis.

2.2.6.3 Transmission electron microscopy

The protocol for fixation by Bordeaux Imaging Center in brief the cells for all strains were grown overnight in 5 ml liquid cultures under standard conditions. The next day the cells were diluted to 5×10^5 cells/ml in fresh YES media and grown to a cell density of 5×10^7 under standard conditions at which point they were harvested. 1 ml of culture was centrifuged at top speed in a microcentrifuge to collect cells. The supernatant was removed to leave a paste of cells at the bottom. The cells were prepared for TEM using a standard protocol developed for budding yeast. The cells were subjected to freeze substitution with on a Leica AFS2 leica with 1% OsO₄ in acetone. The substituted cell mixture was washed 4 times with acetone kept at ambient temperature. The cells were impregnated with increasing proportions of EPON resin as follows: 30 % EPON in acetone for 12 hours, 66 % EPON in acetone for 8 hours, 100 % EPON for 12 hours, and 100 % EPON for 10 hours. The cells in resin were

Table 2.6. Primer pairs use to determine expression of genes correlated to the *hid* genes.

Primer Name	Sense Primer (5'-3')	Antisense Primer (5'-3')
<i>hid1</i>, positively correlated		
SPAC27E2.09	ACCAAACCTTGTCGAGCCCTCTC	ACAACTGCTTGATCGACCTTCC
SPAC13C5.06c	TGGCCGGTGTAGTCAATTATGC	AACAGGTTTCTGCCCTTGACAG
SPAC30D11.08c	ACGGCACGAGAACTAAATCTGTG	TGTCAACCAGTTGGGAGATCGG
SPAC29E6.01	TGTTTCAGGCGGTACGATGGC	TACACAGTGTGTCTCGCCAGTG
SPAC29B12.07	CCCAAACCAAATTTGCCGGAGAC	TGGGAGCCTCAACTTCCAAGG
<i>hid1</i>, negatively correlated		
SPAC27E2.09	CGTCGTCGTGTCTTCTGTTTGTG	AATGAGACCACCGACAGTAGCG
SPAC9G1.11c	GCATGCCATTCGGATTGTTGG	AACGACCTACCATCAGGAGTACC
SPCC338.07c	ACGATTGGAACGTGCTGAGAAGC	ACTTTACCGGCATACCGTCTCTC
SPBC776.04	GCAGTCATTCCAATTGCTGCTATG	ACGCAGATCGACATGACAATAAGG
<i>hid2</i>, positively correlated		
SPAC19B12.08	TGCTCGATTAAGTGATCAGAACCC	TCCATTCTGGCAACATAAACGTG
SPBC29A3.08	CCCTTGGACGTTCTTCTTAACC	ACGTAGGGCGAGAGATTGATGG
SPCC1235.08c	ACATTCAGGGTCCGCTTATCATGC	TCAGAAAGAAGCATTCGCTTGACC
SPBC6B1.05c	AGTGGCAAGGAATTTGCTGTGC	TGCCTTTGGCAACTTTGCTTTGC
SPAC4D7.11	CAAGACCCTACAGTTCTCTTGCTG	AGCGGTAGAGCGTTTAACTGAGAG
<i>hid2</i>, negatively correlated		
SPAC17H9.11	TACTGGAGACCAAGTGCTACCC	CTTTGTGCACCTGCGACACATC
SPBC1347.02	TGGAATGCAAGAAGGCGGTGAG	AATGCTCTGGTTGCCGTAAGCC
SPAC3H1.07	TGTTGACGCTTGGGATCCCATC	TGCTTACGGAAGTCAAACCG
SPBC19C2.03	TGCGGAGCGCGTAATACAATCC	TAACACGATGCCACACTCTCG
SPCC285.12	AGGAACAGCTTCGCAATCCTGAG	TGTTCCACGAACGACTACAAGACC
SPAC23A1.10	GTGATGCTTGCAATGCTAAGATGG	CAGTGAAAGCTTCAACACACATGG
<i>hid3</i>, positively correlated		
SPAC17C9.07	TCGTGGGAGCTGTCACTCTATG	GCTTTCTCATGTACGTGCCAACC
SPAC23D3.13c	ATTGCTCCATTCGCAAGCC	TCCGCTCACTGACTTATCTGCTAC
SPBC1604.21c	TTCCGGAAACACTTGGTGCTTG	CTTCAACGTCTTACCAGTTGGC
SPAC1002.03c	GTAGAGCCGCTGAACTTACACG	TCCAGAAGCGTTCTTGCCAATC
SPAC21E11.08	TCCTGCCTGTCTTTGTTGACC	TGTCGGCTTTATTTGACTAGCG
SPAC22F8.08	GTGATCCACAATTGCGCTCATGG	AGCTGGGAGAGTTCTCACTACG
SPAC23G3.08c	ACCCGAGAGACGATGGCTTTAC	AACCTCGTCCCAGGATGATTTCG
SPACUNK4.07c	TGTAGATTACGCGGCTTGTGTTG	CTTCGGCTTGTGTCACGGAAG
SPCC1322.12c	GCCAATGACGGAGGAGTTAAGC	AACCCGAACCTCCAGTGGAATG
SPCC962.03c	ACATTCGCGAGTTGTCACTCTTC	TTTCGACACGCACACTCAGGAC
SPCC188.08c.1	AGATTTTCATGGTCACGGGTGAAAC	GTAAAGCGCGTCGCAATTCCTC
SPCC188.08c.2	ACGGTACATTTGGCTTTGGATCGTG	GCACGGTCGAGTATATTCCGGCTTC
<i>hid3</i>, negatively correlated		
SPAC57A10.14.1	CCTGCTTCTTTGCCGGTATTTTC	AGCATTCTTGTGGACTGTTTGGG
SPAC31A2.13c.1	CGAATTGATCGGGCTACTGAAAGC	CCTACGATTGCCAGCACCATTG
SPBC887.05c.1	TTGCACAGTCCTTGCCGTTGAC	AGCAACGTGATCAACACGAACC

polymerised in 200 µm flat molds at 60°C for 24 hours. The blocks were cut using a Leica UC7 microtome and the slices incubated in a mixture of lead citrate/uranylacetate in order to increase contrast. Cells were visualised on a Tecnai Spirit 2 FEI operating at 120 kV.

2.3 Characterisation of gene expression

2.3.1 Preparation of material and isolation of mRNA

The strategy for combining the various mutant genotypes to yield 3 biological replicates for wild-type, negative control and *hid1Δ* and *hid3Δ* strains is given in Chapter 4. In brief, cells for each strain were grown in 50 ml volumes of YES at 30°C until each culture had reached a final cell count of 5×10^7 cells/ml. The *hid3Δ* mutants required nearly 2 hours more growth time than other genotypes to reach this cell number. Cells were harvested by centrifugation for 5 minutes at 6,000 r.p.m at 4°C and immediately frozen in liquid N₂. Frozen cell samples were combined at the step of homogenization to yield three independent biological replicates each containing a mixture of two mutant or control genotypes. Three independent cultures were combined to give each biological replicate of wild-type. Total RNA was isolated from ~50 mg of powdered cells as described in Section 2.1.2.2 and quantified and the quality determined preliminarily by OD as specified in Section 2.1.2.3. The quality of the total RNA preparations was determined using an Agilent 2100 Bioanalyzer located at the CGFB sequencing facility. All samples exhibited RIN values greater than 0.9 and so were used for RNAseq and RT-qPCR quantification.

2.3.2 Calculation of gene correlations

Relative transcript amounts were obtained from the gene expression resource of the Bähler laboratory (<http://128.40.79.33/meta-analysis/meta-analysis.htm>, (Pancaldi et al., 2010). Using Excel, Pearson correlation coefficients were calculated pairwise between each *hid* genes and all other genes over all the conditions included within the dataset. The correlation coefficients were ranked in order to reveal those genes most strongly correlated in both positively and negatively with each *hid* gene.

2.3.3 Primer design for quantitative PCR of genes correlated to the *Hids*

All primers for quantitative PCR were designed using the QuantPrime program (Arvidsson et al., 2008) accessible from (<http://www.quantprime.de/>). Primer were designed to be suitable for a SYBR® Green detection system with a primer length between 20-24 bases, an amplicon size between 60 and 150 bp, a primer melting temperature between 61°C and 67°C, and a G/C content between 45% and 55% (Table 2.6).

2.3.4 Gene expression by RT-qPCR

A quantity of 1 µg of total RNA was reverse transcribed by random priming using the QuantiTect™ Reverse Transcription kit as described in Section 2.1.2.4. RT reactions were diluted to 10 ng/µl on the assumption that the RT reaction was 100% efficient. PCRs were run in 20 µl volumes containing 1 x iQ SYBR Green Supermix (BioRad Ltd.), 10 µM of each sense and antisense primer, and 10 ng of cDNA. Reaction mixes were loaded into hard-shell, 96-well PCR plates (BioRad) and sealed with Microseal® optically clear sealing film (BioRad Ltd). PCR was performed on a CFX96™ Real-Time PCR Detection System (BioRad Ltd). The PCR cycle protocol consisted of an initial denaturation step at 95°C for 3 minutes, followed by a 40-cycle amplification sequence of 94°C for 15 seconds, 60°C for 10 seconds and 72°C for 30 seconds. After a final denaturation step at 95°C for 10 seconds, the melting curve was done between 65°C and 95°C in steps of 0.5°C.

2.3.4.1 Analysis of RT-qPCR data

The comparative threshold (Ct) method was used to determine differential gene expression between mutant and wild-type (Schmittgen and Livak, 2008). The reference gene used was *act1*. The mean value for 3 biological replicates was used to calculate $2^{\Delta Ct}$ for a biological sample as long as the Ct values for all three replicates were within 1 Ct value of the mean. Reactions were not included within the data analysis if the melting curve properties were abnormal. All data was subjected to quantification by LinRegPCR (Ramakers et al., 2003), a quantification method based on extrapolating the initial fluorescence of the reaction from the fluorescent amplification profile. Only samples reaching the plateau phase of amplification and having at least four points with which to calculate the slope of the extrapolation line were included in the analysis.

2.3.5 RNAseq and differential expression analysis

Total RNA as specified in Section 2.3.1 was depleted of ribosomal RNAs prior to creation of the RNAseq libraries. RNAseq of *hid1Δ::natMX6* and *hid3Δ::natMX6* and corresponding vector controls (VCN) was performed at the CGFB on an Illumina MySEQ™. RNAseq of *hid2Δ::ura4⁺* and associated vector controls (VCU), and of wild-type and *hid3Δ::natMX6* (intron expression analysis) were conducted by GeT-Biopuces (Toulouse, France) on an ion torrent™ system (life technologies). Reads were aligned to the *S. pombe* genome obtained from Pombase (www.pombase.org) using TopHat v 2.0.11 (Trapnell et al., 2009) and differential gene expression between samples was performed using Cufflinks v 2.2.1 (Trapnell et al., 2012). Reads aligning to specific regions of the genome were visualized with IGV Viewer using the BAM file outputs of reads yielded by TopHat. Comparative analysis of biological samples by Principal Component Analysis was done using the entire gene output from CuffNorm. Cuffdiff was employed to

obtain differential gene expression ratios with loci referenced to the common *S. pombe* transcriptome file yielded by Cuffmerge to provide gene identification. The software packages for conducting the differential expression were located on the server of CIBiB at the University of Bordeaux's Centre for Functional Genomics.

For GO analysis of transcriptomic data, a file was used which had the sum of all transcripts identified by the CuffDiff software (6877 genes). The DE genes for samples *hid1Δ* and *hid3Δ* were placed into a file containing 2 columns - one with the gene identifier and another with designation if the gene was UR or DR. Data for both genotypes were submitted together to GoMiner web server (<http://discover.nci.nih.gov/gominer/GoCommandWebInterface.jsp>). The settings used were data course Pombase.org, organism *S. pombe*, 100 randomisations, 5-500 being the size of the category. Analyses for Biological Process, Cellular Component and Molecular Function were done separately.

2.3.6 Small size selected RNAseq and novel ncRNA transcript detection

Total RNA was extracted as specified in Section 2.3.1 from wild-type and *hid3Δ* mutants and depleted of ribosomal RNAs. Size selected libraries with mean RNA fragment sizes ~80 nt were created by GeT-Biopuces (Toulouse, France) and sequenced on an Ion Torrent™ system (Life Technologies). Reads were mapped to the *S. pombe* genome version ASM294v2.27 using Tophat with the library type chosen as strand specific, maximum intron size of 800 bp and no transcriptome annotation. The Cufflinks pipeline was used to find novel transcripts using the same genome build, but without the *gtf* annotation file and a maximum intron size of 800 bp. Transcript.gtf files from wild-type and mutant were merged with bedtools to obtain one file covering all regions containing any evidence for transcription. The mean FPKM count for merged features, strandedness and the initial names of transcripts were preserved. Transcripts longer than 45 nt (the size of the smallest snoring annotated for *S. pombe*) and FPKM higher than 120 were manually inspected in IGV, and only transcripts with consistent signals for the start and end of transcription were considered as real. This process resulted in 36 novel ncRNAs.

2.3.7 snoRNA prediction from novel ncRNA transcripts

The list of 36 novel ncRNA transcripts identified as expressed in small RNAseq library was used for prediction of CD box snoRNAs with the snoscan 0.9b (Lowe and Eddy, 1999). Methylation sites were predicted by snoscan using all rRNAs annotated in Pombase. Remaining ncRNAs were used as queries to search for orthologs in all *Schizosaccharomyces* species: *japonicus*, *octosporus* and *cryophilus* employing WU-BLASTN with following settings: -hspsepsmax 200, -E 0.01. For those ncRNAs found in all four *Schizosaccharomyces* sp., alignments were build using mafft v7.215 with default settings (Kato and Standley, 2013), folded with RNAalifold and inspected in RALEE (Griffiths-Jones, 2005) for H/ACA

box signatures – two hairpins interspaced with H box and terminal ACA box. The rRNA target sites for H/ACA snoRNAs were predicted using RNAsnoop (Tafer et al., 2010). RNA structures were drawn using VARNA applet v3.93 (Darty et al., 2009).

2.4 General Data Analysis.

Principal Component Analysis and other statistical analyses were performed in R using various work packages (Lê et al., 2008). Specific aspects of data analysis and how they were performed are described in the appropriate chapters.

Chapter 3

Production of *hid* Δ mutants of *S. pombe*

3.1 INTRODUCTION

A core aim of this project is to determine the function of the *S. pombe* orthologues of the *HID1* gene from humans. The biological function of the *HID1* gene remains unknown, although it has been implicated as a neuropeptide transport in *C. elegans* (Yu et al., 2011). It is known that *HID1* in humans and other animals is present in the membrane of the Golgi apparatus in the *medial*- and *trans*-face, so it could be implicated in peptide excretion. Interestingly, the *HsHID1* gene has been shown to be lacking or its expression is down-regulated in a number of different cancer types, and, thus, it has been described as a class II tumour-suppressor gene called *Down-regulated in Multiple Cancers-1 (DMC1)* ((Harada et al., 2001). In order to establish the biological role of *HID1/DMC1*, we are using *S. pombe* a model organism. *S. pombe* is particularly attractive for use as a model organism for studying Hid function. *S. pombe* only possess Hid and not Dym, thus Hid function can be studied without the potential complication of functional complementation. In addition, it is unique in that it has three paralogues of *HID1*, that we have termed *hid1* (SPAP27G11.12), *hid2* (SPBP19A11.07c), and *hid3* (SPAC17A5.16). It is possible that some aspects of function may be partitioned amongst the paralogues. I am involved in a multidisciplinary approach based on holistic profiling technologies in order to find the processes in which the Hid proteins function. A critical aspect of this work is to create the mutant strains, in which the proteins are missing to alterations in biological processes from growth responses to gene expression to be monitored.

This chapter describes production of the mutants, whereby each of the *hid* genes has been eliminated using one-step gene disruption methods based on homologous recombination (Baudin et al., 1993; Grallert et al., 1993; Grimm and Kohli, 1988; Rothstein, 1983; 1998). All three *hid* genes were deleted individually as well as certain combinations of double mutants. Replacement of the target *hid* gene with a selectable marker was determined by both PCR-based genotyping and RT-PCR. Initial characterization of the growth properties of the mutants revealed differences to the growth of wild-type for strains lacking *hid2* and *hid3*, but not *hid1*. There appeared to be no phenotypic difference in the mutants regarding cell size or shape, but the Golgi apparatus of the strains lacking *hid3* appeared to be dramatically modified in structure.

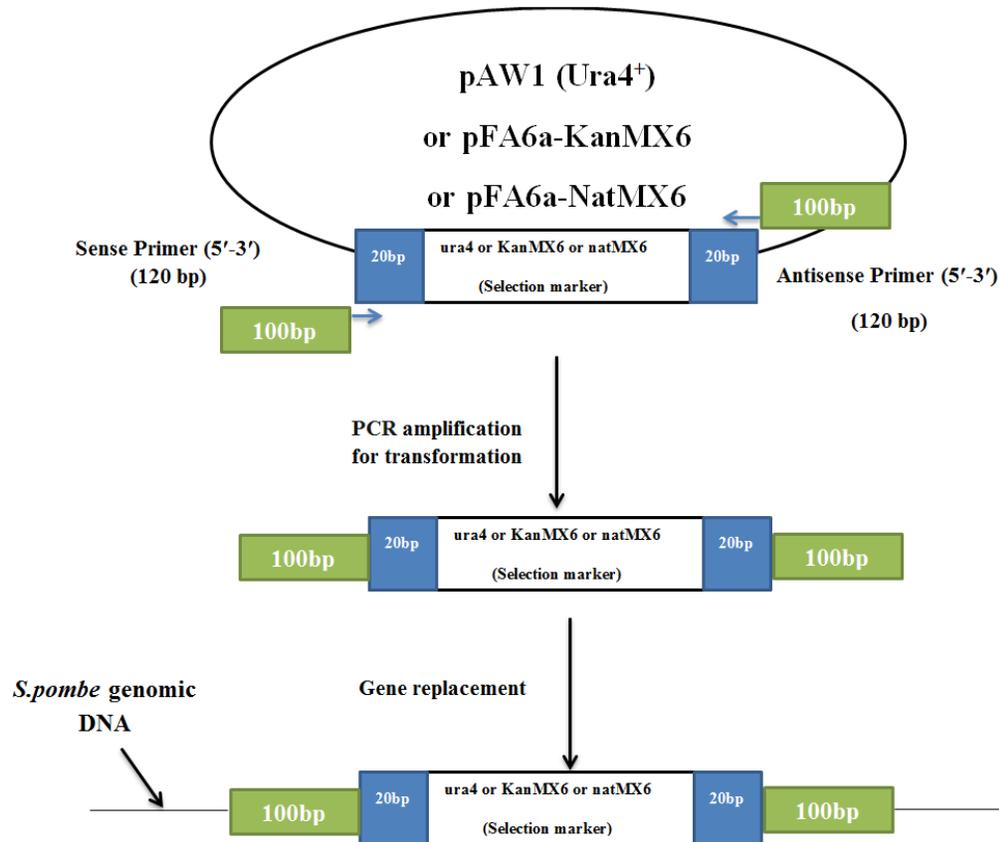


Figure 3.1: Diagram illustrating the process of mutant creation through gene replacement. The plasmids pAW1, pFA6a-KanMX6, and pFA6a-NatMX6 were used as templates for amplification by PCR of DNA fragment containing the appropriate selectable marker to be used for transformation. DNA fragments were produced in mass using 8 x 50 μ l reactions, and the reactions pooled and the PCR products gel-purified prior to transformation. Transformation was carried out as described in Bähler et al. (1998).

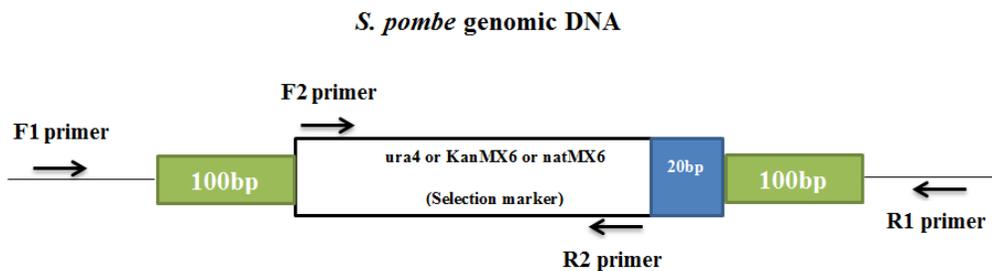


Figure 3.2. Location of primers for confirmation of gene replacement by PCR. Conceptual model showing the location of primers relative to the integration site of the selectable marker gene within the genomic DNA. The external primer pair F1/R1 produces a fragment proportional to the size of the selectable marker gene, whereas primer pair F2/R2 confirms the presence of the selectable marker gene or absence of the target gene. Combinations of primers were used to confirm the presence of appropriate junctions.

3.2 RESULTS

3.2.1 Creation of mutants by gene replacement

Gene deletion strains of *S. pombe* were produced by homologous recombination according to (Bähler et al., 1998). PCR fragments for transformation were generated using 120-mer oligonucleotides that contained 100 bp of sequence homologous to the flanking regions of the target *hid* gene and 20 bp corresponding to the flanking region of one of the selectable marker cassettes in the plasmids pAW1 (*ura4*⁺), pFA6a-KanMX6 or pFA6a-NatMX6. Replacement of a gene by *ura4*⁺ would allow growth in the absence of uracil on minimal medium, since the wild-type strain carried the *ura4-D18* mutation. The other two selectable markers confer resistance to antibiotics, thus selection was possible on standard YES media. The PCR products used for transformation were of sizes ~2.0 kb for the *Ura4*⁺ gene, ~1.6 kb for the aminoglycoside 3'-phosphotransferase (*Kan*^R) gene, and 1.4 kb for the noursethricin acetyl transferase (*Nat*^R) gene. Production of the proper PCR fragments was first verified by agarose gel electrophoresis prior to continuing with product purification and transformation (data not shown). The process of mutant creation through gene replacement is outlined in Figure 3.1.

3.2.2 Verification of gene deletion and replacement by PCR-genotyping and RT-PCR

The success of mutant strain creation by gene replacement was verified both by genotyping the mutant strains and demonstrating the lack of expression of the corresponding transcript.

Genotyping by PCR was conducted on DNA isolated from 1 ml of a standard overnight culture grown in YES medium. Replacement of the target gene by the appropriate selectable marker was confirmed by both detection of the marker gene and loss of the target gene. This was accomplished through a set of 5 primers constructed for each of the target *hid* loci (Figure 3.2). An external primer pair F1/R1 was genomic sequence and it was used to estimate the size of the modified locus containing with the selectable marker gene. Two different pairs of internal PCR primers F2/R2 were used to detect the lack or presence of the marker gene or the endogenous *hid* gene, respectively. The external primer pairs F1/R2 and F2/R1 were used to detect the presence and determine the correct amplicon size of the junction between the marker gene and genomic DNA.

The gene-specific internal primers were also suitable for use to detect transcript by RT-PCR of isolated RNA. Routinely, total RNA was isolated from aliquots of the same cultures used for DNA isolation as described in Chapter 2.

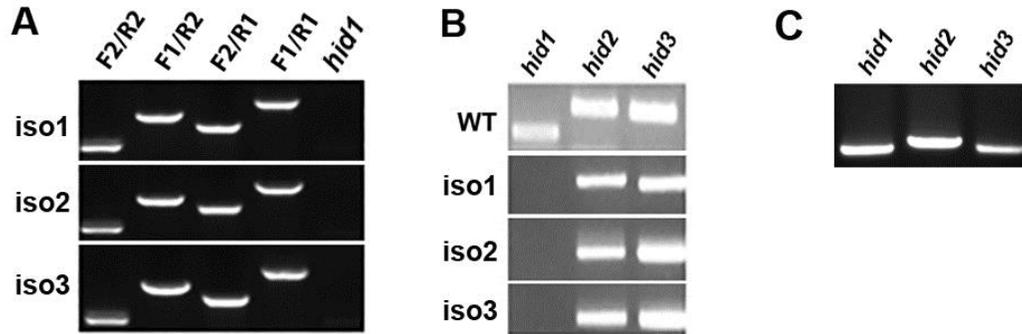


Figure 3.3: Molecular characterization of *hid1Δ::natMX6* strains. All replacement mutant strains were given number designations (iso1, iso2 iso3) according to the order they were selected from plates for subsequent genotyping. (A) Agarose gel showing amplification of sequences at the *hid1* locus. The primer designations are given in Figure 3.2. (B) Agarose gel showing the lack of amplification of *hid1* by RT-PCR. The primers pair used were F2/R2 specific for each gene. The size of all amplified fragments were determined using the Hyperladder™ 1kb (Bioline, BIO-33053) size markers, but the ladders were not included in the images shown. (C) Agarose gel showing the amplification of the individual *hid*⁺ genes from wild-type DNA.

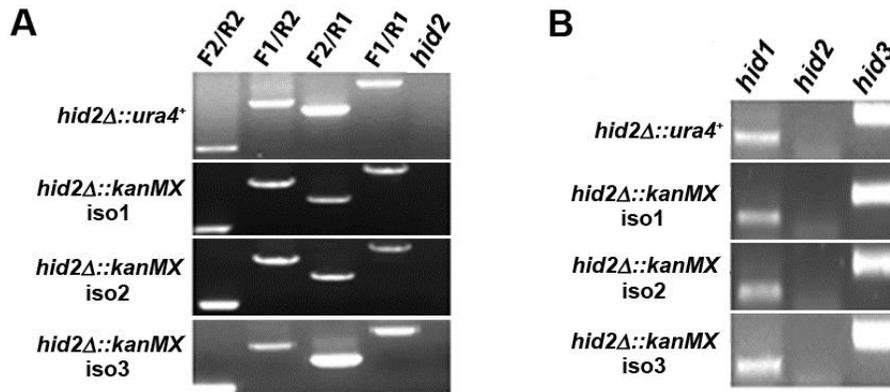


Figure 3.4. Molecular characterization of *hid2Δ::kanMX6* mutant strains. The details for the figure are the same as described in the legend to Figure 3.3, except that they apply to the *hid2Δ* mutant loci. (A) Agarose gel showing amplification of sequences at the *hid2*⁺ locus. (B) Agarose gel showing lack of transcript in *hid2Δ* mutant strains.

3.2.3 Construction of single mutants

3.2.3.1 Confirmation of strains lacking *Hid1*

The *hid1Δ* knockout strains were created using the *NatMX6* cassette of pFA6a-NatMX6. PCR screening for the presence of the *Nat^R* gene suggested three positive strains that were taken on for further molecular analysis. Each strain showed the presence of bands with primers specific for the *Nat^R* gene (Figure 3.3). The primer pairs F2/R2, F1/R2, F2/R1 and F1/R1 would yield fragments of 495, 1378, 1106 and 1989 bp, respectively. Amplification of the *hid1* gene was negative on the same DNA samples, thus showing lack of the *hid1* gene (Figure 3.5A).

The lack of amplification of mRNA by RT-PCR supported the conclusion that gene replacement was successful. *hid2* and *hid3* continued to be expressed in the *hid1Δ* strains indicating that these loci were unaffected. The expected fragment size in the cDNA for *hid1⁺* was 295 bp and those for *hid2⁺* and *hid3⁺* were 448 and 435 bp, respectively. However; the expected fragment size in DNA for *hid1⁺* was 460 bp and those for *hid2⁺* and *hid3⁺* were 531 and 487 bp, respectively (Figure 3.5B).

3.2.3.2 Confirmation of strains lacking *hid2*

The *hid2Δ* knockout strains were created using the heterotrophic marker *ura4⁺* amplified from the plasmid pAW1 (Watson et al., 2008). PCR screening for the presence of *ura4* gave five potential mutant strains that were then analysed. Although each strain no longer required uracil for growth, only one strain showed amplification of sequences for replacement of the gene by *ura4⁺* and this was *hid2Δ::ura4⁺* (Figure 3.4A). The negative strains are discussed below. For *hid2Δ::ura4⁺*, the primer pairs F2/R2, F1/R2, F2/R1 and F1/R1 were expected to yield PCR fragments of 587, 1777, 1494 and 2684 bp, respectively. Construction of *hid2Δ* strains was also done using the *KanMX6* cassette for PCR amplification of transformation product. Three out of seven colonies selected were positive for insertion of the *Kan^R* gene at the *hid2* locus (Figure 3.4A). For *hid2Δ::kanMX6* strains, the primer pairs F2/R2, F1/R2, F2/R1 and F1/R1 were expected to yield PCR fragments of 433, 1651, 1075 and 2293 bp, respectively. Amplification of the *hid2* gene was negative for *hid2Δ::ura4⁺* and the *hid2Δ::kanMX6* strains showing lack of the *hid2* gene.

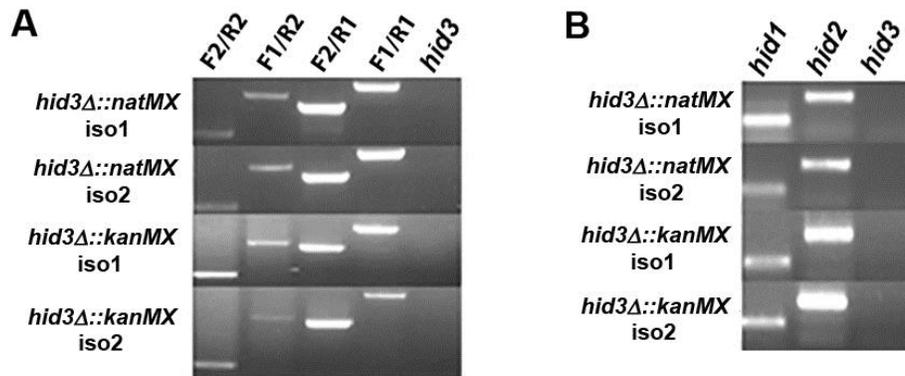


Figure 3.5. Molecular characterization of *hid3Δ* strains. The details for the figure are the same as described in the legend to Figure 3.3, except that they apply to the *hid3* mutant loci. (A) Agarose gel showing amplification of sequences at the *hid3*⁺ locus. (B) Agarose gel showing lack of transcript in *hid3Δ* mutant strains.

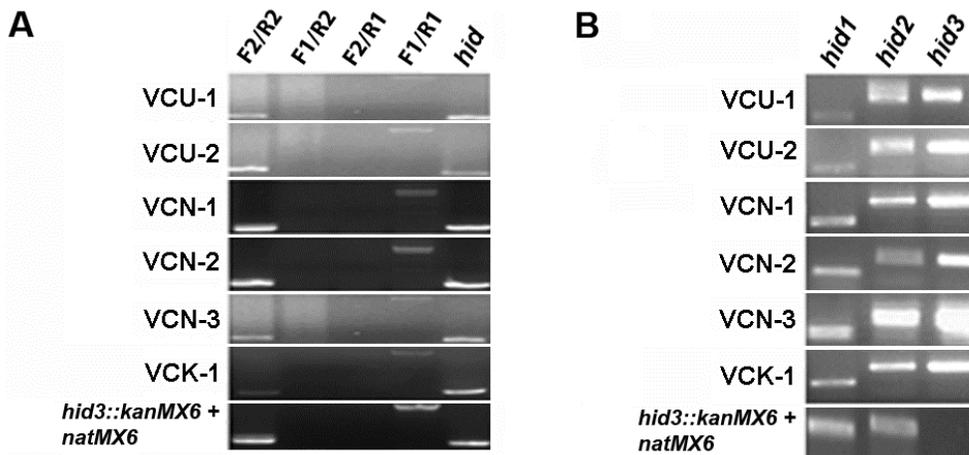


Figure 3.6. Molecular characterization of Vector Control strains. Strains were treated as described in the legend to Figure 3.3. Abbreviations: VC, vector control; N, K and U stand for *Nat*^R, *Kan*^R and *Ura*⁴⁺, respectively. The number represents the order by which the colony was selected. (A) Images of agarose gels showing amplification (or lack of) at mutant loci. Only characterization of the *hid1*⁺ locus is shown for *hid3::kanMX6*⁺ *natMX6*. In the lane *hid*, the fragment shown corresponds to the gene unsuccessfully replaced. (B) Agarose gels showing RT-PCR test for expression of all three *hid* genes in VC strains.

RT-PCR showed that *hid2* transcript was missing in *hid2Δ* mutant strains, but that *hid1* and *hid3* were still expressed (Figure 3.4B).

3.2.3.3. Confirmation of strains lacking *hid3*

With *ura4* replacement initially unsuccessful, *hid3Δ* mutant strains were created using both the *natMX6* and *kanMX6* cassettes (Figure 3.5A). It should be noted that late on in the project, it was possible to get *hid3Δ::ura4⁺* strains, but these were very slow growing and the colonies selected were very small even after 5 days of growth. These mutants have been included in Table 3.1.

For *hid3Δ::natMX6* strains, the primer pairs F2/R2, F1/R2, F2/R1 and F1/R1 were expected to yield PCR fragments of 495, 1293, 1046 and 1844 bp, respectively. For *hid3Δ::kanMX6* strains, the primer pairs F2/R2, F1/R2, F2/R1 and F1/R1 were expected to yield PCR fragments of 433, 1383, 1128 and 2078 bp, respectively. Amplification of the *hid3* gene was negative for all *hid3Δ::natMX6* and *hid3Δ::kanMX6* strains showing that it had been deleted successfully using both types of selectable markers. Again, RT-PCR showed that transcript, this time in *hid3Δ* mutant strains, was missing, but that *hid1* and *hid2* were still being expressed (Figure 3.5B).

3.2.4 Identification of Vector Control strains

Often during the molecular characterization of potential replacement mutants, strains would be observed with the properties of the selectable trait, such as growth without uracil upon transformation with the *ura4* marker or the appropriate antibiotic resistance. Upon analysis by PCR, these lines showed inverse properties to expected mutants. These lines showed PCR fragments using F1/R1 primer pairs corresponding to sizes containing the endogenous gene and that gene-specific primer pairs F2/R2 showed that endogenous genes were still present. However, these lines provided correct size fragments for the marker-specific F2/R2 primer pairs, indicating that the marker gene was present, but fragments corresponding to the junction primer pairs F1/R2 and F2/R1 were not obtained. I obtained VC strains for each of the selectable markers used in the study (Figure 3.6). There were a few instances whereby the transformation to create a second replacement mutation within an already mutant genotype resulted in a mutant expressing a second marker, such as is shown for the *hid3Δ::kanMX6 + natMX6* strain (Figure 3.6). In the case shown, I attempted to replace *hid1* with NatR (see creation of double mutants below).

Each of the VC strains was tested by RT-PCR to check if expression at any of the three *hid* loci was affected during the transformation (Figure 3.6B). Each vector control for the single markers, that is

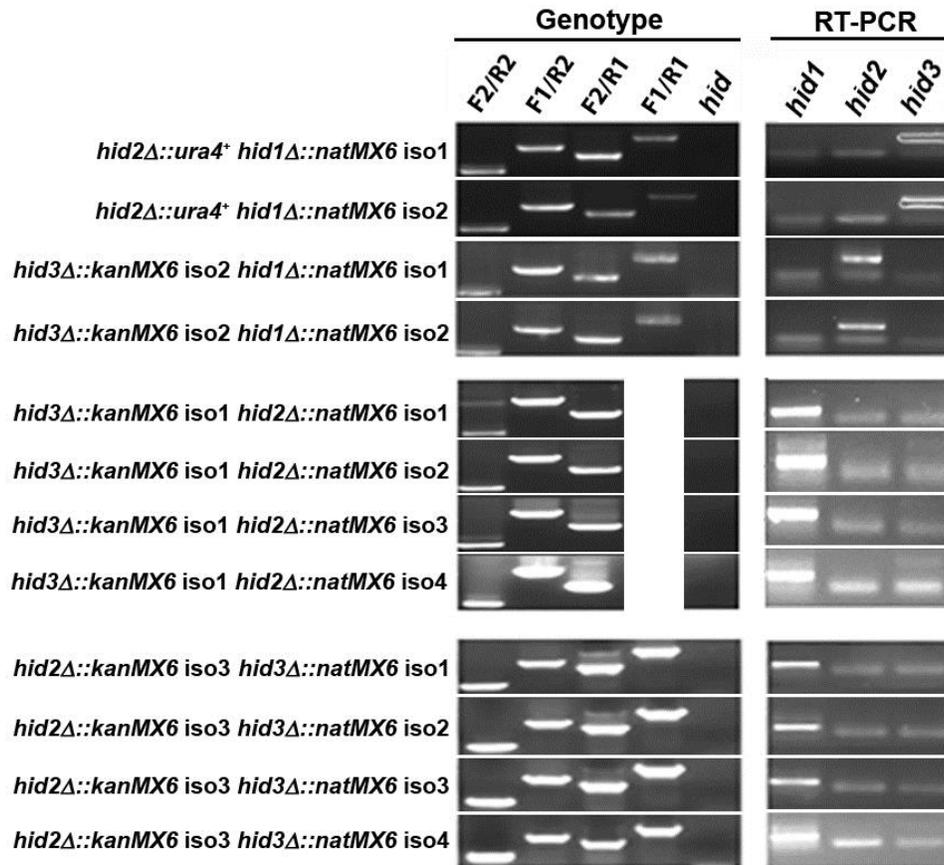


Figure 3.7. Molecular characterization of double mutant strains. Strains were grown as described in Figure 3.3 and DNA and total RNA isolated from the same cultures. Panels A, C and E show PCR characterization of locus of the second replacement. The primer combination F1/R1 was not done for the series *hid3Δ::kanMX6 iso1 hid2Δ::natMX6* mutants. The lane ‘Hid’ shows the presence or absence of the endogenous gene at the second mutant locus. Panels B, D and F show the RT-PCR expression analysis of all three *hid* genes in the double mutants. The fragment sizes are the same as for the single gene-replacement mutants.

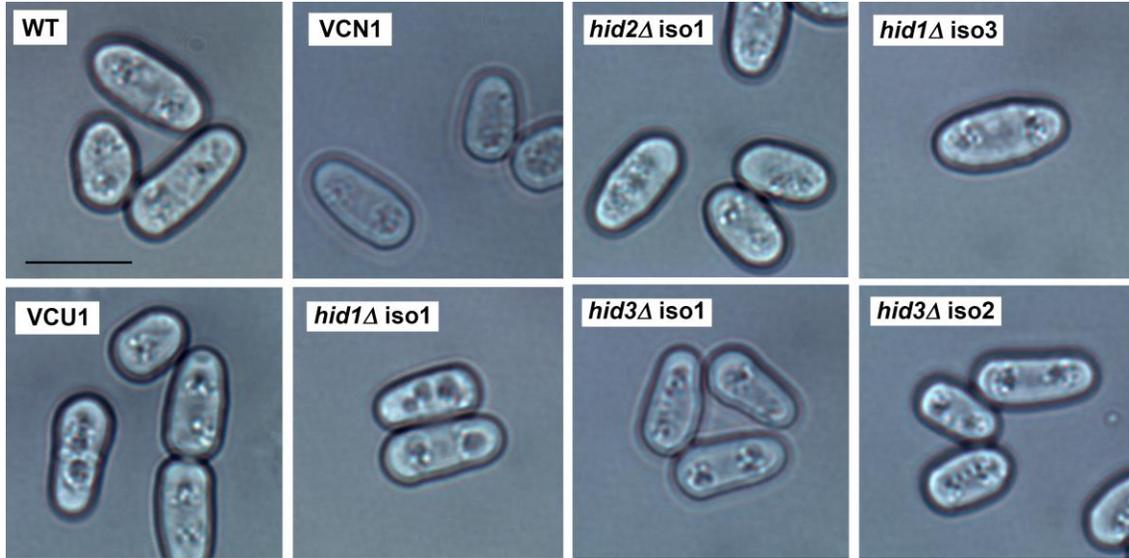


Figure 3.8. Visible phenotype of cells. Mutant strains were grown in duplicate in liquid YES for ~ 16 hours at 30°C with shaking at 200 r.p.m to an approximate cell count of 5×10^7 cells/ml. Images were taken of cells using a Lieca light microscope with a 100x oil-emersion lens. Five images were taken of each strain and one image of ten selected at random. The largest group of cells was selected to show. The scale bar in wild-type image is represents 10 μm .

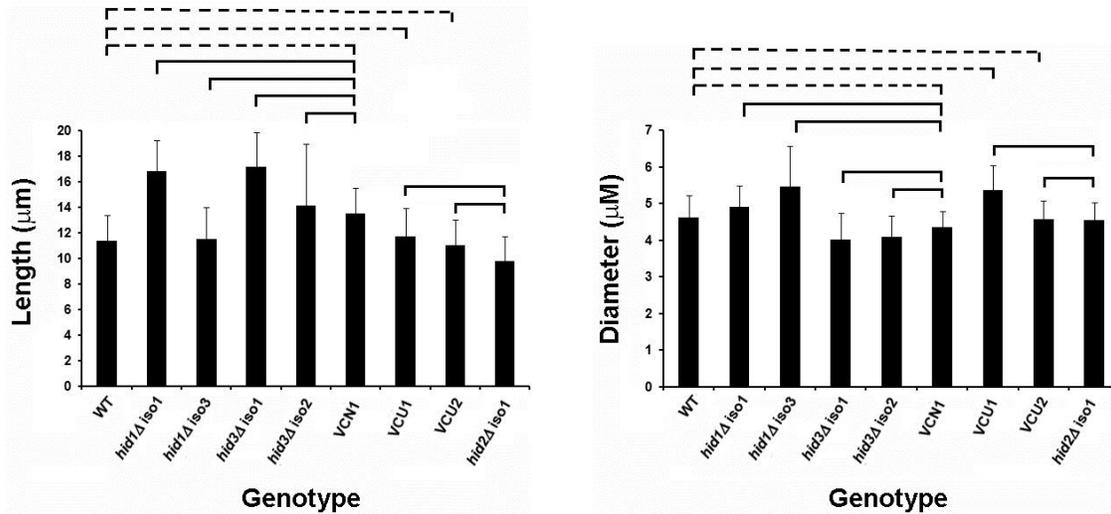


Figure 3.9 Size properties of mutant strains. The mean length (A) and the mean diameter (B) were determined at 40x resolution using a 10 mm ruler at the same magnification. The mean number of pixels for 10 mm was determined from 10 independent images of the ruler. This value was used to set a scale from which the dimensions of cells were determined. All values were determined using ImageJ (Schneider et al., 2012). The solid bars represent the pairs of samples for which statistical analysis were conducted by Student's t-Test (n=30). No still statistical significance is shown.

VCU1, VCN1, VCK1 and so on showed transcript for each of the *hid* genes. As expected, VC strains from double mutant construction showed expression of *hid* genes, except for that corresponding to the mutant background used for the transformation.

3.2.5 Construction of double mutants

The strategy to create double and triple mutants was simply to replace a second *hid* gene by homologous recombination in an already existing mutant background, but using a different selectable marker. This strategy has been shown to work well with *S. pombe* to create multiple mutations (Hentges et al., 2005). Thus, for example, the mutant *hid2Δ::ura4⁺* was transformed with a PCR product derived from the *natMX6* cassette using primers with *hid1* homologous adapters as for the creation of *hid1* replacement mutants. Using antibiotic resistance for selection of the second mutant locus, I successfully isolated multiple strains for each combination of double replacement (Figure 3.7). The each double replacement strain displayed the correct amplification of products at the second locus as observed for the respective single mutant. Gene expression analysis by RT-PCR demonstrated that expression of *hid* came only from the remaining intact locus. Use of *ura4* replacement to create the *hid1Δ hid2Δ hid3Δ* in various double mutant backgrounds was unsuccessful.

3.2.6 Initial physiological studies of gene-replacement mutants

3.2.6.1 Morphology of single mutants

Light microscope images were taken of the mutant strains to see if there were any visible morphological differences. Two *hid1Δ*, two *hid3Δ* and one *hid2Δ* mutants were compared to wild-type and corresponding vector controls at 100x magnification. Nothing regarding cell shape or internal compartmentation was obviously different among any of the strains visualised (Figure 3.8).

It was expected that for cultures in log phase where cells are rapidly dividing, there would be some size differences due to the stage of development, such as newly divided daughter cells or elongated cells just entering mitosis. Therefore, the average cell length and diameter for each strain was determined from the images (Figure 3.9). For the cell size measurements, a second vector control for *ura4⁺* was included. The mean cell length varied between 9.5 and 17 μm (Figure 3.9A) and the mean cell diameter varied between 4 and 5.5 μm (Figure 3.9B).

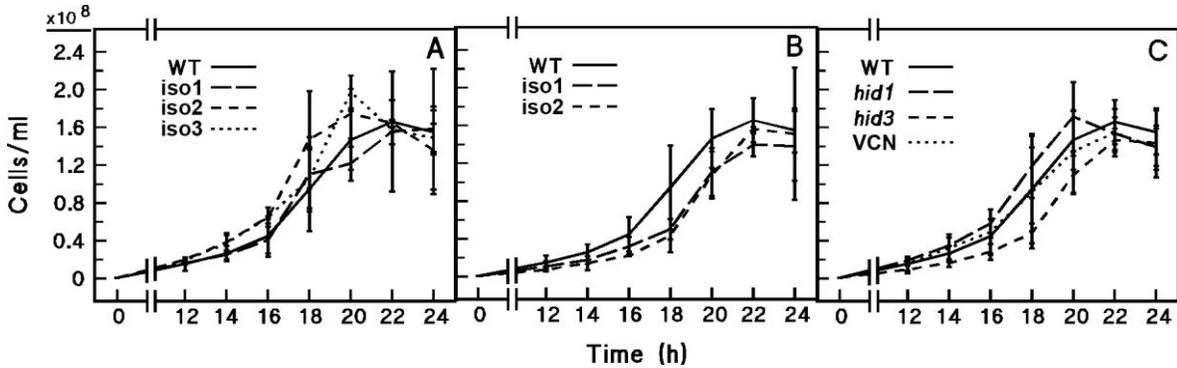


Figure 3.10. Growth curves for *hid1Δ::natMX6* and *hid3Δ::natMX6* strains. Cells were grown in liquid culture with a starting cell density of 5×10^5 cells/ml. Aliquots were taken every 2 hours and the cells counted using a hemocytometer under a microscope. (A) The *hid1Δ::natMX6* strains. (B) The *hid3Δ::natMX6* strains. (C) Representation of average growth curves calculated from the data shown in (A) and (B) with the vector control strains added. In (A) and (B) the error bars represent the variation from duplicate cultures for each strain. In (C), the *hid1* profile was produced from those of *hid1Δ* iso1 and *hid1Δ* iso3 shown in A with error bars representing the sd of $n=4$. Error bars for VCN represent high and low values obtained from two independent vector control strains.

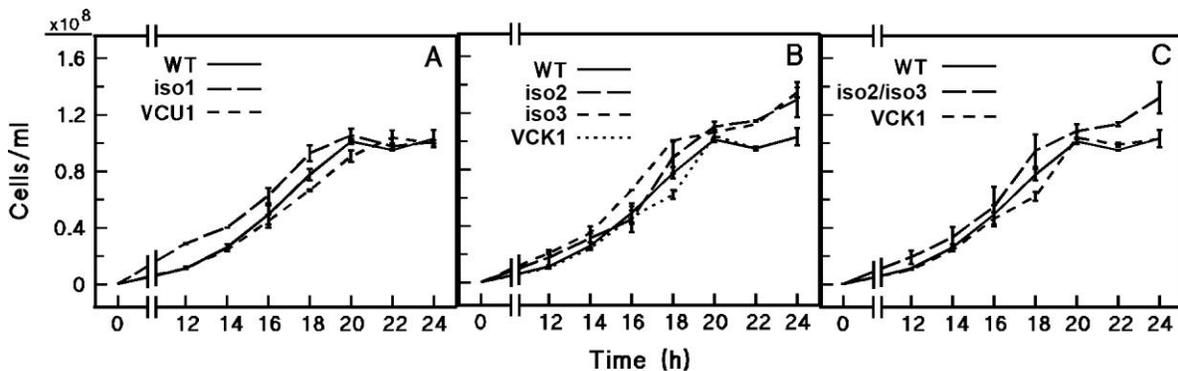


Figure 3.11. Growth profiles of *hid2Δ* mutant strains. Cells were grown as described in Figure 3.10. Cells were counted automatically using a Cellometer mini™ (Nexcelon Inc.). (A) *hid2Δ::ura4⁺* strain and VC containing *ura4⁺* (VCU). (B) *hid2Δ::kanMX6* strains and VC containing *kanMX6*. Each strain was done in duplicate (C) Showing the average (line) \pm sd (error bars) of $n=4$ of *hid2Δ* iso2 and *hid2Δ* iso3 strains. The data for wild-type and VCK1 strains is the same as that shown in (B).

Although there appear to be substantial differences between strains (see Figure 3.9A *hid1Δ::natMX6 iso1* vs. wild-type or *hid3Δ::natMX6 iso1* vs. wild-type), these can only be considered as trends, since there were no significant differences between mutant and corresponding vector control strains. In addition, the differences were not consistent for the similar mutants. For example, *hid1Δ::natMX6 iso3* or *hid1Δ::natMX6 iso1* showed no statistical difference to wild-type or corresponding vector control strain. The lack of difference in diameter among strains was more apparent than for length (Figure 3.9B).

3.2.6.2 Characterisation of proliferate properties.

All mutants isolated were studied for alterations in proliferation rate. Strains were grown under standard conditions of YES medium at 30°C with shaking at 200 r.p.m. proliferation was routinely measured as the number of cells per unit volume at selected times. Cell counts were initially done using a hemocytometer, but then progressed to automated counting in order to increase reproducibility of counts. The growth curves for *hid1Δ::natMX6* and *hid3Δ::natMX6* mutants compared to unmodified wild-type and the vector control strains are shown in Figure 3.10.

The *hid1Δ::natMX6* strains showed similar growth profiles to the unmodified wild-type (Figure 3.10A) or the VCN (Figure 3.10N). Although *hid1Δ::natMX6 iso2* appeared to exhibit a growth rate greater to wild-type, this was not seen for either *hid1Δ::natMX6 iso1* or *hid1Δ::natMX6 iso3*. In contrast, the *hid3Δ::natMX6* mutants exhibited consistent delays in growth compared to wild-type (Figure 3.10B) and vector control strains (Figure 3.10C), although the overall shape of the profiles were similar to both controls. The growth curves for the *hid2Δ* mutants are shown in Figure 3.11.

Each *hid2Δ* strains appeared to proliferate faster than the corresponding wild-type and corresponding vector only control strain. The *hid2Δ::kanMX6* also seemed to attain an overall greater cell density after 24 hours, when the growth curves had leveled off (Figure 3.11B), but this was not seen for the *hid2Δ::ura4⁺* mutant (Figure 3.11A). It should be noted that the cell densities of the stationary phases for the two the cultures shown in Figure 3.10 and Figure 3.11 were approximately 1.6×10^8 and 1.0×10^8 cells/ml, respectively. This difference can be given to use of the two different means of counting the cells. With visual counting of cells newly divided cells or two cells in a v-shape could be counted as separate cells. Typically, these types of cells were counted as single cells using the automated system, thus the total cell counts were usually less. The positive aspect was that the automated system was superior in reproducibility of counts.

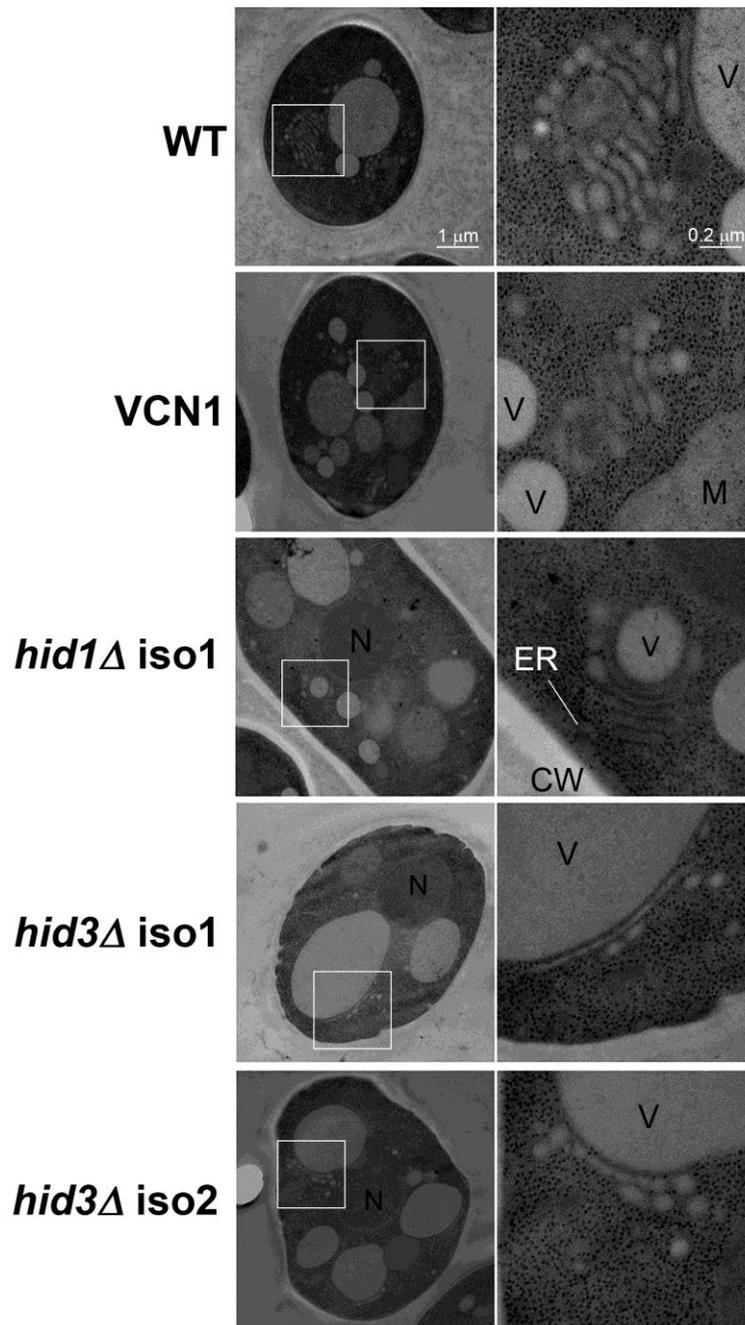


Figure 3.12. Transmission electron microscopy of thin layer sections of *S. pombe* cells. Strains were cultured under standard conditions to a cell count of 5×10^7 cells/ml. The cells were freeze substituted with 1% OsO_4 in acetone. The cells were impregnated with EPON resin were cut using a Leica UC7 microtome. Slices were treated with lead citrate/uranylacetate for contrast and were visualised on a Tecnai Spirit 2 FEI operating at 120kV.

The proliferate properties of the double mutants were also determined, but exhibited a phenotype of one of the single mutants. The *hid2Δ::ura4⁺ hid1::natMX6 iso1* was like *hid2Δ::ura4⁺* and the *hid3Δ::kanMX6 iso2 hid1Δ::natMX6* was like the *hid3Δ::kanMX6 iso2* and *hid3Δ::natMX6* single mutants. Interestingly, the *hid2Δ hid3Δ* reciprocal mutants were like the *hid3Δ::kanMX6 iso2* and *hid3Δ::natMX6* single mutants too.

3.2.6.3 Preliminary ultrastructural characterisation of the mutants

Wild-type, VCN, *hid1Δ*, and *hid3Δ* were investigated for any cellular abnormalities, particularly to the Golgi structure. Wild-type, VCN and *hid1Δ* appeared to have a normal subcellular physiological appearance when compared to other TEM studies on *S. pombe* (Osumi, 2012). As expected, the shape of the Golgi differed from image to image due to the orientation of the cell in the matrix during the freeze substitution and at where through the cell the slice was made. Nevertheless, the Golgi was clearly visible in a number of cells and was ribbon shaped with multiple cisternae for wild-type, VCN and *hid1Δ* (Figure 3.12). The Golgi appeared to be closely related to clear round bodies that are likely to be vacuoles. A similar Golgi structure was not visible in cells of *hid3Δ*. The main structure I could find that resembled a Golgi was one elongated ribbon with what appeared to be associated vesicles. Even if this was not the Golgi, then the Golgi had been completely disrupted in *hid3Δ*.

3.3 DISCUSSION

Conducting a study on the biological function of the Hids from *S. pombe* require the creation of different mutant strains. Screening for cellular and biochemical function using post-genomic technologies of mutants has potential problems. Expression of heterologous marker proteins may itself have effects on metabolism and protein and RNA levels. Thus, in order to compare the effects of missing each Hid, the marker selection used would have to be the same. In addition, it would be necessary to have independent strains for each locus and selectable marker in order to confirm that any biochemical phenotype would be due to the missing protein. However, in order to produce the double and even triple mutants, each *hid* would need to be replaced by a different selectable marker. To compare mutants to wild-type would require insertion of the marker at a locus that would not have an effect on phenotype. Therefore, my initial intention was to create individual mutants for each *hid* using each of the three selectable markers, and if necessary to create strains where markers were inserted at independent loci.

It was largely unsuccessful to obtain mutant strains using replacement with the heterotrophic marker *ura4*. It was simply fortunate that I was able to obtain the *hid2Δ::ura4⁺* strain. Repetition of the transformation

Table 3.1. Mutant strains created during the course of the study, the selectable marker and the number of independent isolates. All strains were created from the parental genotype *h⁻ ade6-M26 ura4-D18 leu1-32* (BP90, McFarlane Collection).

Genotype	Selection	number of isolates
<i>hid1Δ::natMX6</i>	Nat ^R	3
<i>hid2Δ::ura4⁺</i>	Ura4 ⁺	1
<i>hid2Δ::kanMX6</i>	Kan ^R	3
<i>hid3Δ::natMX6</i>	Nat ^R	2
<i>hid3Δ::kanMX6</i>	Kan ^R	2
<i>hid3Δ::ura4⁺</i>	Ura4 ⁺	2
<i>hid2Δ::ura4⁺ hid1Δ::natMX6</i>	Ura4 ⁺ /Nat ^R	2
<i>hid3Δ::kanMX6 hid1Δ::natMX6</i>	Kan ^R /Nat ^R	2
<i>hid3Δ::kanMX6 hid2Δ::natMX6</i>	Kan ^R /Nat ^R	4
<i>hid2Δ::kanMX6 hid3Δ::natMX6</i>	Kan ^R /Nat ^R	4
<i>ura4⁺</i>	Ura4 ⁺	2
<i>natMX6</i>	Nat ^R	3
<i>kanMX6</i>	Kan ^R	1
<i>hid3Δ::kanMX6 + natMX6</i>	Kan ^R /Nat ^R	1

protocol several times with *ura4* did not give positive mutants. It is clear now why it was difficult to get the *hid* mutants and, particularly the *hid3* mutant strains using *ura4*. Selection on for *ura4* integration required growth on minimal medium without uracil. It has been shown that *hid3Δ* strains do not grow on minimal medium (Alasmari, 2015), thus isolation of these mutants would not be possible. It is possible that the lack of Hid1 or Hid2 may also have metabolic effects that would make isolation of these mutants difficult using heterotrophic selection. Selection using resistance to the antibiotics G418 and noursethricin was more efficient and yielded mutants for each of the three lines. Although we did experience problems with consistent use of the pFA6a-KanMX plasmid to successfully amplify a fragment for transformation for the creation of double mutants. Using the gene replacement protocol, I was able to obtain a sufficient variety and number of single and double mutant strains to conduct basic physiological and biochemical analyses on Hid function. The sensitivity of *hid3Δ* strains It was also fortunate than in a number of cases, I obtained strains where the marker gene had been inserted, but not at the desired locus. These strains would be necessary for determining the effects of marker expression on biochemical processes, and could be used to indirectly compare mutant strains with untransformed wild-type. A summary of the mutant strains produced in this study is given in Table 3.1.

An important addition to the molecular characterization of the mutants was to determine gene expression from each locus in the mutant strains. This was less necessary for showing lack of mRNA production for a replaced gene, but it was very important to show that expression at the other two loci was not greatly affected. Therefore, the conclusions we can make on a mutant pertain to that missing protein and not to indirect effects of modified expression at one of the other loci. Even though I did not include *hid2Δ* strains in the post-genomic studies, it was necessary to exclude altered expression at this locus in the other mutants. It is interesting to note that there seems to be a trend in the relative levels of transcripts for the three *hid* genes. Transcript amount appears to increase in the order *hid1*, *hid2* and *hid3*, which is usually the most abundant. Although this was not quantitative, the presence of transcript for each gene was determined exactly the same way, at the same time and on the same total RNA sample, either wild-type , VC or mutants. We cannot exclude the possibility that differences in primer affinity, etc. could be responsible for the differences, but these can be verified using RT-PCR and the RNAseq data (see Chapter 4).

An initial characterization of the mutants showed viable cell populations with no evidence for morphological differences from wild-type or VC strains. These observations were consistent with those previously reported from global phenotypic screens (Hayles et al., 2013). Furthermore, Kouranti et al. (2010) did not report any growth defect for strains lacking Hid3 (Ftp105).

Although the mutant strains were viable, the *hid2Δ* and *hid3Δ* mutants showed differences in growth to both wild-type and VC strains, whereas *hid1Δ* mutants did not. It is likely that the lack of Hid3 affects the function of the Golgi Apparatus in a way that negatively affects cell division, but not cell growth. Cell growth may be delayed, but the cells reach normal size and have a normal morphology. In contrast, the TEM shows a dramatic change in the Golgi structure. The *hid3* orthologue HID1 of both humans and *C. elegans* has been localized to the *medial*- and *trans*-face of the Golgi apparatus and *trans*-Golgi vesicular network (Wang et al., 2011). Therefore, it seems reasonable to conjecture that structure observed is the remaining *cis*-Golgi and that the *medial*- and *trans*-Golgi have been disrupted. Interestingly, this is similar to the effect of adding the anterograde protein translocation inhibitor brefeldin A (Pantazopoulou and Peñalva, 2009)(Ritzenthaler et al., 2002). It is clear that the effects of lacking Hid3 on the Golgi structure need further characterization, but I can conclude that such disruption of the Golgi apparatus would have profound effects on protein transport and secretion and protein localization. It is known that the lack of Hid3 causes mis-localisation of the deubiquitinase Ubp5 from the Golgi apparatus to the cytosol (Kouranti et al., 2010), but it is likely that the transport or localization of other proteins besides Ubp5 is affected. The positive effect of loss of *hid2* was unexpected and the cause of this remains unclear.

In conclusion, I show that normal cell function is altered with removing either *hid2* or *hid3*, but the mechanisms by which these proteins maintain normal cell function remain unknown. The collection of mutants reported in this chapter will be used to conduct functional genomic studies on gene and protein expression in order to determine their cellular functions.

Chapter 4

Transcriptional properties of the *hid1*Δ and *hid3*Δ mutants of *S. pombe*

4.1. INTRODUCTION

This chapter describes part of a multidisciplinary approach to understand the function of the Hid proteins making use of the genetic mutants described in Chapter 3. The underlying concept of studying gene expression is that either gene expression changes resulting from mutation of an important gene cause an altered phenotype or growth patterns, or that mutations affecting growth or phenotype will indirectly and subsequently alter gene expression patterns. The benefit of a transcriptomics approach is that differentiating between these two possibilities might be possible whereas studying the expression of single or a few genes does not give enough information. In addition, it might be possible to uncover specific aspects of biological function if specific groups or types of related genes change in similar ways. For this study, gene expression was done by the measurement of transcript levels of as many genes as possible using the powerful NGS technology of quantitative transcriptomics called RNAseq.

In the past two decades, there have arisen two primary techniques to determine and identify and quantify the global transcriptome of cells, tissues and organisms, Serial Analysis of Gene Expression (SAGE) and microarrays (Ye et al., 2002). Recently, as a development from high-throughput DNA sequencing using NGS technologies, such as pyrophosphate (454) or sequencing by DNA synthesis (Illumina), these technologies were applied to the sequencing of cDNA libraries in a quantitative manner call RNAseq (Wang et al., 2009). In brief, the process of RNAseq entails converting isolated mRNA into short strands of cDNA in libraries that can be sequenced using linked adaptors with specific primers and a means of detecting the addition of bases to a growing nucleic acid chain in a sequencing reaction. In all RNA-seq technologies, the outputs are millions of short sequences that are quantitatively related to the amount of transcript present in the sample. These amounts of “reads” can be compared between samples to quantify differential expression of genes similar to the way SAGE was done.

There are now a lot of bioinformatic tools and pipelines used to analyse RNA-Seq data, but the common ones are DESeq (Anders and Huber, 2010), DESeq² (Love et al., 2014), Limma (Smyth, 2004), edgeR (Robinson et al., 2010) and Cufflinks (Trapnell et al., 2012). The most commonly used bioinformatic tool

to analyse RNASeq data is the Cufflinks pipeline, which was used to analyse the RNASeq data presented in this chapter. There are three computational processes that RNAseq data must go through to provide data on differential gene expression (DE), alignment of the reads to a reference genome if there is one, assembly of reads into a transcriptome comprised of individual genes, and calculation of DE. All of these can be analyzed through two tools were development, which are TopHat (Trapnell et al., 2009) and Cufflinks (Trapnell et al., 2010). These tools are essentially a sequential series of programmes designed to accomplish the three tasks mentioned above. TopHat aligns the reads for each individual RNASeq sample to the reference genome called mapping. In Cufflinks, the aligned reads for a single sample are merged into transcriptome file. In order to compare gene expression, a set of common transcripts must be identified and this comes through the programme Cuffmerge. The programme Cuffquant then takes the number of reads for each sample and quantifies them according to the common transcripts. At this point, the quantified transcripts for each sample can be statistically normalized for multivariate statistics, such as PCA, or it can be fed into Cuffdiff to determine DE genes between each set replicate samples. The Cufflinks pipeline has been used to report many high-resolution transcriptome studies (Graveley et al., 2011; Lister et al., 2011; Mizuno et al., 2010; Twine et al., 2011). Once the group of DE genes is obtained, they can be studied for biological function like other transcriptomic technologies, such as through individual inspection of genes or through groups of gene using Gene Ontologies (Harris et al., 2004).

The work of Chapter 3 showed that strains of *S. pombe* lacking *hid3* grow more slowly and look like they have a structurally disrupted Golgi apparatus. The effects of missing Hid3 would start once the daughter cells divide, if not before, and remain for the duration of cell life. This would represent a prolonged alteration over the entire population, and so the effects would likely show up in a change in gene expression to counter the situation. I employed the transcriptomics approach to determine how *hid3Δ* cells cope with this situation by changing their gene expression. By being able to compare gene expression with both negative controls expressing the marker gene and with *hid1Δ*, which does not show growth defects, I was able to determine that *hid3Δ* cells appear chronically stressed. This results in the induction of transcription factors (TFs) and plasma membrane protein genes to try to alleviate or cope with the stress. The alteration in gene expression indicates that slow growth may be due to a partial quiescent cell state brought on by stress.

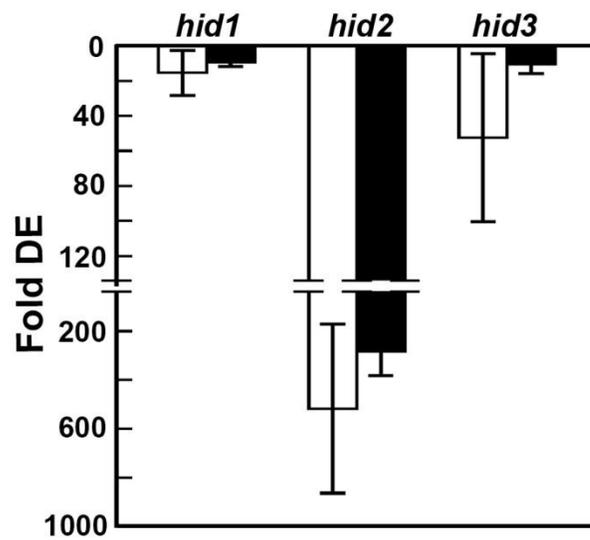


Figure 4.1. Relative transcript levels of *hid* genes in wild-type (□) and VCN (■). Transcript levels were determined by RT-qPCR on a Bio-rad CFX™ real-time PCR machine using Sybr green as the detection dye. The baseline gene was *act1*. For each *hid* gene in each individual sample the differential expression (Fold DE) was calculated relative to *act1*, i.e. $2^{(Ct(hid)-Ct(act1))}$. The values shown are the means \pm sds (n=3) determined across the biological replicates. For any of the three genes, there was no significant difference between wild-type and VCN as analysed by Students t-test.

4.2 RESULTS

4.2.1 Evaluation of *HsHID1* as a Tumour Suppressor gene

HsHID1 is certainly not a driver of tumour formation as it has a very low mutation frequency in known cancers (www.intogen.org; (Gundem et al., 2010). However, this would not be expected for a Tumour suppressor, which only shows a decrease in expression in cancers. The data base Oncomine Gene Browser from Life Sciences Technologies (www.oncomine.org) was searched for the relative expression of *HID1* in cancer vs. normal tissues for under expression in cancer samples. The mRNA search encompassed 83 datasets with 11323 samples using a P-value cutoff of 1×10^{-4} . In the corresponding tissue types from which the cell lines were derived from the Harada et al (2001) study, there is very little indication of down-regulation in tumours compared to the normal tissue, except for some gastric cancers. Even for liver cancers there was no indication of under-expression. In contrast, there was significant over-expression of *HID1* across the range of cancerous tissues, except for the gastric cancers, and, notably, brain and nervous system cancers. For testis and uterine cancers, where there was no indication of over-expression, there was evidence of some specific-tissues under-expression of *HID1*. Interestingly, cell lines showed higher degrees of under-expression of *HID1* than corresponding tissue samples, a good example of which was the melanomas.

4.2.2 Comparative expression of *hid* genes using RT-qPCR

The wild-type and VCN RNA samples to be used for quantitative gene expression by RNAseq were first evaluated for the relative expression of the *hid* genes themselves using RT-qPCR. The primers for RT-qPCR were designed specifically using the programme QuantPrime (Arvidsson et al., 2008), which allowed an annealing temperature of 60°C to be used in order to increase the quality of the PCR amplification. Since the expression of the *hid* genes was being compared within the same biological samples, the expression is given relative to the baseline gene *act1*. Since *act1* was always present in greater quantity than any of the *hid* genes, the values were expressed as Fold-DE decreasing (Figure 4.1).

The RT-qPCR showed that *hid1* and *hid3* were expressed at approximately the same level while *hid2* was much less expressed. At first, this appeared to contradict the expectation from the mutant analysis. However, on inspection of the fluorescence profiles from the qPCR, it was clear that although *hid1* and *hid3* crossed the fluorescence threshold at approximately the same Ct value, the total amount of fluorescence by the end of cycling was much greater for *hid3* and *hid2* than for *hid1*. It is unlikely that differences in primer concentration were the cause, since mutant RT-PCR and RT-qPCR were conducted with different sets of primers. It may be possible that *hid1* is simply less efficient in than the others and

Table 4.1. Genes correlated in expression to each of the *hid* genes. The coefficients corresponding to the 10 most highly positive and highly negative correlated genes are shown as determined by Pancaldi et al. (2010). The coefficients have been ranked in order of highest to lowest value for both positively and negatively correlated genes. The designations given correspond to the *hid* gene, the number assigned to the primer assigned and the direction of correlation, either P (positive) or N (negative).

Gene name	Description	r	qPCR
<i>hid1</i>, positively correlated			
<i>mug105</i>	Meiotically up-regulated gene 105 protein	0.75	
<i>mak2</i>	Peroxide stress-activated histidine kinase mak2	0.73	h1_1P
<i>mim1</i>	Mitochondrial import protein 1	0.73	
<i>rfp1</i>	SUMO-targeted ubiquitin-protein ligase subunit	0.71	
<i>pof11</i>	F-box/WD repeat-containing protein pof11	0.71	h1_4P
<i>SPAC343.06c</i>	Phospholipid scramblase family protein C343.06c	0.70	
<i>phf2</i>	SWM histone demethylase complex subunit phf2	0.70	h1_3P
<i>SPCP20C8.01C</i>	UPF0612 protein	0.69	
<i>SPPC569.01C</i>	UPF0612 protein	0.69	
<i>SPAC977.03</i>	Uncharacterized methyltransferase	0.69	
<i>sec16</i>	COPII coat assembly protein	0.69	h1_5P
<i>mug121</i>	Meiotically up-regulated gene 121 protein	0.68	h1_2P
<i>hid1</i>, negatively correlated			
<i>plp1</i>	Thioredoxin domain-containing protein plp1	0.69	
<i>nat1</i>	N-terminal acetyltransferase A complex subunit nat1	0.57	h1_3N
<i>SPAC1F8.07C</i>	Probable pyruvate decarboxylase C1F8.07c	0.56	
<i>pfk1</i>	6-phosphofructokinase	0.54	h1_1N
<i>SPAPB15E9.01C</i>	Putative GPI-anchored protein PB15E9.01c	0.54	
<i>pma1</i>	Plasma membrane ATPase 1	0.53	
<i>cct5</i>	T-complex protein 1 subunit epsilon	0.53	
<i>rim1</i>	Single-stranded DNA-binding protein rim1	0.53	
<i>spn4</i>	Septin homolog spn4	0.52	
<i>met9</i>	Methylenetetrahydrofolate reductase 1	0.52	
<i>sec232</i>	Protein transport <i>sec23-2</i>	0.51	h1_2N
<i>hid2</i>, positively correlated			
<i>plp1</i>	Thioredoxin domain-containing protein plp1	0.83	
<i>wtf13</i>	Uncharacterized protein wtf13	0.67	
<i>alg2</i>	Alpha-1.3/1.6-mannosyltransferase complex subunit	0.65	
<i>mmm1</i>	Maintenance of mitochondrial morphology protein 1"	0.64	
<i>dsc4</i>	DSC E3 ubiquitin ligase complex subunit 4	0.64	h2_5P
<i>str2</i>	Siderophore iron transporter 2	0.63	
<i>wtf18</i>	Uncharacterized protein	0.63	
<i>SPAC57A10.08C</i>	Abhydrolase domain-containing protein	0.62	
<i>atg7</i>	Ubiquitin-like modifier-activating enzyme	0.62	h2_4P
<i>SPBC543.05C</i>	Putative transporter	0.62	
<i>pdh1</i>	DUF1751 family protein	0.62	h2_3P
<i>pof4</i>	elongin-A, F-box protein Pof4 (predicted)	0.60	h2_2P
<i>atg4</i>	Probable cysteine protease	0.60	h2_1P
<i>hid2</i>, negatively correlated			
<i>SPBP19A11.07C</i>	Uncharacterized methyltransferase	0.58	
<i>gmf1</i>	Actin-depolymerizing factor gmf1	0.46	h2_1N
<i>rpl25b</i>	60S ribosomal protein L25-B	0.40	
<i>rpc10</i>	DNA-directed RNA polymerases I, II and III subunit	0.40	h2_4N
<i>rps24a</i>	40S ribosomal protein S24-A	0.40	
<i>tef1b</i>	Elongation factor 1-alpha-B	0.40	h2_6N
<i>lsm7</i>	U6 snRNA-associated Sm-like protein LSm7	0.36	h2_5N
<i>mrp119</i>	54S ribosomal protein L19. mitochondrial	0.36	
<i>esf1</i>	Pre-rRNA-processing protein esf1	0.35	
<i>gua1</i>	Inosine-5'-monophosphate dehydrogenase	0.35	
<i>aru1</i>	Arginase	0.33	h2_3N
<i>fkbp39</i>	FKBP-type peptidyl-prolyl cis-trans isomerase	0.32	h2_2N
<i>hid3</i>, positively correlated			
<i>alg8</i>	Dolichyl pyrophosphate Glc1Man9GlcNAc2		
alpha-1,3-glucosyltransferase		0.71	h3_1P

<i>lcb2</i>	Serine palmitoyltransferase 2	0.69	h3_5P
<i>bub1</i>	Checkpoint serine/threonine-protein kinase	0.66	h3_9P
<i>ubp7</i>	Probable ubiquitin carboxyl-terminal hydrolase 7	0.66	h3_7P
<i>mon2</i>	Protein MON2 homolog	0.66	h3_2P
<i>hhp1</i>	Casein kinase I homolog hhp1	0.65	
<i>ppk29</i>	Serine/threonine-protein kinase ppk29	0.64	
<i>ptr3</i>	Ubiquitin-activating enzyme E1 1	0.64	h3_3P
<i>pdf1</i>	Palmitoyl-protein thioesterase-dolichyl pyrophosphate phosphatase fusion 1	0.64	
<i>SPBPJ4664.04</i>	Putative coatomer subunit alpha	0.64	
<i>glc2</i>	Glucosidase 2 subunit alpha	0.63	h3_4P
<i>sec24</i>	Protein transport protein	0.63	h3_6P
<i>cta4</i>	Cation-transporting ATPase 4	0.63	h3_8P
<i>cut15</i>	Importin subunit alpha-1	0.62	h3_10P
<i>ubp5</i>	Ubiquitin carboxyl-terminal hydrolase	0.53	h3_11P
			h3_12P
<i>hid3</i>, negatively correlated			
<i>sgf11</i>	SAGA-associated factor 11	0.53	h3_1N
<i>cwf29</i>	U2 snRNP component ist3	0.52	h3_2N
<i>SPBC713.09</i>	Uncharacterized protein	0.50	
<i>cox13</i>	Cytochrome c oxidase subunit 6A, mitochondrial	0.50	
<i>SPAC1805.02C</i>	Probable electron transfer flavoprotein subunit	0.50	
<i>atp7</i>	ATP synthase subunit d, mitochondrial	0.49	
<i>SPBC8D2.12C</i>	Probable transcriptional regulatory protein	0.49	
<i>hnt3</i>	Aprataxin-like protein	0.47	
<i>qcr10</i>	Cytochrome b-c1 complex subunit 10	0.47	
<i>SPBC1773.01</i>	Uncharacterized WD repeat-containing protein	0.46	
<i>sft1</i>	Protein transport protein	0.43	h3-2N

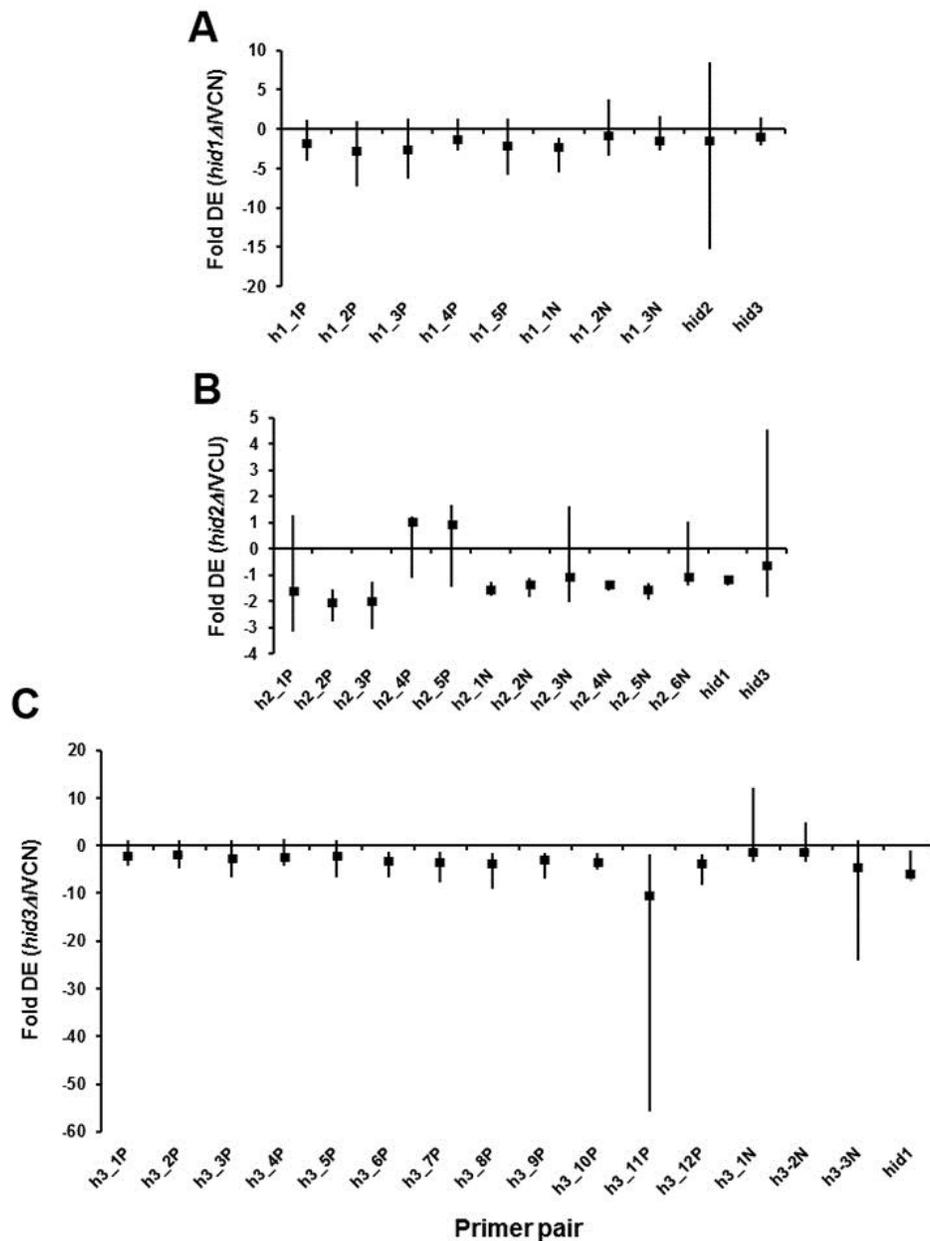


Figure 4.2. Differential expression of genes highly-correlated to the hid genes. The Fold DE of a gene in a mutant genotype is given relative to the appropriate VC marker genotype: VCN for the *hid1Δ:matMX6* mutant (A), VCU for the *hid2Δ:ura4⁺* mutant (B) and VCN for the *hid3Δ:matMX6* (C). The Fold DE was calculated using the $2\Delta\Delta\text{Ct}$ method (Livak and Schmittgen, 2001). Each primer was tested 4 times in each of the three biological replicate strains of a mutant genotype. The mean Ct value for a biological replicate was used to calculate the mean and standard deviation of among the biological replicates. The gene *act1* was used as reference control to calculate ΔCt and the $\Delta\Delta\text{Ct}$ was calculated between mutant and VC genotypes. The black squares represent the mean of the $2\Delta\Delta\text{Ct}$ and the lines the range based on the standard deviations.

becomes increasing less efficient as primer concentration decreases. We can conclude that *hid1* and *hid3* are expressed relatively equally and greater than *hid2*.

4.2.3 Evaluating the potential regulation of co-expressed genes by RT-PCR

The microarray correlation table as described in (Pancaldi et al., 2010) was downloaded from the Bähler Resources website (www.bahlerlab.info/resources). The data corresponds to the normalized expression values for 5250 elements including mRNAs, ribosomal RNAs, and non-coding RNAs from 1162 hybridisation experiments. The table gives the linear correlation coefficients determined between each pair of genes, which then can be ranked for any given gene. Lists of ranked genes from highly positively correlated to highly negatively correlated were made for each *hid* gene (Table 4.1).

The genes chosen for analysis, generally, were based on the most highly correlated for a correlation group, but also function was taken into account. Signalling genes, like the F-box protein gene *pof11* and the stress-activated histidine kinase gene *mak2* that were positively correlated to *hid1* and the Checkpoint serine/threonine-protein kinase gene *bub1*, were preferentially selected. I also selected those genes that were involved in ubiquitination processes, like the ubiquitin carboxyl-terminal hydrolase 7 gene *ubp7*, and the Ubiquitin-activating enzyme E1 gene *ptr3* that were highly correlated with *hid3*. More positive-correlated genes were selected for *hid3* in order to test the large number of interesting genes with potentially interesting associations and test a greater range of coefficient values. With certain exceptions, for all three *hid* genes, the negatively correlated genes were less clear in terms of associated function and genes were selected based on correlation coefficient. It was interesting that the most highly negatively correlated gene for *hid1*, the thioredoxin domain-containing protein gene *plp1* was the most highly positively correlated to *hid2*. This indicated that *hid1* and *hid2* were inversely expressed.

The RT-qPCR primers were designed using QuantPrime and the analysis conducted on the same samples to be used for the RNAseq. It was expected that genes whose expression depended on, or was repressed by, the presence of a Hid protein, then the amount of transcript for that gene would change in the corresponding mutant. Figure 4.2 shows the fold-change in expression of the selected genes in each mutant sample. Those genes where the variation crosses the x-axis were not considered to be DE, which was the case for all genes tested for *hid1Δ*, except h1_1N, the gene encoding phosphofructokinase (Figure 4.2A). However for h1_1N, DE was down for this gene and DE was only between 2-5 fold. For genes tested with *hid2Δ*, there was a general trend in DR, but the amount of change was small and negatively correlated genes were also DR (Figure 4.2B). Only for genes tested in *hid3Δ* was there evidence for DR of positively correlated genes, 2 of the 3 negatively correlated showed some indication of UR, and the degree of change for a number of genes was more than those for *hid1Δ* and *hid2Δ* (Figure 4.2C).

Genotype	Sample number		
WT	WT 1	WT 2	WT 3
<i>hid1</i> Δ iso1	1 <i>hid1a</i>	2 <i>hid1b</i>	3 <i>hid1c</i>
<i>hid1</i> Δ iso3	1	2	3
<i>hid3</i> Δ iso1	1 <i>hid3a</i>	2 <i>hid3b</i>	3 <i>hid3c</i>
<i>hid3</i> Δ iso2	1	2	3
VCN1	1 VCNa	2 VCNb	3 VCNc
VCN2	1	2	3

Figure 4.3. Scheme of preparation of samples for post-genomic studies. Cell cultures were growth to a density of 5×10^7 cells/ml. The cells were collected by centrifugation, immediately frozen in liquid N₂, and ground to fine powder using a Cryomill. The genotypes were combined at this step, where equal masses of powder were combined for the two genotypes specified. The procedure yielded from 1.2 to 1.5 g of powdered cells. The circles represent the genotypes and the sample numbers that were combined. Wild-type was grown in triplicate.

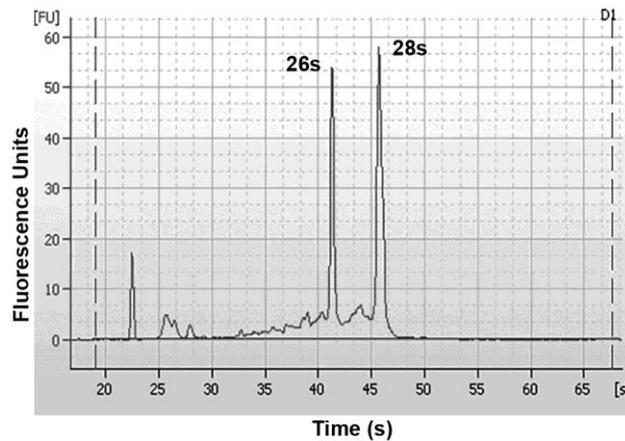


Figure 4.4. Quality determination of total RNA preparations. The quality of total RNA was performed at the Centre for Functional Genomics at Bordeaux using an Agilent 2100 Bioanalyzer™. The image shown corresponds to sample *hid3a* and which had a RIN value of 8.1. The large peaks correspond to the ribosomal RNAs 26s and 28s.

The greatest decrease in expression for a positively correlated gene was observed for *ubp5* (primer h3_11P), which was nearly 60-fold in some reactions. However, the decrease in expression of *ubp5* was tested with a second primer pair (h3_12P) and the overall decrease was more like the expression of the other positively correlated genes.

4.2.4 Global analysis of gene expression changes in *hid1Δ* and *hid3Δ* by RNAseq

4.2.4.1 Strategy of sample selection and evaluation of RNA quality

From the growth data presented in Chapter 3, it was clear that because of proliferation profile differences among the various strains that proper selection of mutants and the time of selection of samples would be important. For example, in order to obtain an “average” sample for *hid3Δ::natMX6*, it was necessary to let cultures stay under growth conditions for approximately 2 hours longer than for the other strains. Therefore, cells for post-genomic studies were taken based on cell count and not culture time. Cells were cultured in 50 ml YES as described in Chapter 2 with an initial density of approximately 5×10^5 cells/ml. Cells were harvested at a density of 5×10^7 cells/ml. This cell density was chosen as it was at the transition between lag- and log-phase growth, and it would clearly present the differences among strains (Figure 3.10). Cells were grown in large cultures to ensure there was sufficient material in order to conduct all aspects of the post-genomic study, transcriptomics, proteomics and metabolomics, on the same samples. In order to include biological variation within the study, each sample was made of two different genotypes according to the scheme shown in Figure 4.3. This approach gave 12 samples from which total RNA was isolated for RNAseq and RT-PCR analyses of transcript levels.

Total RNA corresponding to the 12 samples was isolated from frozen powdered material as described in Chapter 2 and the quality verified by an Agilent 2100 Bioanalyzer™ 2100. The phenol-based extraction procedure gave high quality RNA that typically showed RIN values > 8. Figure 4.4 shows the worst-case output from all 12 samples. A good sample will show large, clearly defined ribosomal RNA peaks with a smaller broad peak corresponding to the mRNA.

4.2.4.2 RNAseq reveals greatest differences in gene expression for *hid3Δ*

For each sample, the reads from the Illumina MySEQ™ DNA sequencer came in two fastq files corresponding to forward and reverse adapters for paired-end reading. For each sample, the reads contained within the two files were mapped together to the *S. pombe* genome Ensembl build EF2 using Tophat v. 2.0.9. The statistics for the number of reads obtained for each sample and the percentage of

Table 4.2. Statistics for read mapping. This is a summary of the percentage of reads mapped for each sample using Tophat v 2.0.9. The input # reads represents the total number of reads for that sample for which both left and right priming was initiated. Under Mapped Reads are given the number of reads mapped for each paired-end primer and the percentages of reads mapped and the percentage of those mapped to more than one sequence.

Sample	Input # reads	Mapped Reads	
		Left ^a	Right ^a
WT 1	2116612	1974817 (93.3, 11.1)	1942198 (91.8, 11.0)
WT 2	2664071	2501691 (93.3, 11.7)	2455439 (92.2, 11.6)
VCNa	2035967	1893322 (93.0, 11.1)	1864488 (91.6, 11.1)
VCNb	2235052	2101346 (94.0, 13.8)	2066680 (92.5, 13.8)
VCNc	2540626	2394197 (94.2, 17.2)	2332817 (91.8, 17.1)
<i>hid1Aa</i>	2387953	2249338 (94.2, 16.7)	2223351 (93.1, 16.7)
<i>hid1Ab</i>	2296483	2162705 (94.2, 16.7)	2128662 (92.7, 16.6)
<i>hid1Ac</i>	940907	886581 (94.2, 15.0)	872427 (92.7, 14.9)
<i>hid3Aa</i>	1480614	1393083 (94.1, 16.7)	1372959 (92.7, 16.7)
<i>hid3Ab</i>	2005502	1822468 (90.9, 10.4)	1786277 (89.1, 10.3)
<i>hid3Ac</i>	1890765	1741271 (92.1, 22.0)	1724848 (91.2, 22.0)

^a The number in parentheses represents the percentage of reads mapped and the percentage of reads with multiple alignments, respectively.

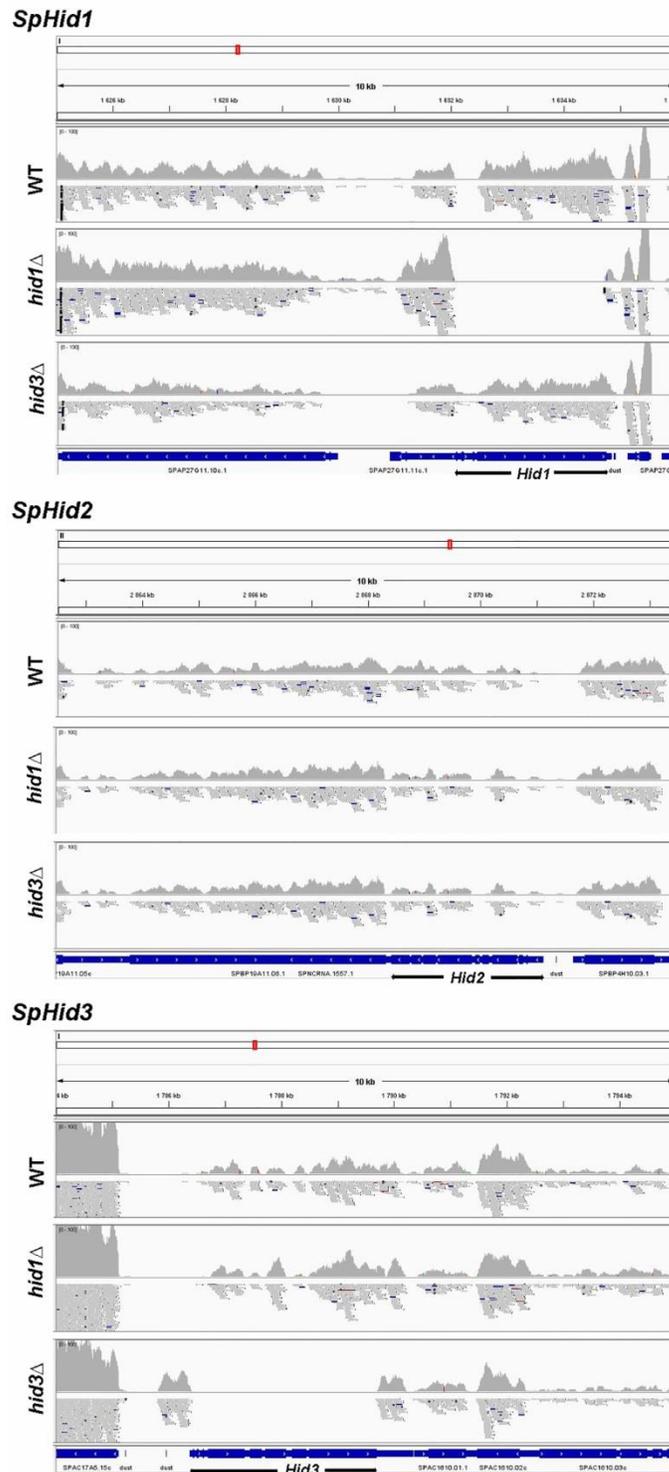


Figure 4.5. Visualisation of mapped reads at *hid* loci. The images shown are representative of the triplicate mutant samples. The locus shown is given above the image and the precise location of the *hid* gene shown below. The reads were mapped using Tophat v2.0.11 and the images were created from the ‘.bed’ files using the Integrated Genomics viewer (Robinson et al., 2011b). The location of the genes is from the EF2 version of the pombe genome downloaded from iGenomes (www.broadinstitute.org). The windows have been expanded to the same scale to provide an estimate of relative expression level.

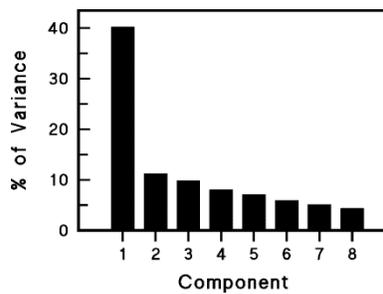
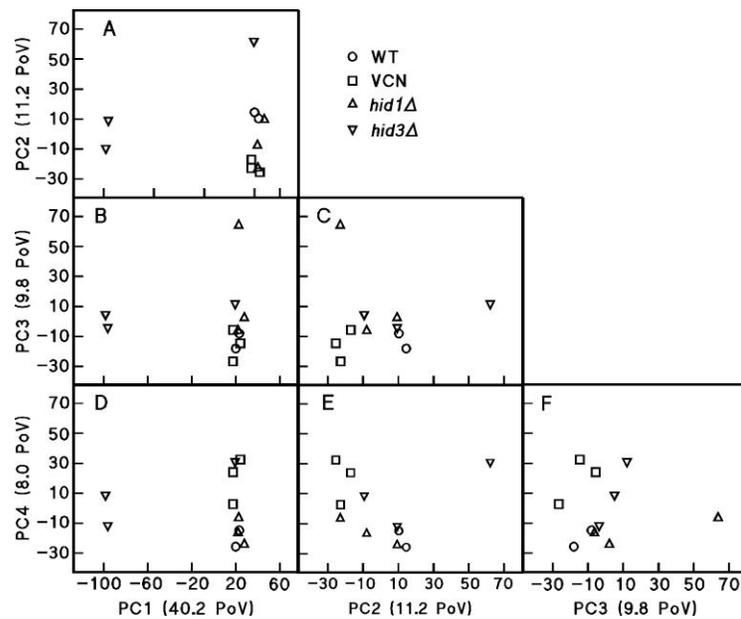
A**B**

Figure 4.6 PCA of combined gene expression data from RNAseq. PCA was conducted in R using the FactoMineR package, which exploits the resident ‘princomp’ tool resident in R. All rows with missing data had to be removed as ‘princomp’ does not recognise NaN. The number of data feature after missing data was removed was 5272. (A) Plot of variance for the first 8 PCs. (B) Pairwise relationships between the first four PCs. Only four PCs were analysed since the plot of PC4 vs PC3 clusters samples around 0. The N in the legend for sample names represents replacement of the gene by the Nat^R gene. PoV stands for Percentage of Variance.

reads mapped is given in Table 4.2. One of the wild-type samples did not work at the level of library creation for unknown reasons.

With the exceptions of *hid1Δc* and *hid3Δa* around 2 million total reads were obtained with more than 90% of them being able to be mapped to the genome. Two essential data files were created, one was the ‘.bed’ file that was used to visualise the mapped reads directly and the second was the ‘.bam’ file used for quantifying the mapped reads. The reads around the mutant loci were visualised as a straightforward test to check the accuracy of the mapping as it would be expected that no reads would be mapped to genes replaced in the respective mutant, but that reads would be present at the other gene loci (see Chapter 3). For the *hid1Δ* and *hid3Δ* mutants, no reads were mapped to the altered loci, but were mapped to the two unaltered loci showing that those genes were expressed (Figure 4.5). From the visualization, qualitatively I could assess that expression of non-mutant *hid* loci within any particular mutant was unaffected compared to wild-type, which is consistent with the RT-PCR results of the mutants presented in Chapter 3. Also, expression immediately at the flanks of the mutant loci also appeared unaffected, indicating precise insertion of the marker gene.

The next step in the analysis was to quantify the reads that had been mapped. This was done using the Cufflinks workflow described above. The individual annotation files ‘.gtf’ were fed into Cuffcompare in order to generate a ‘combined.gtf’ file for normalisation of data across samples using Cuffnorm. Cuffnorm gave 6007 features in the genes_fpkms_table file that were common amongst the 11 biological samples. The Cuffnorm output was exported as a text file and imported into the program R for multivariate statistical analysis. After removal of features without data, 5252 features were left. The PCA analysis provided an overall picture as to differences among the samples. First, the contribution of PCs to the total variance among samples was determined. The first four eigenvectors accounted for 69.2% of total variance with PC1 accounting for about 40% of that (Figure 4.6). A broad spread of variance in subsequent PCs 2-11 is not unusual noting the large amount of data included and the variation that data was likely to have.

The plot of PC2 vs. PC1 shows that the *hid3* samples differed substantially from the other genotypes (Figure 4.6B). However, *hid3Δa* was comparable to the other genotypes in the first dimension and only differed in the second. There is very little separation of non-*hid3Δ* genotypes. In any dimension except for *hid1Δc*, where its separation may be due to a lower number of reads being present. It was expected that the differences observed for *hid3Δ* would also show up in the differential gene expression analysis.

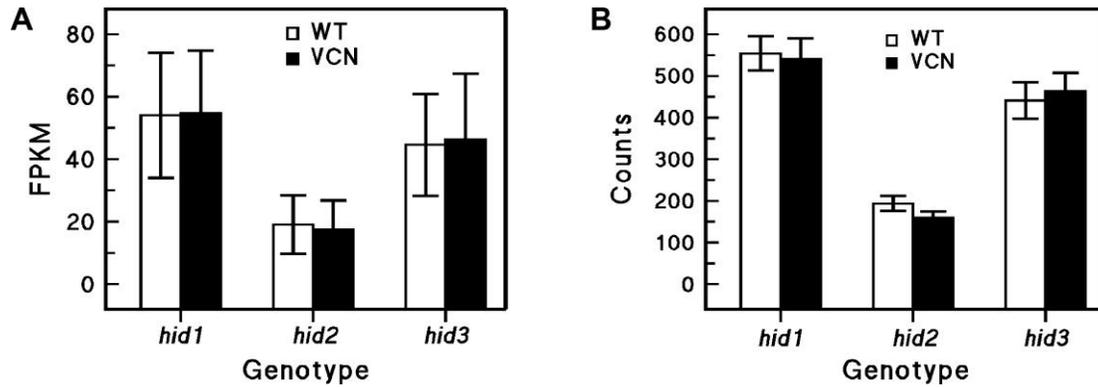


Figure 4.7. Relative expression of *hid* genes based on read counts. The reads were taken from either the gene.fpkm_tracking file from the Cuffdiff output (A) or from the counts.txt output file from featureCounts v1.4.6 (B). The data in Cuffdiff is represented as mean \pm the high and low values, whereas for featureCounts the data is represented as mean \pm range (WT, n=2) or sd (VCN, n=3). Counts for *hid2* in VCN samples are statistically significantly different ($P < 0.05$) from those for *hid1* and *hid3* based on Student's t-tests.

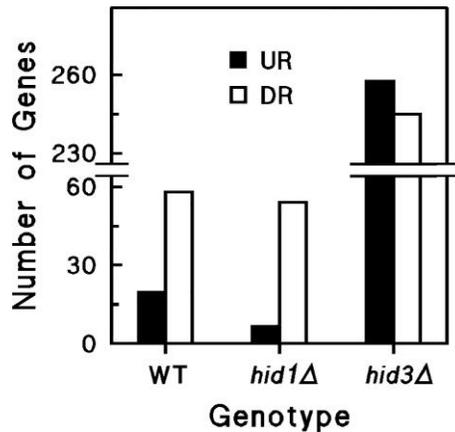


Figure 4.8. Number of genes differentially expressed in the mutants. The number of DE genes is based on the statistical calculation within Cuffdiff and doesn't necessarily reflect a strict fold-cutoff value. Each genotype has been compared to the VCN samples. UR is up-regulated expression and DR is down-regulated expression.

Table 4.3. DE genes selected for qRT-PCR for *hid3Δ*. UR and DR stand for up- and down-regulated respectively. The fold-expression is the ratio of average counts as calculated by Cuffdiff for *hid3Δ* over VCN.

Systematic ID	UR/DR	Fold-expression
SPBC4F6.09	UR	3.3
SPCC1742.01	UR	2.8
SPAC1F7.07c	UR	2.8
SPAC1F8.06	UR	2.2
SPAC1F7.08	UR	2.2
SPAC186.06	DR	4.3
SPNCRNA.942	DR	3.7
SPAC977.18	DR	2.8
SPBC685.02	DR	2.5
SPNCRNA.863	DR	2.5
SPCPB16A4.07	DR	2.3

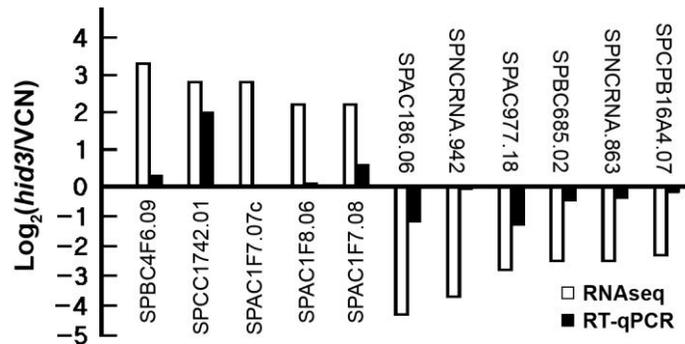


Figure 4.9. Comparison of relative transcript quantification by RNAseq and RT-qPCR. One μg of the total RNA from each sample was converted to cDNA and the equivalent of 20 ng used for qPCR. The values are presented as the \log_2 of the ratio of the amount of transcript in *hid3Δ* to that in VCN. The values shown for the RT-qPCR are the means of the independent RNA samples as shown in Figure 4.1 ($n=3$). The names are the Systematic IDs of the genes selected.

Cuffmerge was used to create the global ‘merged.gtf’ from the Cufflinks output that was subsequently used to quantify transcripts by Cuffquant and calculate differential gene expression by Cuffdiff. Cuffdiff compared genotypes pairwise and by taking into account replicates determined statistically significant DE genes. The output from Cuffdiff includes a ‘tracking’ file that provides the number of counts at each annotated locus for each sample. From this file, the number of normalized counts for the three *hid* genes was extracted in order to compare relative expression levels with the RT-qPCR results presented in (Figure 4.7A). A second means of measuring gene expression was directly through exon hit counting using featureCounts (Liao et al., 2014). The levels of mRNA for *hid1* and *hid3* are approximately equal and that for *hid2* is nearly 3-fold lower than the other two. The counting also confirmed the greatly reduced expression of *hid1* and *hid3* in the corresponding mutant strains. However, there was residual expression of *hid3* in the *hid3Δ* strain, because the knock-out procedure to remove the coding sequence left the end of the last exon.

The number of DE genes was determined for each genotype expressed relative to the VCN samples (Figure 4.8). In order to understand the effect of the marker gene, the UR and DR genes for wild-type must be inverted. Therefore, expression of the marker gene at heterologous loci resulted in 58 and 20 genes being UR and DR, respectively, compared to wild-type. The total number of genes changing expression in *hid1Δ* was no different to that of wild-type, but *hid3Δ* showed substantial amounts of altered gene expression with 258 and 245 genes being UR and DR, respectively. The amount of changing gene expression was not great with highest UR gene for *hid1Δ* and *hid3Δ* being 3.1- and 9.8-fold respectively. The greatest DR genes for *hid1Δ* and *hid3Δ* were 10.8- and 20.2-fold, respectively.

4.2.4.3. Validation of RNAseq data by qRT-PCR

Primers for qRT-PCR were produced for the top five UR and DR genes for *hid3Δ* compared to VCN (Table 4.3). Quantitative RT-PCR was conducted using the same total RNA samples as used for the RNAseq analysis. *act2* was used in order to normalise expression for each primer pair. For all UR and DR genes taken from the RNAseq quantification, the RT-qPCR either showed a corresponding directional change in expression although the degree of change was consistently less (Figure 4.9).

4.2.4.4. Analysis of correspondence with genetic interaction data provided in PomBase.

The sets of DE genes were searched with the gene identifiers for those genes for which a genetic interaction has been proposed and has been listed in PomBase. For the 6 interactions listed for *hid1*, none were DE. STRING database lists three other genes, which have been linked to *hid1* through co-expression: SPCC18.03, a shuttle craft like TF; SPAC6B12.07c.1, an ubiquitin-protein ligase E3; and

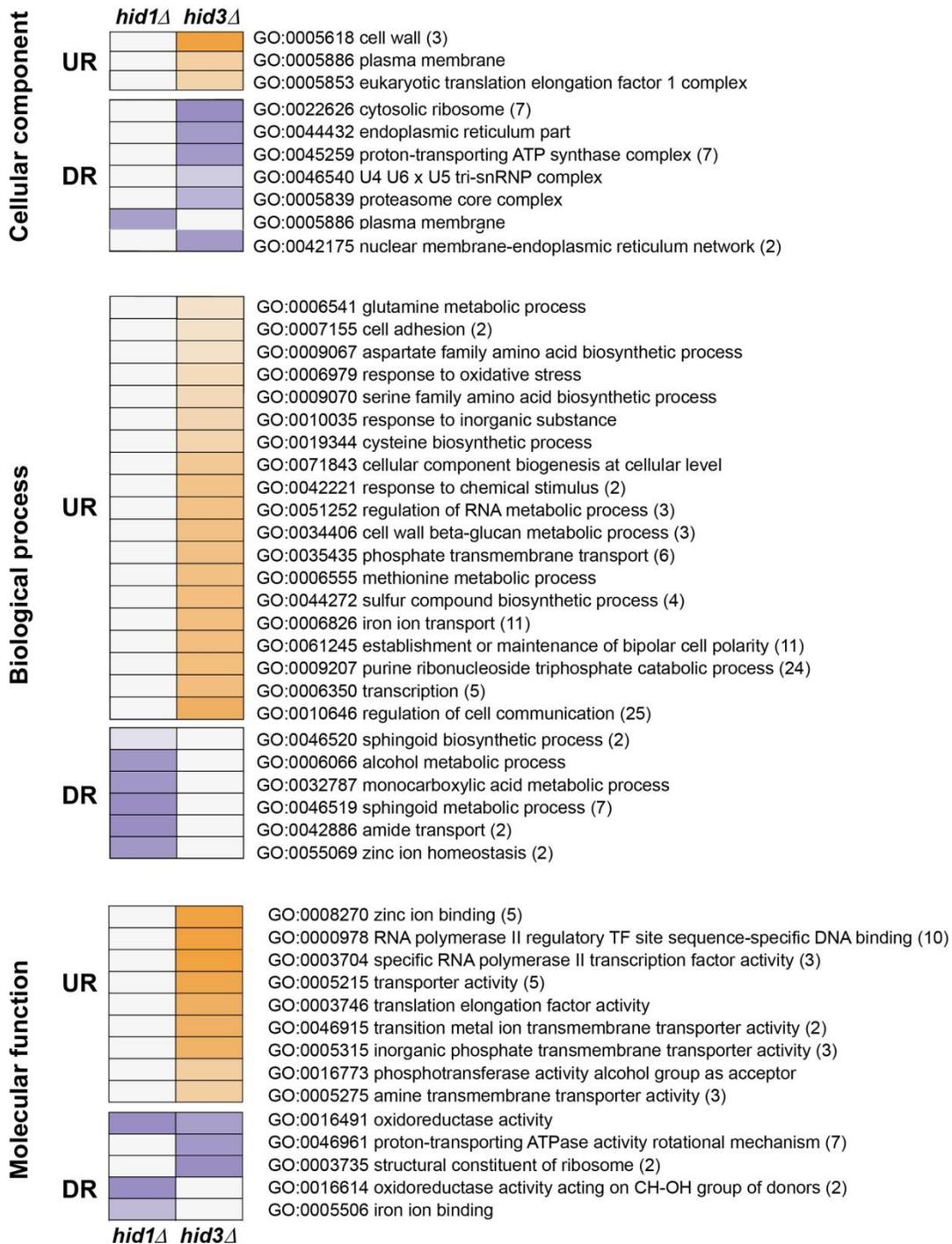


Figure 4.10. Gene Ontology analysis of DE genes. GoMiner was given the sets of DE genes for *hid1Δ* and *hid3Δ* and the set of all genes for which there were reads to use as the baseline. GoMiner determined the classes of DE genes that were overrepresented for both UR and DR genes. The GO terms have been pruned by REVIGO (revigo.irb.hr, (Supek et al., 2011)) to reduce the number of displayed GO term and ease the visualisation. The numbers in parentheses represent the number of classed condensed to give the final go term. The degree of colour is indicative of the statistical significance.

Pof11, an F-box protein. None of these were DE. Of the 101 genes with potential genetic interactions with *hid3*, seven genes were in the list of DE. Six of the seven genes were UR and none by more than 2-fold. The one gene with a negative interaction that was DR was the phosphatidyl-N-dimethylethanolamine N-methyltransferase *cho1*, which is involved in lipid metabolic processes of the ER. The low-confidence, co-expressed genes listed in the STRING database were also not represented in the set of DE genes.

4.2.4.5. Finding biological function through gene ontologies

The DE genes were run through GoMiner (<http://discover.nci.nih.gov/gominer/index.jsp>, (Zeeberg et al., 2003) in order to determine if the changing genes were associated with any biological or molecular function. The queries were made using the systematic IDs so no identification of particular genes was made at this time. GoMiner yielded those GO classes that are observed to have more genes altered in expression compared to their representation in the background data set of all genes.

The output entailed the three major GO categories, Cellular Component, Biological Function and Molecular Function. Seeing that *hid3Δ* had 503 DE genes compared to 78 for *hid1Δ*, it was expected and shown that *hid3Δ* had more GO classes changing than *hid1Δ* (Figure 4.10). For *hid1Δ*, the number of DR genes was greatly more than UR genes and this is represented in the GO analysis with there being no particular class of genes being up-regulated in each of the three categories. The DR genes in *hid1Δ* for both biological and molecular function related to metabolism processes. The Cellular Component category for *hid3Δ* was primarily represented by DR genes encoding proteins related to ribosome function GO:0022626 and mitochondrial ATP synthesis GO:0045259. The Biological Processes appear to be represented by a wide variety of classifications, but they centre around particular themes. Amino acid metabolism is represented by seven different GO classifications including for glutamine (GO:0006541), aspartate (GO:0009067), serine (GO:0009070) and the two sulfur amino acids cysteine and methionine (GO:00019344, GO:0006555, GO:00044272). Stress response genes (GO:0006979, GO:0010035, GO:0042221), membrane transport (GO:0035435, GO:0006826), RNA and nucleotide metabolism, mainly catabolism (GO:0009207, GO:0051252) and cell communication and maintenance (GO:0007155, GO:0061245, GO:0006350, GO:0010646) were also represented. The Molecular Functions were represented by the themes transcriptional activity (GO:0008270, GO:0000978, GO:0003704) and membrane transport (GO:0005215, GO:0046915, GO:0005315, GO:0005275). The Molecular Function classes DR involved metabolic (GO:0016491, GO:0046961) and ribosomal protein (GO:0003735) genes.

Table 4.4. Transcription Factor genes induced in *hid3Δ*. Gene corresponding to GO:0000978 as determined by GoMiner using the list of DE genes from Cuffdiff. The fold-change corresponds to the \log_2 value of the ratio of normalized reads of *hid3Δ* to VCN. TF stands for transcription factor.

Name	Description	Fold change
<i>sfp1</i>	TF <i>sfp1</i> (predicted)	1.11
<i>SPAC2H10.01</i>	TF zf-fungal binuclear cluster type (predicted)	1.08
<i>gsf1</i>	TF zf-fungal binuclear cluster type Gsf1	1.00
<i>SPBC530.11c</i>	TF, zf-fungal binuclear cluster type (predicted)	0.97
<i>esc1</i>	TF <i>esc1</i> (predicted)	0.91
<i>loz1</i>	TF zf-C2H2 type (predicted)	0.91
<i>scr1</i>	TF <i>scr1</i>	0.86
<i>atf1</i>	TF Atf-CREB family <i>atf1</i>	0.85
<i>SPBC56F2.05c</i>	TF (predicted)	0.80
<i>pho7</i>	TF <i>pho7</i>	0.80
<i>gaf1</i>	TF <i>gaf1</i>	0.76
<i>rst2</i>	TF <i>rst2</i>	0.66
<i>SPCC1393.08</i>	TF, zf-GATA type (predicted)	0.56
<i>hsr1</i>	TF <i>hsr1</i>	0.55
<i>SPAC1327.01c</i>	TF, zf-fungal binuclear cluster type (predicted)	0.48

The ontological category GO:0000978 represents those genes assigned the function of ‘RNA polymerase II core promoter proximal region sequence-specific DNA binding’. The molecular functional category contains 90 TFs of which 15 are significantly induced in *hid3Δ* (Table 4.4).

One of the induced TFs was *atf1*, which is a CREB/ATF family TF. This transcription factor is integral to the induction of stress-related genes. (Chen et al., 2003) investigated the global expression responses of *S. pombe* in relation to various stresses. A comparison of the 258 *hid3Δ* UR genes with the 456 UR genes from their study showed that 68 (> 25% of *hid3Δ* UR genes) were in common between the two data sets. In addition, they determined which genes the induced was dependent on either of the two integral signaling genes *atf1* or *sty1*, which encodes a MAPK that activates the TF Atf1. Of the 68 overlapping genes, 45 were known to be affected by Atf1 independently or in addition to the presence of Sty1. The expression of *sty1* in our data was unchanged amongst all genotypes, which corresponded with the findings of Chen et al. (2003). The protein Pcr1 is known to physically interact with Atf1 to modify transcription of the *cgs2* by binding to its promoter (Davidson et al., 2004). The gene *pcr1* was not significantly UR in *hid3Δ*, but *cgs2* was induced approximately 1.8-fold, but it was not flagged as statistically significantly UR due to variation in its expression. The stress expression data from Chen et al. (2003) was checked for common patterns of expression for the other TFs. Although most were induced in response to one or more stresses, there did not seem to be any clear pattern of expression as either CESR or SESR genes (Chen et al., 2003), and four UR in my study were initially DR in their data set.

4.2.4.6. Changes in expression of specific genes

Of the 258 and 245 genes UR and DR, respectively, 35 were increased and 72 decreased by at least 2-fold. Therefore, although there were more genes statistically determined to be UR than DR, down-regulation was apparently more pronounced in effect (Table 4.5). There were no genes that were dramatically altered in expression with the largest change observed being the 4.3-fold decrease in expression of an ER unfolded-response protein gene SPAC186.06. The UR genes primarily encoded plasma membrane transporters, cell surface proteins, signaling/nucleic acid binding genes and some stress-related and ubiquitination-related genes. Interesting genes that were UR, but less than 2-fold, were DNA helicase *mot1*, the transcription factor TFIIIC complex subunit gene *sfc4*, and the protein kinase gene *wis1*. The set of DR genes was more varied in composition and contained more metabolic genes including internal metabolite transporter-related genes. The largest class of DR genes encoded ribosomal proteins, and of the 245 total DR genes, 37 encoded 40S or 60S ribosomal proteins.

Table 4.5. DE genes in *hid3Δ*. The genes shown have been calculated by Cuffdiff as significantly UR or DR by at least 2-fold.

Gene Name	log ₂ (<i>hid3Δ</i> /VCN)	Gene Description
Up-regulated		
<i>str1</i>	3.3	siderophore-iron transporter Str1
<i>gsf2</i>	2.8	galactose-specific flocculin Gsf2
<i>fip1</i>	2.8	iron permease Fip1
<i>SPCC663.08c</i>	2.5	short chain dehydrogenase
<i>fta5</i>	2.2	cell surface glycoprotein
<i>fiol</i>	2.2	iron transport multicopper oxidase Fiol
<i>osr1</i>	2.1	short chain dehydrogenase (predicted)
<i>SPCC18B5.02c</i>	2.1	pseudogene
<i>set10</i>	2.1	ribosomal lysine methyltransferase Set10
<i>SPAC27D7.09c</i>	1.9	But2 family protein
<i>pfl3</i>	1.8	cell surface glycoprotein (predicted). DIPSY family
<i>pfl9</i>	1.8	cell surface glycoprotein (predicted). DIPSY family
<i>caf5</i>	1.8	spermine family transmembrane transporter Caf5
<i>SPBPB10D8.01</i>	1.8	cysteine transporter (predicted)
<i>SPNCRNA.1153</i>	1.7	antisense RNA (predicted)
<i>gst2</i>	1.6	glutathione S-transferase Gst2
<i>SPNCRNA.1424</i>	1.3	intergenic RNA (predicted)
<i>SPAPJ691.02</i>	1.3	yippee-like protein
<i>fet4</i>	1.2	iron/zinc ion transporter (predicted)
<i>bit61</i>	1.2	TORC2 subunit Bit61
<i>cys2</i>	1.2	homoserine O-acetyltransferase (predicted)
<i>SPNCRNA.1355</i>	1.2	intergenic RNA (predicted). possible alternative UTR
<i>SPBC409.08</i>	1.2	spermine family transporter (predicted)
<i>bfr1</i>	1.1	brefeldin A efflux transporter Bfr1
<i>SPCC1739.06c</i>	1.1	uroporphyrin methyltransferase (predicted)
<i>21S_rRNA</i>	1.1	21S rRNA mitochondrial
<i>sfp1</i>	1.1	transcription factor Sfp1 (predicted)
<i>SPAC11D3.11c</i>	1.1	zinc finger protein. truncated
<i>ubp16</i>	1.1	ubiquitin C-terminal hydrolase Ubp16
<i>SPAC2H10.01</i>	1.1	TF zf-fungal binuclear cluster type (predicted)
<i>SPBP8B7.15c</i>	1.1	ubiquitin-protein ligase E3 RBBP6 family (predicted)
<i>dfr1</i>	1.0	dihydrofolate reductase protein Dfr1
<i>sap1</i>	1.0	switch-activating protein Sap1
<i>SPCC1672.14</i>	1.0	sequence orphan
<i>adh8</i>	1.0	alcohol dehydrogenase (predicted)
Down-regulated		
<i>SPAC186.06</i>	4.3	ER unfolded protein response protein (predicted)
<i>SPNCRNA.942</i>	3.7	intergenic RNA (predicted)
<i>SPAC23H3.15c</i>	3.0	sequence orphan
<i>SPAC977.18</i>	2.8	conserved fungal protein
<i>exo5</i>	2.5	mitochondrial single stranded DNA specific 5'3' exodeoxyribonuclease Exo5 (predicted)
<i>SPNCRNA.863</i>	2.5	intergenic RNA (predicted)
<i>SPBC23G7.13c</i>	2.4	urea transporter (predicted)
<i>SPCPB16A4.07</i>	2.3	sequence orphan
<i>mug96</i>	2.3	sequence orphan
<i>SPBC1271.07c</i>	2.3	N-acetyltransferase (predicted)
<i>SPBC1271.08c</i>	2.3	sequence orphan
<i>SPBTRNAGLN.04</i>	2.3	tRNA Glutamine
<i>SPAC1399.04c</i>	2.3	uracil phosphoribosyltransferase (predicted)
<i>SPAC186.04c</i>	2.2	N-terminal of transmembrane channel truncated

<i>aca1</i>	2.2	Lazetidine-2-carboxylic acid acetyltransferase Aca1
<i>urg2</i>	2.0	uracil phosphoribosyltransferase (predicted)
<i>urg1</i>	2.0	GTP cyclohydrolase II Urg1 (predicted)
<i>SPBC660.05</i>	1.7	WW domain containing conserved fungal protein
<i>SPAC521.03</i>	1.7	short chain dehydrogenase (predicted)
<i>SPCC70.03c</i>	1.6	proline dehydrogenase (predicted)
<i>SPCC70.04c</i>	1.6	sequence orphan
<i>SPAPJ695.02</i>	1.6	sequence orphan
<i>hem13</i>	1.4	co-proporphyrinogen III oxidase Hem13 (predicted)
<i>urg3</i>	1.4	DUF1688 family protein
<i>SPAC29B12.13</i>	1.4	S-(hydroxymethyl)-glutathione synthase activity
<i>dap1</i>	1.4	cytochrome P450 regulator Dap1
<i>SPAC823.02</i>	1.4	sequence orphan
<i>tam13</i>	1.4	sequence orphan
<i>vma10</i>	1.4	V-type ATPase V1 domain subunit G (predicted)
<i>gto2</i>	1.4	alpha-glucosidase (predicted)
<i>SPAC1039.02</i>	1.4	phosphoprotein phosphatase (predicted)
<i>lsm6</i>	1.3	U6 snRNP-associated protein Lsm6 (predicted)
<i>SPNCRNA.1483</i>	1.3	intergenic RNA (predicted)
<i>SPBC26H8.16</i>	1.3	mitochondrial protein. DUF1674 family
<i>SPAC1039.01</i>	1.3	amino acid permease (predicted)
<i>SPBPB21E7.08</i>	1.3	pseudogene
<i>SPBC12C2.09c</i>	1.3	Haemolysin III family protein (predicted)
<i>rpl4102</i>	1.3	60S ribosomal protein L41 (predicted)
<i>SPCC757.05c</i>	1.2	peptidase family M20 protein
<i>rpl3602</i>	1.2	60S ribosomal protein L36
<i>rbx1</i>	1.2	RING-box protein 1. SCF-complex ubiquitin ligase subunit
<i>mug8</i>	1.2	conserved fungal protein
<i>SPBC29A10.12</i>	1.2	DUF1014 family protein. HMG-box clan member
<i>aes1</i>	1.2	enhancer of RNAMediated gene silencing
<i>rpp202</i>	1.2	60S acidic ribosomal protein A4
<i>srp14</i>	1.2	signal recognition particle subunit Srp14
<i>pmc6</i>	1.2	mediator complex subunit Pmc6
<i>tim8</i>	1.1	TIM22 inner membrane protein import complex subunit Tim8 (predicted)
<i>rps2801</i>	1.1	40S ribosomal protein S28 (predicted)
<i>SPCC16C4.20c</i>	1.1	HMG box protein (predicted)
<i>ght5</i>	1.1	hexose transporter Ght5
<i>SPAC589.10c</i>	1.1	ribosomal-ubiquitin fusion protein Ubi5 (predicted)
<i>SPNCRNA.1626</i>	1.1	antisense RNA (predicted)
<i>SPAC6F12.04</i>	1.1	COPI-coated vesicle associated protein (predicted)
<i>sss1</i>	1.1	translocon gamma subunit Sss1 (predicted)
<i>SPBC800.11</i>	1.1	inosineuridine preferring nucleoside hydrolase
<i>SPNCRNA.1333</i>	1.1	intergenic RNA (predicted)
<i>cyp4</i>	1.1	cyclophilin family peptidylprolyl cis-trans isomerase
<i>SPNCRNA.1292</i>	1.1	antisense RNA (predicted)
<i>rps21</i>	1.1	40S ribosomal protein S21
<i>gut2</i>	1.1	glycerol-3-phosphate dehydrogenase Gut2 (predicted)
<i>mrp21</i>	1.1	mitochondrial ribosomal protein subunit Mrp21
<i>cox13</i>	1.1	cytochrome c oxidase subunit VIa (predicted)
<i>zym1</i>	1.0	metallothionein Zym1
<i>rps2401</i>	1.0	40S ribosomal protein S24 (predicted)
<i>lsd90</i>	1.0	Lsd90 protein
<i>rpl22</i>	1.0	60S ribosomal protein L22 (predicted)
<i>srb7</i>	1.0	mediator complex subunit Srb7
<i>rpl4101</i>	1.0	60S ribosomal protein L41
<i>SPAC11E3.10</i>	1.0	VanZ-like family protein
<i>SPBC29A10.16c</i>	1.0	cytochrome b5 (predicted)
<i>SPAPJ691.03</i>	1.0	mitochondrial inner membrane organizing system protein

Within the DR group of genes, there were two genes, *srb7* and *pmc6*, that encode TF components of the RNA polymerase II complex, and would, therefore, appear to be acting opposite to the RNA polymerase II-related TFs in Table 4.4. A closer look showed that nearly all genes encoding subunits of the Polymerase II mediator complex were either unchanged or DR. Two mediator complex genes were UR, *pmc1* and *srb8*, but both were not under the cut-off value for statistical significance. Within the set of DR genes were also two RNA polymerase I TFs, *rrn7* and *rrn9* and one RNA polymerase III subunit *rpc11*.

4.3 DISCUSSION

With user-friendly databases currently available to search for cancer-related gene expression, it was easy to search for *HID1* for patterns of gene expression in tumour and normal tissues in order to validate the hypothesis that down-regulation of *HID1* may promote cancer formation Harada et al (2001). The observation that *HID1* is more often UR than DR in cancers indicates that it is not involved in tumour suppression. It is possible that *HID1* is induced by stresses that cause cancer, such as DNA damaging agents and that *HID1* is part of a stress sensing and signaling process as it functions in the formation of Dauer larva in *C. elegans* under conditions of stress (Ailion and Thomas, 2003).

In order to postulate on human *HID1* function, this discussion will primarily focus on the results obtained relating to the pombe *hid3Δ* mutant. It is unlikely that *Hid3* is a component of the stress signaling process, because the expression of *hid3* is not affected by stress (Chen et al., 2003). In contrast, *hid1* is induced by various stresses including oxidative, heat and DNA damage, but at this time the reason for this induction is unclear. If we start with the assumption that disruption of Golgi function in *hid3Δ* (Alasmari 2015) induces a chronic stress imposed immediately upon the daughter cells after division, then the effects of this stress would be reflected in altered gene expression profiles serving to compensate for this stress. There are potential pleiotropic effects on gene expression that must be considered. The mislocalisation of *Ubp5* from the Golgi to the cytosol (Kouranti et al., 2010) may have indirect effects on gene expression. Genes found to be positively correlated to *hid3* expression involved ubiquitination processes, such as *ubp7*, *ubp5* and *ptr3*, as well as potential signaling genes, such as *bub1*, *hhp1* and *ppk29*. Therefore, it was logical to determine if the expression of any of these genes was affected in the mutant. Although there was some indication that transcript levels of some of the correlated genes may be affected (and in the expected direction) in *hid3Δ*, the effect was not strong. This was certainly the case for genes correlated with *hid1* and *hid2* when compared with those correlated with *hid3*. Therefore, it is possible to conclude that gene co-expression of other genes is not greatly dependent on the presence or level of *Hid3*. It is also possible that overall transcriptional responses in the *hid* mutants are not very dramatic.

Global transcriptomic analyses of the *hid1Δ* and *hid3Δ* mutants were done using RNAseq. Unfortunately, the delay in obtaining the *hid2Δ* mutants did not allow me to collect the RNAseq data until very recently. This data has not gone through the GO analysis and the effect on individual genes has not been investigated in detail. It appears that the number of genes altered in expression in *hid2Δ* were few like that seen for *hid1Δ*. The RNAseq mapping through Tophat correctly demonstrated the lack of expression of the *hid* genes in the respective mutants, but also supported the relative quantification of the transcript levels of the three *hid* genes that had been determined by RT-qPCR. If RT-qPCR is considered to be the best method for determining the transcript levels of specific genes (Marioni et al., 2008), then inversely, the *hid* gene RT-qPCR helps to validate the RNAseq data. The RT-qPCR validation of the RNAseq data using the greatest UR and DR genes for the *hid3Δ* data was not as clear, except that the direction of expression change was correct for nearly every gene tested. According to the RT-qPCR only 1 gene, SPCC1742.01, would be considered to be DE if a strict cut-off of $\log_2 = 2$ was used. I attribute this difference to the fact that the primers and the reaction conditions for the other UR and DR genes were not as well optimized as those for the *hid* genes. Optimization was not done, because 1) of the corresponding *hid* RT-qPCR validation and 2) it was not worth the effort to optimise all reactions, since this would not change the way the RNAseq data was analysed. In addition, RNAseq data tends toward false-negative results, whereby genes that are DE are not identified as such (Liu et al., 2011), thus we can have reasonable confidence on the biological conclusions made from the set of DE genes that have been determined.

The PCA conducted on the combined normalised data clearly revealed that the gene expression changes were greater in *hid3Δ* compared to *hid1Δ*. This is not surprising with the obvious effect of *hid3* deletion on reducing growth and there is no effect of deleting *hid1*. There were modest gene expression changes between wild-type and VCN, but these changes were not apparent in the PCA scores plot, where VCN samples clustered with wild-type and *hid1Δ* samples. Therefore, I can conclude that there was no apparent effect of the presence of the marker protein on global gene expression. This is in contrast to protein and metabolite levels that exhibit a noticeable change in their respective PCA scores plots in the PC2 dimension (Alasmari 2015).

The greater change in transcriptional activity in *hid3Δ* was confirmed by the finding through the Cuffdiff DE analysis that more than 6-times more genes were significantly DE in *hid3Δ* than *hid1Δ* when compared to VCN. No genes postulated to have a genetic interaction with *hid3* or *hid1* (Roguev et al., 2008; Ryan et al., 2012) were shown to be DE, which is consistent with the RT-qPCR result of correlated

genes. This suggests that any genetic interaction does not come from a change in gene expression, but that any interaction must be a physical interaction between proteins that is perturbed by the lack of the corresponding Hid protein. The levels by which genes were DE were not extreme with only 9.8-fold and 19.7-fold being the greatest changes observed for a UR and DR gene, respectively. This might be expected noting that the cells were being grown under non-stress conditions prior to sampling and the only effect would be due to the mutation. However, the differences in gene expression provide evidence that *hid3Δ* cells are stressed. A large number of DR genes without clear functional categorization suggests a broad repression of gene expression, which is indicative of cellular stress (Gasch, 2007). There were also many ribosomal genes DR (not shown in Table 4.5), which is also a symptom of cells under stress. The surprising outcome from the expression data was the induction of specific classes of genes. The induction of the set of transcription factors that are part of the RNA polymerase II complex may explain the broad changes in gene expression for *hid3Δ* and those that are related to stress signaling in particular. The gene *atf1* encodes a TF that is part of the the MAPK stress-signaling pathway. Atf1 interacts with Pcr1 to induce the expression of *cgs2⁺* to affect cAMP-dependent signaling (Davidson et al., 2004). We do not observe induction of *pcr1⁺*, but *cgs2⁺* is induced 1.8-fold although it was not flagged as being significantly UR. *atf1⁺* is activated upon phosphorylation by Sty1 as part of the MAPK stress-signalling pathway, but we did not observe induction of *sty1*. Sty1 has other targets for phosphorylation, like Bfr1, Pap1 and Ntp1. Bfr1 is a plasma membrane ABC-type transporter responsible for resistance to drugs like tributyltin (Akiyama et al., 2011a, 2011b) and the gene encoding this protein was induced more than 2-fold. The genes for the TF Pap1 and the alpha-trehalase Ntp1 were also induced, 1.6-fold and 1.2-fold, respectively, but not flagged as being statistically significant. It is known that the genes for these two proteins are also induced in response to stress. The gene for the MAP kinase kinase protein active in stress signaling was significantly induced at 1.9-fold and the gene for the stress response regulator kinase Mcs4 was induced 1.3-fold, but it was not significantly different. The genes *sty1⁺* and *wis1⁺* have been identified as encoding proteins involved in maintaining cells in a quiescent state (Sajiki et al., 2009). From a screen for cell shape and/or survival, they identified a total 38 proteins that appear to be necessary for maintenance of cellular quiescence in *S. pombe*. In our study 10 of these genes were induced (4 significantly) and the vast majority of others were unchanged in expression. A few were slightly repressed, but only *ypt1*, which is a GTPase involved in vesicular transport from ER to Golgi, was significantly repressed at 1.8-fold. The onset of cellular quiescence in *S.pombe* is triggered by stress, in particular by nutrient limitation such as the lack of nitrogen. I observed that a specific set of TFs was significantly induced and this included the stress response gene *atf1*. I can speculate that the function of the other induced TFs is to induce the expression of the UR set of genes. For example, general stress-related genes were also induced, such as those encoding the short-chain dehydrogenases SPCC663.08c

and Osr1 and the glutathione reductase Gst2. In contrast, the set of gene encoding proteins involved in the RNA polymerase II mediator complex were DR suggesting that the set of genes affected by this complex could be actively repressed and the mechanism remains to be identified.

The other types of gene induced in my study were those encoding proteins that are plasma membrane resident or pass through the secretory pathway, such as transporter proteins and glycosylated proteins. The iron permease Fip1 and the iron transport multicopper oxidase Fio1 are known to physically interact (Askwith and Kaplan, 1997) and both of these genes are two of the highest UR genes. Therefore, there are examples within the data of genes with similar functions being co-induced. From the gene expression data I can speculate that disruption of Golgi function leads to diminished protein transport to the plasma membrane. The effects are to decrease transport processes leading to starvation and responses including the induction of genes to alleviate these process and in the meantime induce a quiescent cellular state. These hypotheses are presented in more detail in the overall discussion of the topic presented in Chapter 6.

Chapter 5

Analysis of structural features of introns and their expression in fission yeast based on the *pcs3⁺* gene family

5.1. INTRODUCTION

The regulation of gene expression has mainly been investigated from the aspect of the production of mRNA. So traditionally, transcription factors that bind to RNA polymerases and either activate or prevent production of mRNA have been intensely researched. For the expression of protein-coding genes, the end-point is not really the production of mRNA, but the making of the protein. In this sense, gene expression includes all the processes from gene to protein, including transcription, mRNA modification and degradation, translation, post-translational modification and even protein turnover. For mammalian systems, estimates place around 40% as the upper limit for the number of proteins that whose levels are strictly controlled by mRNA levels (Tian et al., 2004). Therefore, many factors must be considered to understand how gene expression leads to protein levels (Mata et al., 2005).

The regulation of gene expression by non-coding RNAs has become a very intense topic of research within the last few years (Griffiths-Jones, 2007). Non-coding RNAs (ncRNAs) are transcribed, but do not code for proteins. It is well known that ribosomal RNAs (rRNAs) help form the peptide bond during protein translation (Gutell et al., 2000). Small nuclear RNAs (snRNAs) are critical components of the large spliceosome complexes that remove introns from heteronuclear pre-mRNAs (Will and Lührmann, 2011). snoRNAs are involved in the processing of rRNAs, tRNAs and snRNAs and help in their correct folding (Fayet-Lebaron et al., 2009b; Samarsky et al., 1999b). In higher eukaryotes, micro RNAs (miRNAs) negatively control gene expression post-transcriptionally by adjusting mRNA levels through their degradation or by affecting protein translation (Lee et al., 1993). In eukaryotes, it was demonstrated that exogenously supplied small, double-stranded RNA can affect protein amounts when it is transferred into cells (Fire et al., 1998). In mammals it has been reported that snoRNAs can also control the alternative splicing of mRNA (Kishore and Stamm, 2006b). Introns can be considered non-coding RNAs and they often they have a function, e.g., they encode small RNAs. This means that after the intron is expressed as part of the pre-mRNA it is removed by splicing and it remained within the cell and processed further to the size of small ncRNA. The range of size for introns in *S. pombe*

is between 30 and 700 bp, with the average being around 80 bp (Wood et al., 2002a) and around 50% of genes contain multiple introns (Prabhala et al., 1992). The consensus sequence for splicing in *S. pombe* is similar to mammalian splice sequences and introns in mammal mRNAs have been shown to be correctly spliced out in *S. pombe* (Käufer et al., 1985).

It has been proposed that the last common ancestor of fungi was an intron-rich organism (Stajich et al., 2007). Many studies pointed out multiple intron loss events but there is very little evidence for intron gains (Rogozin et al., 2012). Taken together those observations raise the question, since introns are becoming extinct in fungal genomes, what is the significance of the intron currently remaining? Introns might be either evolutionary remnants and they will be eventually lost or they are kept because they have some function. Introns can function in multiple ways, such as through their sequence as a binding region for proteins, their secondary structure as a splicing regulating structure or as a separate ncRNA, or the process of splicing itself can affect the gene expression by affecting the timing of mRNA processing. Here, I will consider the possible functions fulfilled by introns after splicing via RNA secondary structure. Normally intron lariats are quickly degraded (Conklin et al., 2005), but secondary structures can protect intron lariats from degradation after splicing. Since RNA secondary structures are based on base-pairing rules, it is possible, although not trivial, to computationally predict what structures an intron sequence might adopt.

As a side part of my thesis project, and to obtain training in computational biology, I investigated the possibility that introns within the *S. pombe psc3⁺* gene may produce functional small RNAs that may have regulatory function. This project stems from a previously published genetic characterization of the *psc3⁺* gene of *S. pombe* (Ilyushik et al., 2005). The gene *psc3⁺* encodes a subunit of the cohesion complex that holds sister chromatids together during metaphase of cell division. Psc3 works during mitotic cell division and it has a paralog Rec11 that functions in meiosis. Determination of the gene structure of the *psc3⁺* gene showed some interesting intron features. One of the 5' *psc3* introns is located towards the 3' end of the gene with a large gap separating it from 3 5' introns. This is unusual because normally we observed more introns retained at the 5' of mRNAs according to the reverse-transcription mediated intron loss, 3' introns are most likely to be lost first (Mourier and Jeffares, 2003). It appears that in *psc3⁺*, the middle introns have been lost but the distant 3' one was retained. Does the retention of the 3' intron of *psc3⁺* signify a possible function for the 3' *psc3⁺* intron? In this chapter, I present the results of computer-based experimentation to investigate possible stable structures in introns of *psc3⁺* and its paralog *rec11⁺*. I describe the required bioinformatics steps to obtain and align orthologous intron sequences from

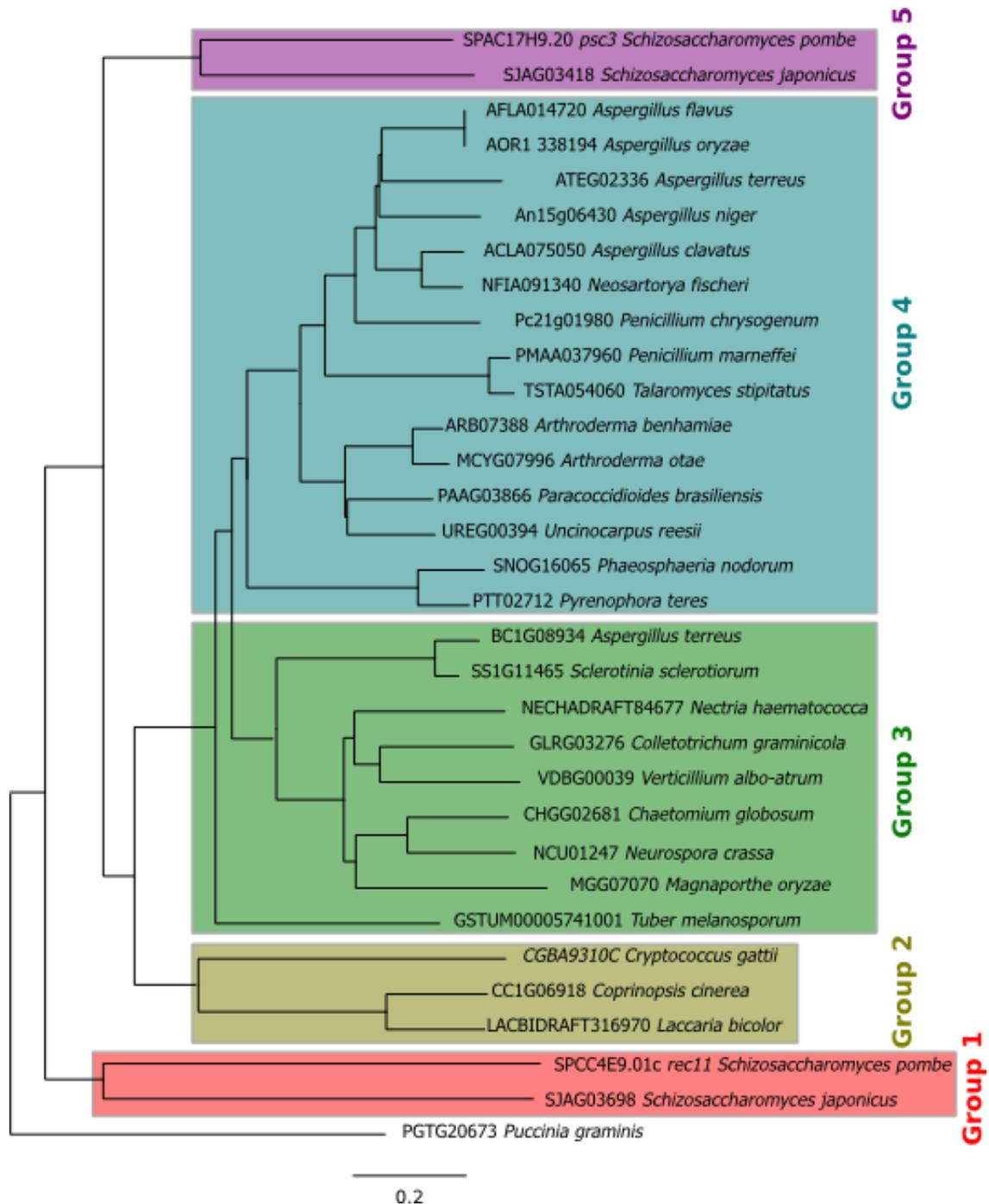


Figure 5.1: Phylogenetic tree of cohesion gene orthologues in fungi. The *psc3* and *rec11* proteins sequences were combined into one set and the phylogenetic tree determined using MEGA5 (Tamura et al., 2011). Group 1 contains the two *rec11* sequences and Group 2-5 contain *psc3* sequences.

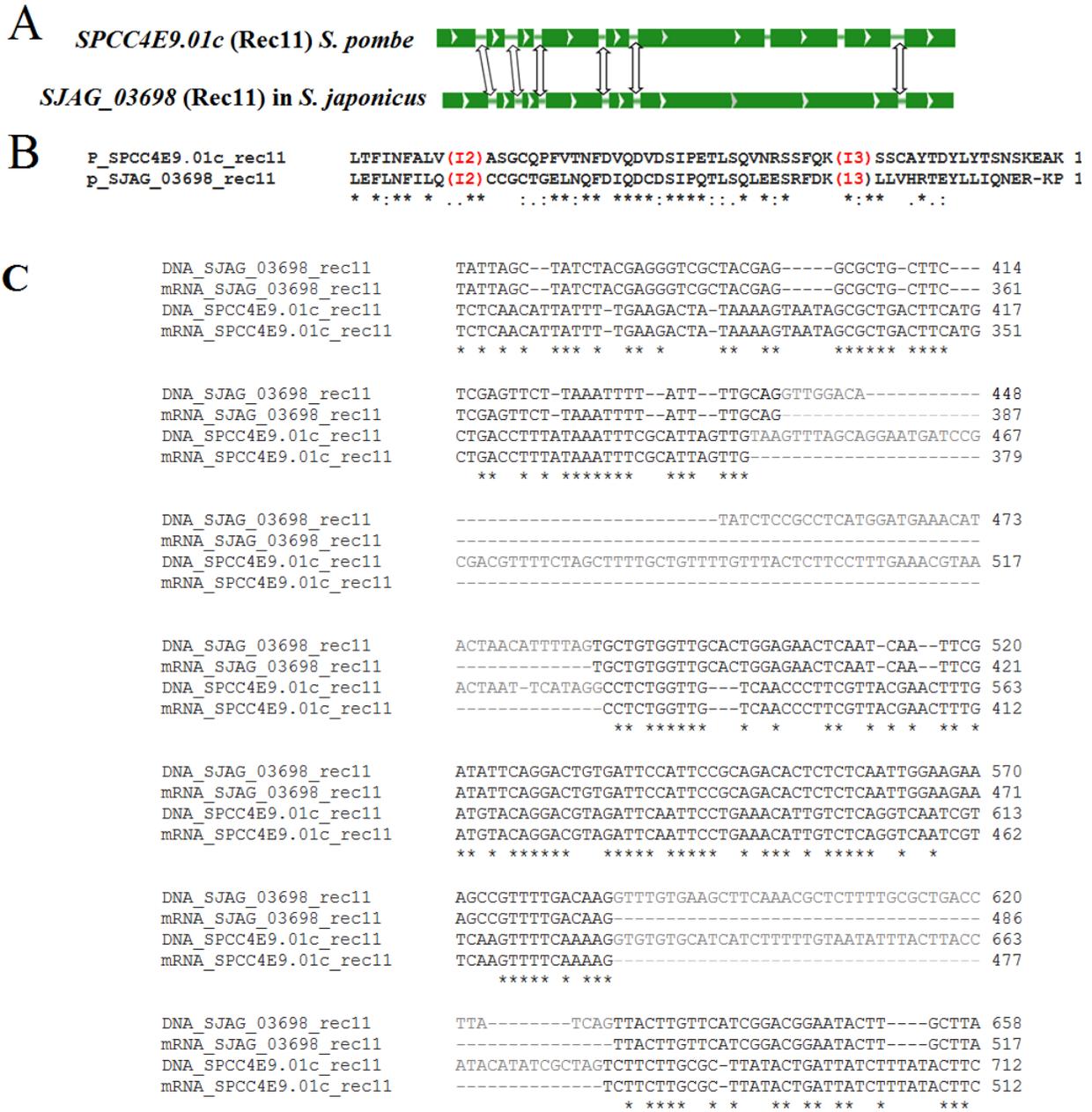


Figure 5.2: Testing for intron orthology using Group 1 as per Figure 5.1 (A) Assignment of intron orthologs based on the gene structure. (B) The alignment of proteins in Group 1 confirms the placement of intron 2 and 3. (C) Sites of intron 2 and 3 confirmed by mRNA and DNA alignment.

resources. After the RT-PCR screen for remaining intron transcripts of these genes proved negative, I adopted a new approach of using RNAseq of small RNA libraries to see if any full introns were expressed in *S. pombe*. Only a few were, but I did observe the expression of a number of new features, mainly snoRNAs, that had not been previously annotated in the *S. pombe* genome.

5.2. RESULTS

5.2.1. Identifying orthologous introns in fungal cohesins

The first step in the analysis was to find fungal orthologs of the protein Psc3. The general taxon ‘fungi’ (taxid:4751) was searched with Psc3 as query using BLASTP (Altschul et al., 1990). Most returns only specified cohesin, so the top 50 hits were selected and found to contain mainly Psc3 homologues, but also the Rec11 paralogs of both *S. pombe* and *S. japonicas*. This was fortuitous, because the *rec11*⁺ gene with 8 introns would serve to develop the sequence of bioinformatic steps for analysing the *psc3*⁺ introns. The gene sequences of these proteins were retrieved and 18 rejected, because their genes either did not contain introns or the sequences were duplicates. Using the tools in NCBI, the sequence relationship among the remaining genes was performed as a group. In order to facilitate identifying orthologous introns and afterwards their secondary structure, we divided the structural analysis according to each of the five groups identified (Figure 5.1).

Using the two *rec11*⁺ genes of Group 1 as an example, the process of finding orthologous introns within each group was as follows. Group 1 contained the two *rec11*⁺ genes from *S. pombe* and *S. japonicus*. Frequently, inspection of the gene structure is sufficient to establish the intron orthologs. The two genes had six orthologous introns that appeared to align and there were two introns in the *S. pombe* gene that were not present in *S. japonicus* gene (Figure 5.2A). The correspondence of introns was confirmed by aligning the introns to the protein sequences as orthologous introns would map to the same places in the protein sequences (Figure 5.2B). The final way to ensure that the orthologous introns had been correctly defined was to align DNA and mRNAs for all the genes in selected group (Figure 5.2C). The same procedure of finding orthologous introns was employed to all the *psc3*⁺ groups (2-5).

5.2.2. Alignment of orthologous introns and secondary structure prediction

As shown in Figure 5.2, Group 1 has two genes for *rec11*⁺ which have six orthologous introns. Since this project was performed, other genome sequences of the species *Schizosaccharomyces* have been sequenced, namely: *S. cryophilus* and *S. octosporus*. Therefore, I compared *rec11*⁺ in those four species using the annotation provided by Ensembl Fungi. First, ClustalOmega was used to align each orthologous

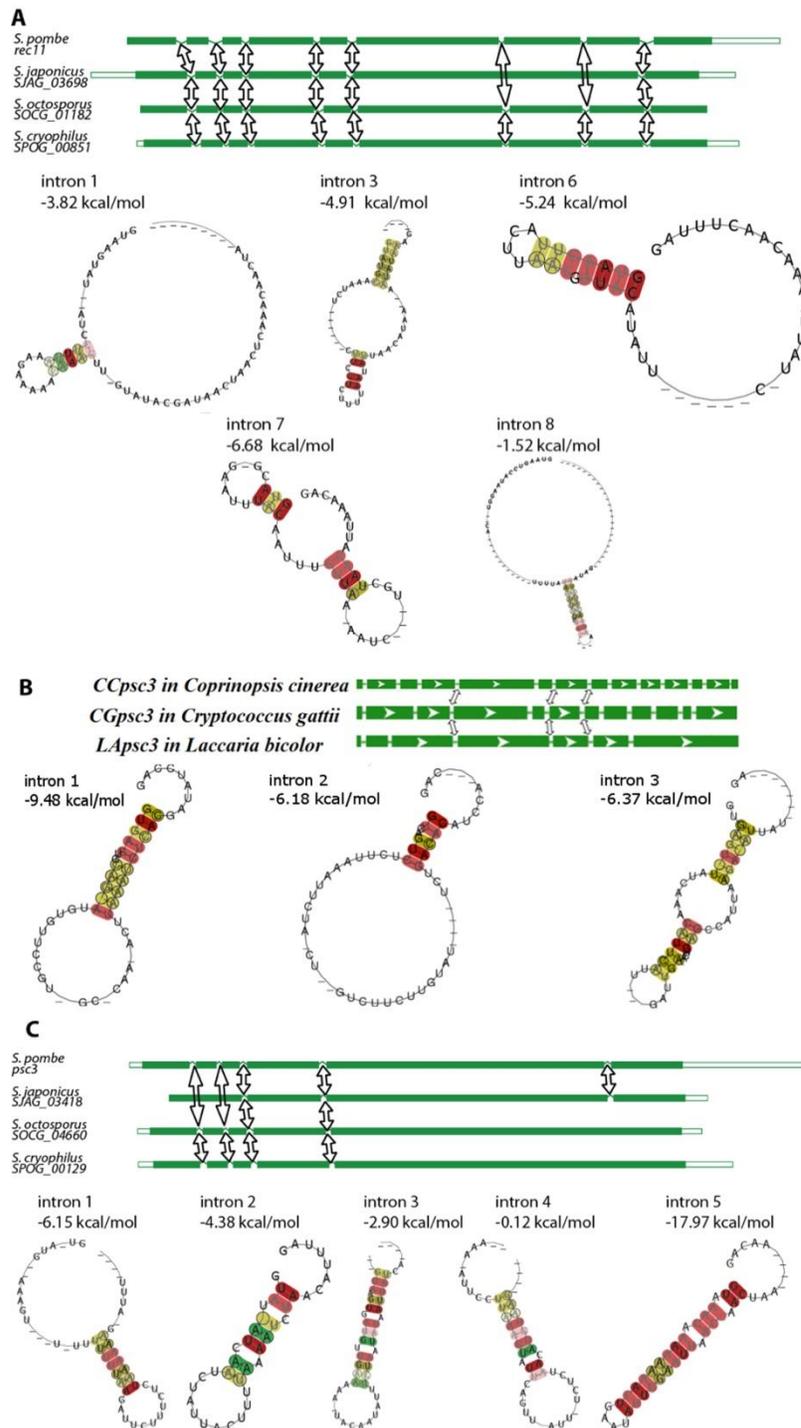


Figure 5.3: Secondary structures for the orthologous introns. Each figure A-C shows the gene alignment with the orthologous intron designated by an arrow. Introns are numbered from left to right. Alignments were done using either ClustalW or ClustalOmega and structures were determined using RNAalifold. The structure of each intron for a group is provided as well as the MFE of folding. (A) Group 1 *rec11* genes of *Schizosaccharomyces* sp. (B) The *psc3* gene of Group 2 as shown in Figure 5.1. (C) Group 5 *psc3* genes for *Schizosaccharomyces* sp.

intron sequences and then the secondary structure for each alignment was predicted by RNAalifold web server (Figure 5.3). The intron positions among the four species were well conserved but not their sequences. This results in predicating RNA structures with a relatively high MFE values. The lowest was only -6.8 kcal/mol for intron 7 (Figure 5.3A). Group 2 contained the *psc3*⁺ genes from three different organisms, which were *Coprinopsis cinerea* (*CCpsc3*), *Cryptococcus gattii* (*CGpsc3*) and *Laccaria bicolor* (*LApsc3*). This group of genes from these species had three orthologous introns (Figure 5.3B, panel a). Different introns have been lost in each of the species, but in *L. bicolor* we see an increased intron loss at the 3' end of the gene, which is consistent with the proposed cDNA reverse-transcriptase intron-loss model (Fink, 1987). The relatively small number of orthologous introns packed into the 5' end of the gene and coupled with the low sequence and structure conservation pointed to the conclusion that the introns in this group are probably not functional. This was supported by the MFE scores for the three introns with Intron 1 having the lowest MFE of -9.8 kcal/mol (Figure 5.3B). Group 3 had nine genes from different organisms, but these only shared two orthologous introns. Group 4 consisted of one orthologous intron among the 15 genes in the various organisms in this group. The low number of orthologous introns and their placement near the 5' end of the gene strongly suggested that this intron would not have a function. The structures for the orthologous introns for Groups 3 and 4 were determined and the average MFE scores for the introns of these two groups were on average 0.00 kcal/mol. The sequences of Group 5 were expected to be more similar than groups 2-4, since the four species come from the same genus. For reference, the systematic ID of the *psc3*⁺ gene for each of the *Schizosaccharomyces* species is given in Figure 5.3. However, of the five introns of *S. pombe Psc3*⁺ only introns 3 and 4 were conserved in all four species. Introns 1 and 2 are missing in *S. japonicus* and intron 5 is only present in *S. pombe* and *S. japonicas* (Figure 5.3C). Structure predictions were made for each of the introns according to the orthologous pairing. It was interesting that only for the 5' intron of *psc3*⁺ of *S. pombe* and *S. japonicus* was a structure returned where the MFE value at nearly -18 kcal/mol was low enough for it to be considered a stable enough structure to remain in the cell.

5.2.3. Investigation of intron expression by RT-PCR

The structural prediction of the stability of intron 5 was certainly a good basis by which to proceed with the experiments to determine if it was expressed. The logical approach was to see if it could be amplified by RT-PCR. The three different sets of primers were designed to determine the expression of intronic sequences of the *psc3* gene and to distinguish them from unspliced and spliced transcripts. For example, amplifying just the intron may yield a product from a small number of pre-mRNAs that had not yet been spliced or even contamination from DNA. One pair of primers was designed so that the oligos sat in the exons on either side of the intron, which I called the exon primers. Another set was designed only to

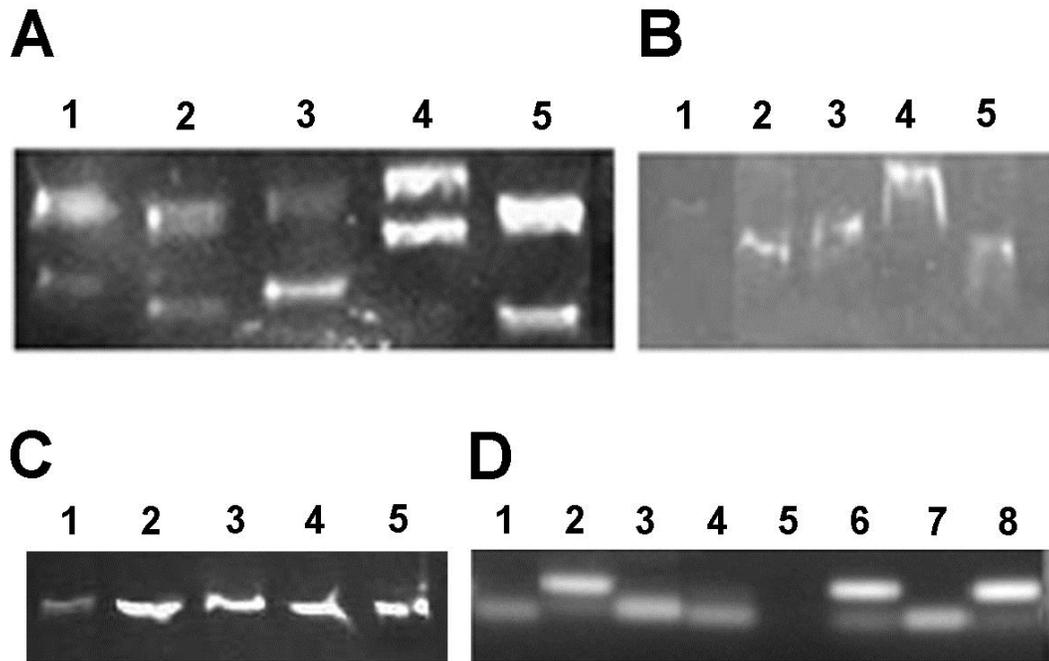


Figure 5.4. Gel-based verification of product amplification by RT-PCR. (A) Fragments from reactions conducted on cDNA using primers spanning the each of the introns from the *S. pombe psc3+* gene. The approximated fragment sizes are given in the text. (B) Fragments from reactions on cDNA made from DNase-treated RNA. (C) Junction primers for *S. pombe* gene *psc3+* tested by normal PCR on genomic DNA. The fragment sizes are given in the text. For panels A-C, the lane numbers 1-5 correspond to the intron number as specified in Figure 5.3C. The bands were resolved on 15% polyacrylamide gels. (D) RT-PCR of known ncRNAs in cDNA samples used in panel B. Lanes 1 & 2 are the primer pairs for SPNCRNA.184 with fragment sizes of 45 and 96 bp, respectively. Lanes 3 & 4 are the primer pairs for SPNCRNA.185 with fragment sizes of 55 and 92 (expected) bp, respectively. Lanes 5 & 6 are the primer pairs for SPNCRNA.186 with fragment sizes of 47 (expected) and 106 bp, respectively. Lanes 7 & 8 are the primer pairs for SPNCRNA.487 with fragment sizes of 49 and 103 bp, respectively. Bands were separated using a 2% TAE-agarose gel.

amplify the intron sequences of each of the five *psc3*⁺ introns. A third set was designed where the forward and reverse primers spanned the intron/exon junctions 3' and 5' on each side of the intron. This set of primers would only amplify unspliced pre-mRNA transcripts. A group of primer pairs was made to amplify a number of ncRNAs that are postulated to be expressed from loci near to *psc3*⁺. These primers were designed to give product sizes approximating those of the *psc3*⁺ introns sizes.

The exon primer pairs, when used against cDNA made from total RNA from wild-type cells as template, were expected to give fragment sizes of 60 bp, 52 bp, 57 bp, 81 bp and 52 bp for introns 1-5, respectively, for mature mRNAs. Bands of these fragment sizes were observed, but I also observed bands for fragments of larger size (Figure 5.4A). As mentioned above, bands of larger fragment size would be present if either DNA had contaminated the RNA samples or there was unspliced transcript present. Either way, the larger bands would give fragments of sizes of 105 bp, 90 bp, 105 bp, 131 bp and 100 bp for introns 1 to 5, respectively, and these corresponded to the sizes of the upper bands visible in each lane. It was found that DNA had contaminated the RNA preparations, because DNA treatment of the RNA samples prior to cDNA synthesis removed the upper bands (Figure 5.4B). The presence of only one band of the predicted size for the cDNA indicated that unspliced transcript was not present and the mature cDNAs resulted from perfect splicing. The junction primers, when tested on genomic DNA, gave the predicted fragment sizes of 65 bp, 58 bp, 67 bp, 70 bp and 68 bp for introns 1 to 5, respectively (Figure 5.4C), but the cDNA samples from DNase-treated RNA gave no bands. This confirmed that unspliced pre-mRNA for *psc3*⁺ was not present in the samples. Finally, and most importantly, the intron primers did not give amplification products and so no data was shown. As a final control, I investigated if the RT-PCR would amplify ncRNAs that had been annotated in the *S. pombe* genome. I selected and designed primer pairs for three ncRNAs that were located on Chromosome 1 in the vicinity of *psc3*⁺, *SPNCRNA.184*, *SPNCRNA.185* and *SPNCRNA.186*, and one ncRNA located on Chromosome 3, *SPNCRNA.487*. For each ncRNA, two primer pairs were made, and they were designed to give products of approximately the size that would be expected of an average intron, around 80 bp (Wood et al., 2002a). Six of the eight primers pairs appeared to work for the ncRNAs, except for primer pair 2 of *SPNCRNA.185* and primer pair 1 for *SPNCRNA.186*. This indicates that in our wild-type cells, ncRNAs are generally expressed and in sufficient quantities to amplify using standard RT-PCR (Figure 5.4D).

5.2.4. Investigating the potential for intron expression by RNAseq

Since the RT-PCRs did not confirm the expression of *psc3*⁺ introns, I decided to take a global approach and employ RNAseq to assess the expression of *S. pombe* introns. For that I requested from the sequencing facility at GeT-Biopuce in Toulouse, France to create a size-fractionated 'Small' libraries.

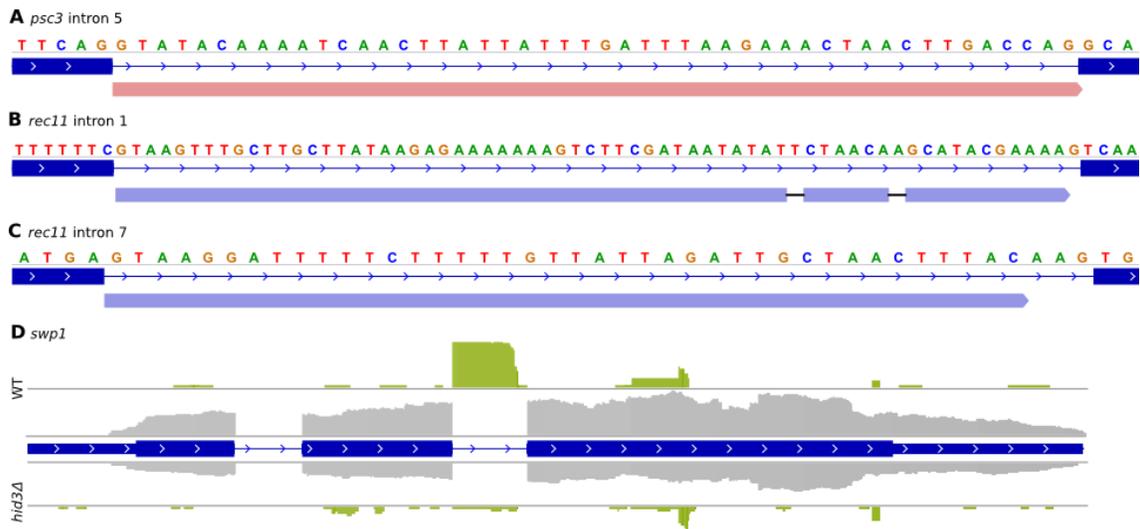


Figure 5.5. Intron expression in RNaseq performed on small RNA library. (A) The only read mapping to the *psc3* intron. (B) (C) Reads mapping to *rec11* introns. (D) Example of a highly expressed intron of *swp1*. Dark blue boxes represent exons of genes with thin blue line connecting them representing the introns. Grey histograms show the expression of the full gene *swp1* in WT and *hid3Δ* RNaseq done on polyA selected RNA, whereas the green histogram shows the expression in the size-selected small RNA library.

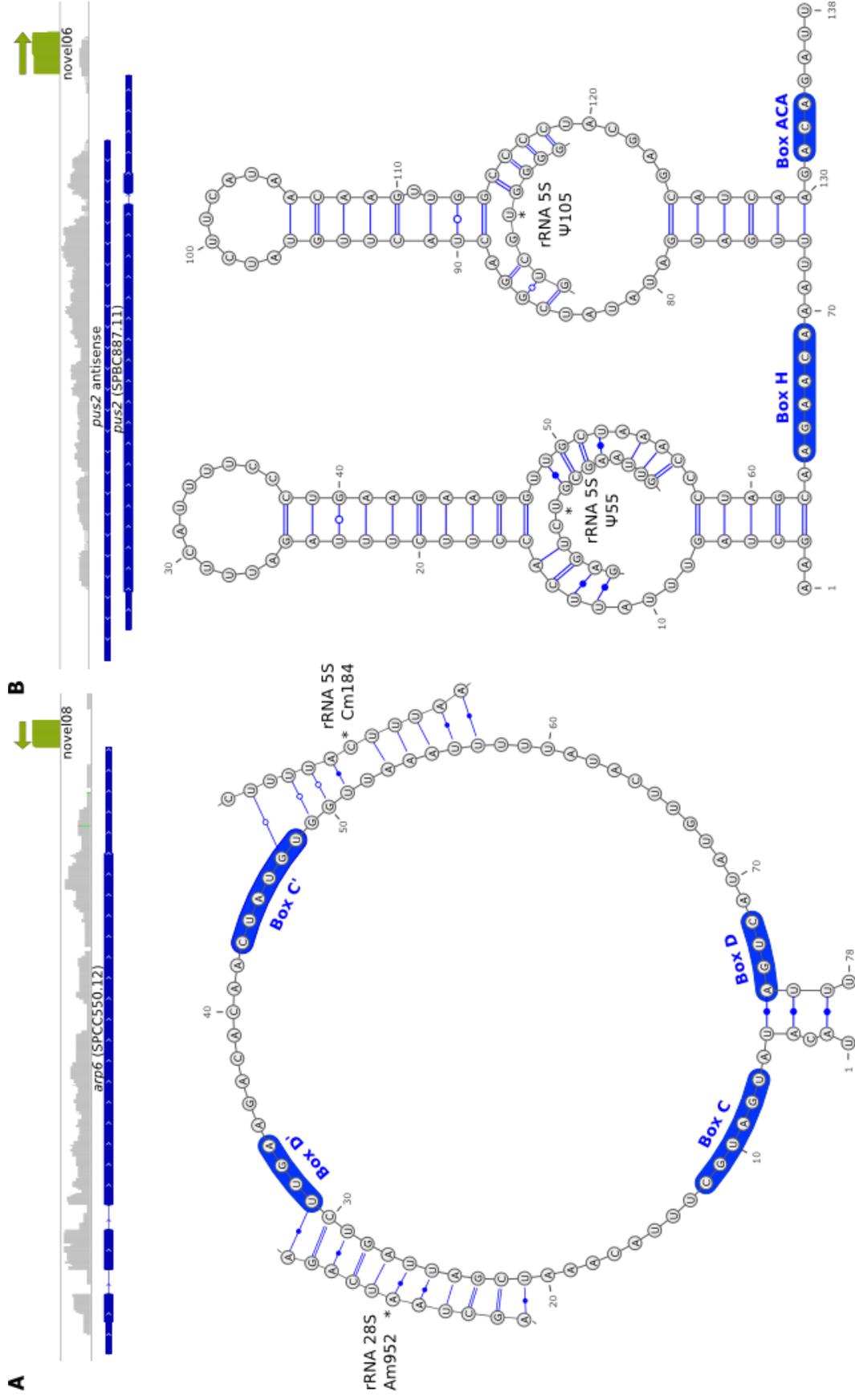


Figure 5.6. Examples of novel expressed transcripts identified as snoRNAs. (A) Expression of the novel08 C/D box snoRNA with transcripts shown in the small (green) and total (grey) RNAseq libraries (top panel). The bottom panel shows the common structure of the snoRNA and putative interactions with the rRNA. The blue boxes indicate the C and D boxes within the snoRNA (A) and its putative interaction with 5S rRNA. The blue boxes show the H and ACA boxes. (B) Expression of the novel06 H/ACA box snoRNA with transcripts shown in the small (green) and total (grey) RNAseq libraries (top panel). The bottom panel shows the common structure of the snoRNA and putative interactions with the rRNA. The blue boxes indicate the H and ACA boxes within the snoRNA (B) and its putative interaction with 5S rRNA. The blue boxes show the H and ACA boxes.

Two libraries were made, one from wild-type and one from *hid3Δ*, with a mean RNA molecule size of 80 and 100 nt, respectively, which were then sequenced using an IonTorrent system (Life Science Technologies). In the wild-type library, I found only one read mapping to the intron 5 of *psc3* (figure 5.5A). No reads for other *psc3* introns were observed. For *rec11* I found two reads one for intron 1 and one for intron 7 (figure 5.5B-C). This observation supports the results from the RT-PCR that the introns of both *psc3* and *rec11* are most quickly degraded after splicing. In fact, I observed only a few introns where there were sufficient reads to consider that they were expressed (Figure 5.5D). The best example was that for the gene encoding Swp1 (SPCC553.06) a predicted oligosaccharyltransferase delta subunit. The second intron of this gene had transcript remaining whereas no transcript was detected for the first intron. This was observed in both wild-type and *hid3Δ* samples and the reads for the intron were around half of those observed for the gene in the wild-type and *hid3Δ* samples shown in Chapter 4 determined by classic poly A+ RNAseq for quantification of mRNA. The reads for the intron were not observed in the classic analysis. Interestingly, the high expression of this intron was only observed for wild-type and not the *hid3Δ*.

Through visual inspection of the mapped reads on each chromosome using IGV viewer, I noticed high levels of expression for distinct small regions (Figure 5.6A-B). These distinct regions likely represented snoRNAs. Additionally, there were 36 novel expressed regions with a very clear start and end sites suggesting of a consistent production of novel ncRNAs. Like for *swp1*, the vast majority of the novel identified transcripts were either not expressed in RNAseq performed on polyA+ RNA or it was difficult to recognise them as separate transcripts, because of the presence of other expressed genes. However, the comparison with the expression of polyA library allowed us to find out which of the transcript might reside in the introns of either protein coding genes or other ncRNAs. I found that 16 transcripts came from the introns of other genes, which was expected for snoRNAs, however, some exonic snoRNAs were also observed. Some snoRNAs came from predicted intergenic regions (Figure 5.6 A-B).

To investigate the function of the novel ncRNAs found in the RNAseq, I used two common programs to investigate the structure of the snoRNAs. One called snoscan (Lowe and Eddy, 1999) predicts C/D-box snoRNAs and of the 36 novel transcripts, 16 of them were predicted to have this motif C/D box snoRNAs. The novel C/D box snoRNA that I call novel08 was found near the actin-like protein gene *arp6* (Figure 5.6A, top panel). It has the classic loop structure to interact with rRNAs such as the 5S and 28S rRNAs (Figure 5.6A, bottom panel). Then, I used BLAST to search for the orthologs of the remaining 20 novel transcripts and found that five of them were conserved in all four *Schizosaccharomyces* species. Among these five well-conserved and expressed ncRNA transcripts, there

were two ambiguous and one very probable H/ACA box snoRNA according to the program RNAsnoop (Tafer et al., 2010). The *S. pombe* ortholog of this H/ACA box snoRNA was found down-stream of the *psu2* gene that encodes a putative beta-glucosidase (Figure 5.6B, top panel). The predicted structure of this snoRNA has the classic two-hairpin secondary RNA structure and H and ACA boxes for 5S rRNA binding (Figure 5.6B, bottom panel).

5.3 DISCUSSION

Secondary structural predictions and assessment of expression of introns of the *psc3*⁺ and *rec11*⁺ genes of *S. pombe* were started based on the interesting gene structure of the *psc3*⁺ gene where the terminal 5' intron has been retained. The structural prediction only gives an idea if an RNA sequence can adopt a structure stable enough to avoid degradation. It does not say anything about predicting potential function as a regulatory ncRNA. Therefore, there are certain conditions by which RNA structures are usually determined in order to increase confidence that a predicted structure is biologically real. The search for conserved intron orthologs is essential because an RNA that is regulatory is likely to be conserved, at least within closely related species. The more distantly related the organisms are for which orthologs for an ncRNA can be found give greater confidence for a biological role. Therefore, most efforts in the structural prediction of intron structure for *S. pombe psc3* introns were spent identifying the orthologous introns in other species. Once the orthologs are found, structures for the individual ncRNA sequences can be compared to a structure for the consensus sequence. If they match, then it is likely that this structure has been evolutionarily conserved for a regulatory function (Griffiths-Jones, 2007; Hooks and Griffiths-Jones, 2011).

It was interesting that for all 13 introns comprising the *psc3*⁺ and *rec11*⁺ genes, the only intron with a potentially stable structure was the unique 3' intron of *psc3*⁺. Assignment of a stable structure to this intron has to be taken with caution. First, the structure was determined using RANalifold, because it was the only one that will output structures and scores for short RNA sequences. In general, RNA structures should be measured using a combination of prediction programs to ensure that the predicted structure is real. Second, there were only the two intron sequences for *S. pombe* and *S. japonicus* used for the consensus sequence. Because they were similar, individual structures would compare well against the consensus and give a lower MFE score. Third, the question still remains as to how significant a predicted structure is from a biological point of view if it is present in only one or closely related organisms. However, the loss of 5' introns in the *S. japonicus psc3* gene compared to *psc3*⁺ of *S. pombe* suggests that the close evolutionary relationship between the two organisms is not a reason for common retention

of the unique 3' intron. Based on the latter two points, it was necessary to determine if transcript for this unique intron was present in *S. pombe*. Although the test by RT-PCR was negative, it did not mean that the intron is not expressed. The gene *psc3*⁺ is generally weakly expressed during the cell cycle, so it is likely the intron would also be in low quantities if retained (Ilyushik et al., 2005). Also, it was not possible to test expression in a range of conditions where the *psc3*⁺ gene may be expressed, such as under stresses. The final possibility that could have been tried was to attach adaptors to the RNA and to try to amplify the introns using specific primers overlapping the junction between intron end and the adaptor. However, this would not work if the introns were processed in any way. Intron processing could explain why the intron primers themselves didn't work.

Therefore, it was decided to take a more global approach using RNAseq to investigate if introns in general, including those for *psc3*⁺, were present. Since pombe introns are on average quite small at 80 bp, it was decided to increase the possibility of detecting short RNA fragments by analysing libraries made from small RNA with insert sizes around 80 bp. Although the sequencing did give reads corresponding to the intron sequences of the *psc3*⁺ and *rec11*⁺ genes, they were very low. Thinking positively, it still remains to be determined if the 3' intron of *psc3*⁺ is expressed at some point. An interesting finding from the RNAseq was finding more than 30 non-annotated potential snoRNAs. Some of these clearly had the feature of C/D and H/ACA boxes. C/D box snoRNAs are involved in rRNA processing through methylation, which can stabilise rRNA structures by strengthening individual hydrogen-bonded base pairs and H/ACA boxes process rRNAs through modification of uridine bases, which is thought to stabilise the overall structure of the rRNA (Helm, 2006). In fact it is not clear what affect the majority of rRNA modifications are doing, so it is difficult to discuss potential roles for the novel snoRNAs here. Further characterisation would require deleting them and observing effects on cell physiology and protein and gene expression.

Chapter 6

General Discussion and Outlook

My work on investigating the function of the *hid* genes started from the idea that the down-regulated expression of the *HsHID1* gene could promote the development of cancer as proposed by Harada et al. (2001). The human gene has been referred to by the systematic designation of a gene of unknown function *c17orf28* that had persisted in databases, including Genbank, until about 2011 when there appeared a number of publications that proposed a function of the *C. elegans* orthologue *HID-1*. At the time of the Harada publication, the orthologue in *C. elegans* had not yet been identified, and a study by Ailion and Thomas (2003), identified the *hid-1* gene through a screen to identify mutants of *C. elegans* that formed dauer larva under the conditions of slightly elevated temperature that would not normally cause the nematode to take a dauer form. It was surprising that this gene had not been characterized previously, since even under normal conditions the mutant exhibited pleiotropic phenotypes, such as delayed defecation, uncoordinated movement, and mutants affecting these processes had been screened for extensively (Hart, 2006).

From what is currently understood about *HID-1* function in *C. elegans*, it should be possible to consider if *HsHID* could function as a tumour suppressor gene. Due to other types of mutants isolated as part of the screen, they proposed that *HID-1* functions in the insulin-like signalling pathway in *C. elegans*. There is more and more evidence showing the involvement of insulin signalling in tumorigenesis (Gallagher and LeRoith, 2011). The down-regulation of expression of the forkhead TF *FOXO4* has recently been shown to contribute to gastric cancer cell line growth and an increase in metastasis (Su et al., 2014b). The *C. elegans* orthologue of *FOXO4* is the forkhead TF *DAF-16*, which is a positive regulator of dauer formation. It has been identified in screens to identify mutant defective in dauer formation (Gottlieb and Ruvkun, 1994). The down-regulation or mutation of *daf-16* to promote growth is consistent with a potential role of *FOXO4* as a tumour suppressor gene. In *C. elegans*, the *daf-16* mutation suppresses the *hid-1* constitutive Dauer-formation phenotype. This means that they would appear to work in opposite directions with *HID-1* expression required for normal cellular growth, which would contradict an implied function as a tumour suppressor.

There is a regulatory step between *HID-1* and *FOXO4* that must be considered in order to completely address the question of contradictory modes of action. *SpHid3*, known in Pombase as *Ftp105*, retains the deubiquitinase *Ubp5* to the Golgi apparatus (Kouranti et al., 2010). The human orthologues of *SpUbp5* is

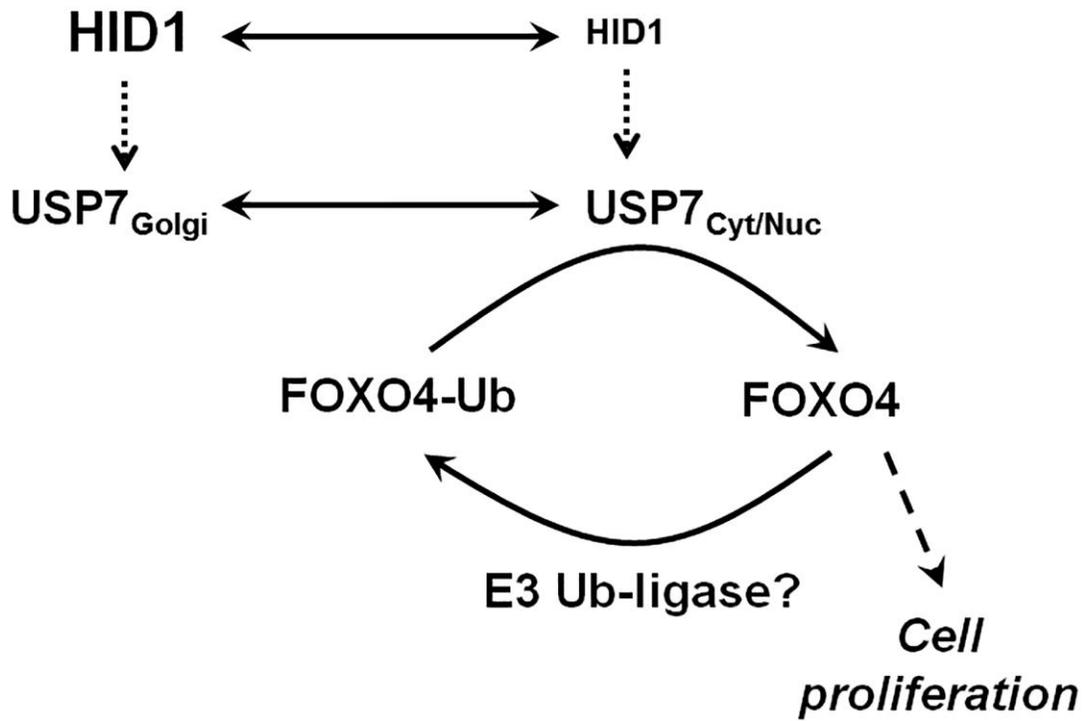


Figure 6.1. The regulation between HID1 and FOXO4. When HID1 levels go down, USP7 is released and can migrate to the nucleus to stop activation of tumour inhibiting pathways. For example of this through the human transcription factor FOXO4, which is necessary for inducing cell apoptotic pathways.

USP7, a deubiquitinase that interacts with FOXO4. FOXO4 is regulated by a system of monoubiquitylation/deubiquitylation in which FOXO4 is both ubiquitinated and deubiquitylated in response to oxidative stress (van der Horst et al., 2006). FOXO4 is ubiquitinated by an unknown E3 ligase within a few minutes after exposure to oxidative stress. Ubiquitination of FOXO4 leads to its recruitment by USP7 deubiquitinase in order to deactivate its transcriptional up-regulating properties. An imbalance of USP7 would ensure that FOXO4 is in an inactivated state and decreasing aspects of gene expression (Figure 6.1). This would potentially lead to the up-regulation of growth and metastasis. By this model lower *HsHID1* expression would favour tumorigenesis or proliferation.

A potential complicating factor to this model is that it does not seem clear that there is a specific functional homologue of FOXO4 in *S. pombe*. There are a number of forkhead-like TFs in *S. pombe*, such as the forkhead TFs Fhl1 and Fhk2, and analogues of the Akt/PI3K signalling pathway of insulin signalling in humans are present in *S. pombe*, such as Sck1 and Sck2, for the regulation of glucose metabolism. No forkhead-like TFs were altered on the gene expression level (or protein expression level) whereas there were clear increases in expression of certain groups of TFs (Chapter 4) in the *hid3Δ* mutants.

Another potential complicating factor is the ultrastructural modifications to the Golgi apparatus in cells lacking HID1 like those that I observed in *pombe* lacking Hid3 (Chapter 3). Modifying the structure and thus function of the Golgi would expect to alter the localization of other proteins besides UPS7. There is evidence that FOXO4 may interact directly with the STEROL RECEPTOR BINDING PROTEIN 2 (SREBP2) in order to stimulate cholesterol synthesis (Zhu et al., 2014). In the Introduction to my thesis I presented the relationship between SREBP signaling from the Golgi and effects on gene expression. In brief, SREBP2 is an ER localized protein that when cholesterol levels are low is transported to the Golgi and proteolytically cleaved, and which can then enter the nucleus to up-regulate genes in cholesterol biosynthesis. Proteolytic cleavage of SREBP2 in the Golgi is regulated by ubiquitylation/deubiquitylation processes (Stewart et al., 2011). The implication is that signaling processes, such as through SREBP, would likely be altered with dramatic Golgi structural changes. This would also be the case for the phosphatidylinositol-4-kinase effector protein GOLPH3 (Buschman et al., 2015). GOLPH3 is present in the *trans*-Golgi and help to promote vesicle formation through microtubule association (Dippold et al., 2009). DNA damage alters Golgi structure, which can be alleviated by over-expression of GOLPH3, therefore it is frequent in cancers that GOLPH3 is induced to prevent DNA damage induced apoptosis (Farber-Katz et al., 2014). The *S. pombe* orthologue is Vps74, but it does not change on either the mRNA or protein level in *hid3Δ*, but it would be interesting to see if Vsp74 reverses the effect of Hid3 removal.

It is clear that there are dramatic changes at the mRNA level in *hid3Δ* strains. Although it was not presented in Chapter 4, a survey of genes related to ubiquitin-mediated processes revealed that 15 genes are significantly DE (8 UR and 7 DR). Of the eight genes that were significantly up-regulated 5 were E3 ubiquitin ligases or deubiquitinases. Two E3 ligases were DR, *SPAC2F3.08* and *uch1* as well as the E2 ubiquitin ligase RAD6 homologue. Lower levels of Rad6 would increase transcription at the mating-type locus. This is another potential example of the link between Golgi function and chromatin remodeling.

It is plausible that in humans and *C. elegans* there are other proteins holding together the Golgi apparatus and so the loss of HID1 would not be so dramatic. As well as HID1, animals have DYM, which is a peripheral membrane protein associated with the *cis*- and *trans*-Golgi stacks (Dimitrov et al., 2009; Wang et al., 2011). Patients mutant in *DYM* have severe morphological deformity of Golgi structure, which can be reversed by the expression of *DYM* (Denais et al., 2011). This would suggest that in animals the presence of DYM is critical and a structural role of HID1 is less important. However, a structural role for HID1 cannot be excluded, since the Golgi can form discrete subsets of forms, for example on treatment with BFA (Nizak et al., 2003). DYM was found to localize with a dispersed form of mini-stack and not with the remaining core Golgi upon treatment of cells with nocodazole, a treatment that disrupts Golgi structure (Dimitrov et al., 2009). Loss of DYM leads to severe growth defects and can lead to mental retardation depending on the nature of the mutation. A role for HID1 in Golgi structure can be presented from a different point of view. One question is why is the lack of DYM not lethal? It is interesting to speculate that HID1 may complement to some degree the function of DYM in maintaining Golgi structure and in doing so allowing the flow of proteins through a modified Golgi structure. To my knowledge, in animal cells there has not been a report where both DYM and HID1 have been eliminated and the effects on cell survivability tested. Using *S. pombe*, a next step would be to express the human and *C. elegans* DYMs in *hid3Δ* in order to see if they can reverse the apparent Golgi structural deformities within the mutant. This may show if there is any functional overlap between the DYM and HIDs.

From my results on *S. pombe*, is it possible to address the question of the putative tumour-suppressive properties of HsHID1? The major physical attributes of the *hid3Δ* mutant are its slow growth and altered Structure, but normal cellular morphology (Chapter 3). I measured cell growth through cell counts, thus slow growth could simply be a result of cells dying prematurely. Denais et al. (2011) reported that the apoptotic index was greater in cells lacking DYM. Apoptosis was not spontaneous, but was induced by over-growing cell cultures. Since apoptosis was not seen in normal cells, but only in over-grown cell cultures, they concluded that cells without DYM were already stressed and so more susceptible to

apoptosis under unfavourable growth conditions. If Hid3 has a similar function in *S. pombe* as DYM does in animals regarding Golgi structure, then I cannot remove the possibility that lower cell counts in *hid3Δ* cultures were not due to a certain amount of cell death due to stress. This remains to be tested with viability assays. As I will discuss in more detail below, there is evidence that under normal growth conditions, there is expression of genes that are markers of stress, such as *atf1*, and that this stress would be present over the lifespan of the cell.

Another possibility is that the cells have entered into a type of quiescent growth state. Stress signaling in *S. pombe* appears to be an inherent part of the entry of cells into quiescence and maintaining this state (Sajiki et al., 2009). Besides delayed growth and induction of stress markers, cells lacking Hid3 show distinct molecular signs of a quiescent state. There has been a lot of interest recently into studying cell quiescence as a mechanism for prolonging life. There have been screens looking for mutants with increased lifespan in the situation of low glucose in stationary phase (Roux et al., 2010). Although *hid3+* was not identified directly as contributing to increased lifespan, their gene expression analysis showed the DR and UR of *git3*⁺ and *cka1*⁺, respectively, which could contribute to increased lifespan. In *hid3Δ* cells grown under normal conditions, both genes were expressed less than the 2-fold cutoff to be included in the tables of differentially expressed genes, but they were significantly altered. Git3 is a glucose-responsive G-protein coupled receptor that when lost leads to increased lifespan (Roux et al., 2009). Git3 expression is discussed more below. Cka1 is a serine/threonine protein kinase that extends lifespan of normal, non-quiescent cells when over-expressed (Roux et al., 2010). In addition, it has been reported that the lack of Hid3 may increase the lifespan of *S. pombe* in response to nitrogen starvation (Sideri et al., 2014). Protein synthesis is also decreased in quiescent cells and the GO analysis of DE genes in Chapter 4 showed substantial DR of ribosomal protein genes. Also the transcriptional profiling shown in Chapter 4, there were a number of TFs that were induced specifically in *hid3Δ*. One of these *atf1*⁺ was mentioned previously as a marker of stress. *atf1*⁺ transcription is up-regulated by the MAPK stress-signaling pathway. The gene *wis1*⁺ that encodes the MAPKK in the pathway is also significantly induced in *hid3Δ* by nearly 2-fold. The gene *wis1*⁺ was shown to be necessary for both entry and maintenance of cellular quiescent state (Sajiki et al., 2009). Two other interesting TFs that were induced in *hid3Δ* are *gaf1*⁺ and *hsr1*⁺. In *S. cerevisiae*, the TORC1 signalling pathway promotes cell growth over quiescence in part by repressing the transcription of *GAT1*. The *S. pombe* orthologue of *GAT1* is *gaf1*⁺. The induction of *gaf1*⁺ in *hid3Δ* would suggest derepression due to inactive Torc1-mediated signalling. The TF gene *hsr1*⁺ is the *S. pombe* orthologue of the budding yeast TFs *MSN2* and *MSN4*. In budding yeast, Msn4p and Msn2p are repressed by both the PKA and TORC1 signalling pathways to promote cell growth (Gray et al., 2004) and there is evidence that the PKA pathway represses the expression of *MSN2* (Lenssen et al., 2002).

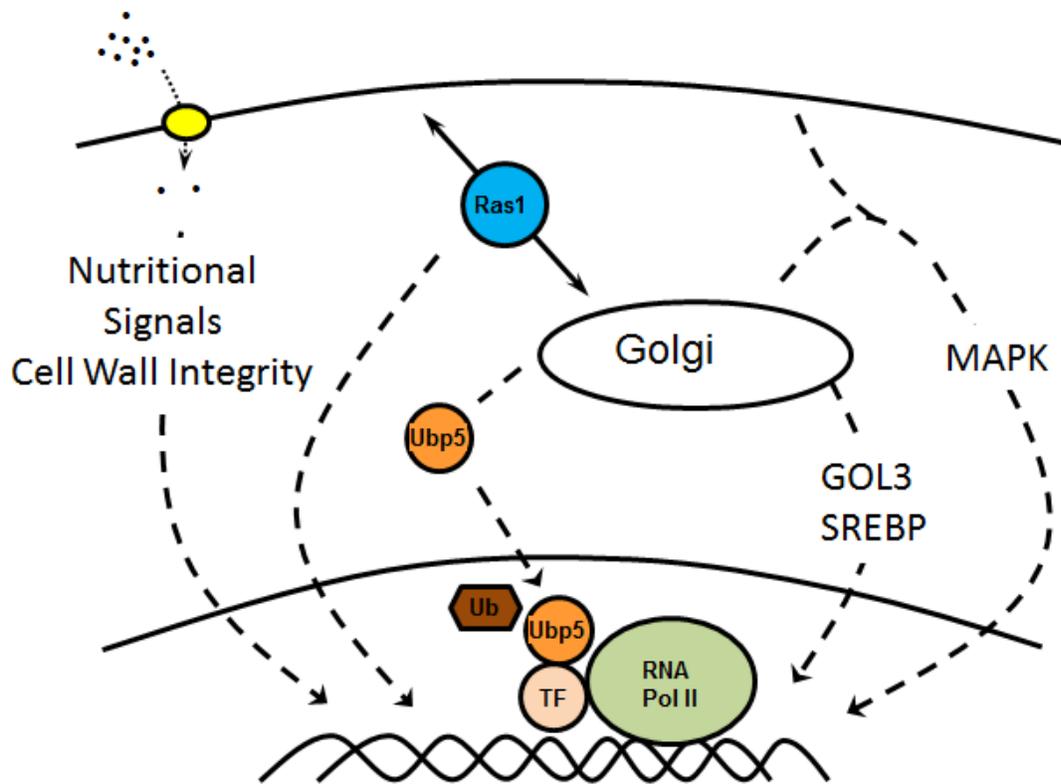


Figure 6.2. Transcription signalling pathways potentially affected in *hid3Δ*. Very restricted types of gene changed: transporters, stress, stress signalling. How does this relate to the variety of signalling pathways present?

If this system was working similarly in *S. pombe* then induction of *hsr1*⁺ is consistent with low activity of PKA-mediated cAMP signalling. The usual mechanisms by which human cells enter quiescence are the limitation of growth factors through withdrawal or serum depletion through prolonged growth, separation from other cells, or inhibition of cell contact signals (Coller et al., 2006; Parr, 2012). For unicellular microorganisms it is nutrient starvation (Gray et al., 2004). Where, it was previously thought that entry into quiescence was cell cycle arrest at G1, it is now apparent that entry into quiescence can take place at any part of the cell cycle and that this is an active process (Laporte et al., 2011). Nutrient starvation usually generates intracellular signals to try to access more of the missing nutrients. For example, iron starvation induces genes for iron transport proteins that would help to scavenge iron from the surrounding medium. This is true for a wide range of nutrients (Conrad et al., 2014). What was very striking from the gene expression analysis was the induction in *hid3Δ* of genes encoding a variety of transport proteins, including for iron, phosphate, and amino acids. Also common among the genes UR were proteins secreted or destined for transport to the plasma membrane. This indicates that disruption of the Golgi through lack of Hid3 causes disruption of processes at the plasma membrane and that this feeds back to increase gene expression in an attempt to produce more protein. There are a number of ways that *S. pombe* can signal changes occurring at the plasma membrane, including MAPK signaling, Ras/cAMP signaling, the Cell Integrity Pathway (CIP) pathway acting through Rho GTPases (Levin, 2011; Sánchez-Mir et al., 2014), and the Pmk1 MAPK signaling pathway through Protein Kinase C (Figure 6.2) (Zaitsevskaya-Carter and Cooper, 1997). In *S. cerevisiae*, nitrogen signaling is mediated through the nitrogen catabolite repression pathway requiring the action of the TF Gat1p (Conrad et al., 2014) and I observed the induction of the pombe orthologue *gaf1*⁺ in *hid3Δ* strains. It is interesting to speculate that the other induced TFs may have roles in the induction of starvation responses or other processes located at the PM. This is not an unreasonable idea. One point that demonstrates the specificity of induction of these set of TFs is that other TFs involved in transcription remained unchanged, for example the Mediator and the Swi/Snf complexes do not have any components that change in expression. However, the core mediator complex is quite stable, but fluctuates in component composition and activates transcription in relation to other TFs that interact with it (Malik and Roeder, 2010). The genes encoding cell surface proteins that are most highly UR in *hid3Δ*, *gsf2*⁺, *fta5*⁺, *plf6*⁺ and *pfl9*⁺, are regulated by the Mediator complex. The logical conclusion is that some of the induced TFs are working to regulate transcription of cell surface protein genes through the Mediator complex.

Although entry of *hid3Δ* into a quiescent cell state appears attractive, there are observations that have to be reconciled to equate slow cell growth with a quiescence state. In quiescent cells of *S. cerevisiae*,

cyclic-AMP (cAMP) levels remain low and signaling through the Protein Kinase A pathway is inactivated (Gray et al., 2004; Jiang et al., 1998). In addition, mutants of *S. cerevisiae* that cannot induce Protein kinase A-mediated signalling in response to cAMP exhibit a delayed growth phenotype (Jiang et al., 1998). However, I observed the significant induction of *cgs1*⁺, which encodes the regulatory subunit of PKA. This finding is difficult to interpret. Although it suggests increased cAMP-dependent signaling, it also contradicts cAMP signalling to serve as a negative feedback loop to repress the expression cAMP signalling genes in the presence of high glucose (Hoffman, 2005). In addition, Alasmari (2015) observed a large increase in Ras1 in *hid3Δ* and Ras1 activates the cAMP signalling pathway through PKA. However, Ras-mediated signalling of various pathways depends on its subcellular location as discussed below. Interestingly, mutant strains of *S. cerevisiae* in which active Ras2 is mis-localised from the plasma membrane to the cytosol also show a delayed growth phenotype

Related to Ras-signalling is the interesting question as to why *hid3Δ* cells are not affected morphologically when growth rate is? The answer for this comes from the observation by Alasmari (2015) that Ras1 protein levels are more than 3-fold greater in *hid3Δ* than other genotypes (Tamanoi, 2011). Ras1 activates two different pathways, the Byr1 pathway that is responsive to pheromones and controls mating and the Scd1/Ral1 pathway responsible for spindle formation and cytokinesis. The Scd1/Ral1 pathway is also important in determining cell shape. Cells lacking Ras1 are generally round as well as sterile. Although Alasmari (2015) did not see any obvious induction of the proteins in this pathway, no decrease was observed either. As long as protein levels are maintained, the pathways could be activated. One result suggesting preferential activation of the Scd1/Ral1 pathway is my observation that *erg25*⁺ is significantly induced (1.7-fold), but *ste6*⁺ remains unchanged, in *hid3Δ*. Both Erg25 and Ste6 are GEF proteins that activate Ras1, but they activate the spindle formation/cytokinesis and pheromone signalling pathways, respectively. If it follows that increased expression of *erg25*⁺ leads to more protein, then Ras activity would favour induction of the pathway giving normal morphology. In *S. cerevisiae*, Ras has other potential actions, such as activating influencing lifespan, the cell cycle and stress response pathways. The regulation of signalling pathways by Ras is dependent on its subcellular location (Onken et al., 2006). Ras1 located at the plasma membrane activates the Byr1 pathway and Ras1 located in the endomembrane system preferentially activates the Scd1/Ral1 pathway. It is possible that in *hid3Δ* transport properties are sufficiently disrupted to retain Ras within the endomembrane system. It is critical to determine the subcellular location of Ras in the *hidΔ* mutants, particularly in *hid3Δ*. Furthermore, mis-localisation of Ras and induction of its protein levels due to disruption of Golgi structure would coincide with the idea that the localisation of proteins other than Ubp5 could be affected in *hid3Δ*.

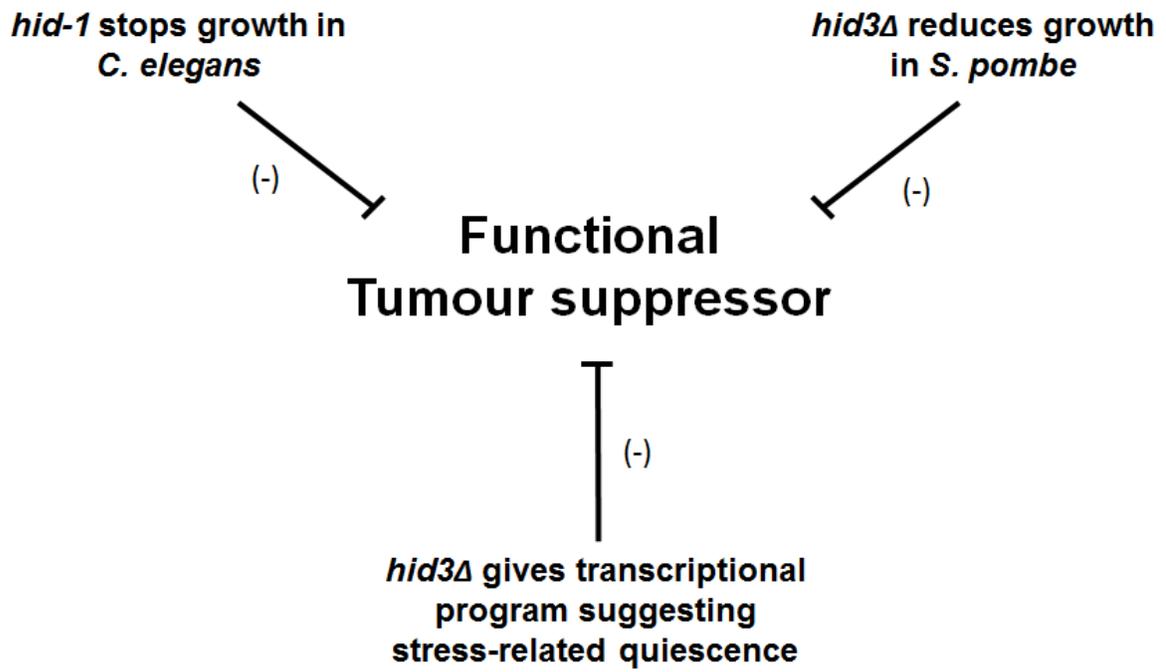


Figure 6.3. Analysis of tumour suppressor properties of HID1.

In conclusion, under the model by which I began the project, HID1 may be functioning as a tumour suppressor whose decrease in expression and lower level of protein contributes to increased growth and proliferation. This would imply the reverse in that increased expression of *HID1* would aid, or at least not prevent, apoptosis. My observation from a survey of publicly available data on gene expression in cancers was that HID1 appears to be UR in more cancers than DR. This suggests that HID1 contributes to maintaining cell growth and it is consistent with the function determined for HID-1 in *C. elegans* and its requirement for normal growth in *S. pombe*. Based on the results here I can make a mechanistic argument that down-regulation of HID1 in humans would be beneficial by contributing to a quiescent cellular state that would impede cell proliferation (Figure 6.3). In fact, cellular quiescence has been observed as a protective mechanism from tumorigenesis and proliferation (White et al., 2014). However, it is not yet clear what effect complete removal of HID1 from human cells would have and certainly in combination with other mutations affecting Golgi function, protein transport or cell/stress signaling.

Outlook and Future Work

It is apparent that there are many experiments that could be conducted in order to provide more detailed information on the ideas presented here. This is the nature of omics work to raise questions and hypotheses for direct testing. I will only some of them here. Some of these I am raising to point out potential limitations in the work also.

I feel that there is a need to create mutants without the marker genes present. Although the effects of the expression of the Nat^R gene did not show up in the PCA of the transcript data, it did show up in other data (Alsamari 2015). It was greatly unexpected to find glucose levels in the wild-type much greater than in any of the strains expressing the selectable marker Nat^R . This difference would have substantial apparent effects on signalling processes, particularly involving metabolic signalling. For example, the gene *git3⁺* that encodes the G-protein receptor kinase involved in high glucose signalling is significantly DR in *hid3Δ* compared with VCN. Interestingly, *git3⁺* expression is significantly induced in all marker-containing genotypes compared to wild-type, which is consistent with the levels of glucose being lower in these genotypes. The observation that *git3⁺* expression is reduced in *hid3Δ* compared to VCN is consistent with the differences in the amount of glucose measured between these genotypes (Alasmari 2015). Therefore, it appears that there is some glucose regulation of gene expression occurring in the mutant genotypes, and it is not possible to exclude some contribution from the cAMP signalling pathway. It is possible to create mutant strains without the marker gene using the CreLox system or CRISPR/Cas9 system of genome editing (Sauer, 1998). This would allow direct comparison of mutant and wild-type strains.

The main continuation of the project would be to study the relationship between Golgi function and the various signalling pathways, gene expression and the potential entry into quiescence. For example, the relationship with the MAPK signalling pathway can be investigated by determining the phosphorylation state of the various components in the *hid3* mutants. Since Wis1 is induced on both the mRNA and protein levels and *atf1*⁺ is also induced, it would be particularly interesting to see if phosphorylation of Sty1 is increased in the *hid3Δ* mutants, and if this is altered in the *hid3Δ wis1Δ* double mutant. It would also be interesting to determine if the expression of *atf1*⁺ remains induced in the *hid3Δ sty1Δ* double mutant indicating induction by a Sty1-independent mechanism. Also, it is essential to determine the activity state of the cAMP-dependent kinase signalling pathway.

It will be interesting to determine if the various induced TFs in *hid3Δ* contribute to the induction of expression of plasma membrane and secreted proteins, which can also be done by using combinations *hid3Δ::tf* double mutants and mutants of the Mediator co-activator complex. It will be possible to test if the reduced growth of *hid3Δ* is indicative of cells entering quiescence through induction of Wis1 or Cdc2 by analysing the transition into quiescence of the *hid3Δ wis1Δ* and *hid3Δ cdc2Δ* double mutants. Non-complementation of the slow growth phenotype in the double mutants would suggest that the slow growth is due to something else other than a nutrient starvation-induced transition into quiescence, such as a direct effect on Golgi function.

Due to the rich results obtained from doing the transcriptional profiling of the *hid3Δ* mutants and the questions it has raised, there remains much to be investigated. In addition, I have not discussed the consequences of the non-coding RNA work outside of the discussion of that Chapter. Since it was a side project and has produced preliminary, but interesting results, I feel it is outside the scope of this discussion to comment on it. However, non-coding RNAs present an interesting avenue toward understanding the regulation of gene expression in *S. pombe*.

REFERENCES

- Aalto, A.P., and Pasquinelli, A.E. (2012). Small non-coding RNAs mount a silent revolution in gene expression. *Curr. Opin. Cell Biol.* *24*, 333–340.
- Agarwal, R., Jurisica, I., Mills, G.B., and Cheng, K.W. (2009). The emerging role of the RAB25 small GTPase in cancer. *Traffic Cph. Den.* *10*, 1561–1568.
- Ailion, M., and Thomas, J.H. (2003). Isolation and characterization of high-temperature-induced Dauer formation mutants in *Caenorhabditis elegans*. *Genetics* *165*, 127–144.
- Akiyama, T., Iodi Carstens, M., and Carstens, E. (2011a). Transmitters and Pathways Mediating Inhibition of Spinal Itch-Signaling Neurons by Scratching and Other Counterstimuli. *PLoS ONE* *6*.
- Akiyama, T., Carstens, M.I., and Carstens, E. (2011b). Enhanced responses of lumbar superficial dorsal horn neurons to intradermal PAR-2 agonist but not histamine in a mouse hindpaw dry skin itch model. *J. Neurophysiol.* *105*, 2811–2817.
- Alasmari, A. N (2015). Proteomic studies of the *hid1Δ* and *hid3Δ* mutants of *Schizosaccharomyces pombe*. PhD thesis. University of Bordeaux.
- Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., and Walter, P. (2002a). The Structure and Function of DNA.
- Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., and Walter, P. (2002b). From DNA to RNA.
- Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. *J. Mol. Biol.* *215*, 403–410.
- Amaral, P.P., and Mattick, J.S. (2008). Noncoding RNA in development. *Mamm. Genome* *19*, 454–492.
- Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. *Genome Biol.* *11*, R106.
- Arvidsson, S., Kwasniewski, M., Riaño-Pachón, D.M., and Mueller-Roeber, B. (2008). QuantPrime--a flexible tool for reliable high-throughput primer design for quantitative PCR. *BMC Bioinformatics* *9*, 465.
- Askwith, C., and Kaplan, J. (1997). An Oxidase-Permease-based Iron Transport System in *Schizosaccharomyces pombe* and Its Expression in *Saccharomyces cerevisiae*. *J. Biol. Chem.* *272*, 401–405.
- Aspegren, A., Hinas, A., Larsson, P., Larsson, A., and Söderbom, F. (2004). Novel non-coding RNAs in *Dictyostelium discoideum* and their expression during development. *Nucleic Acids Res.* *32*, 4646–4656.
- Aziz, R.K., Breitbart, M., and Edwards, R.A. (2010). Transposases are the most abundant, most ubiquitous genes in nature. *Nucleic Acids Res.* *38*, 4207–4217.
- Bachellerie, J.P., Cavallé, J., and Hüttenhofer, A. (2002). The expanding snoRNA world. *Biochimie* *84*, 775–790.

- Bähler, J. (2005). Cell-cycle control of gene expression in budding and fission yeast. *Annu. Rev. Genet.* *39*, 69–94.
- Bähler, J., Wu, J.Q., Longtine, M.S., Shah, N.G., McKenzie, A., Steever, A.B., Wach, A., Philippsen, P., and Pringle, J.R. (1998). Heterologous modules for efficient and versatile PCR-based gene targeting in *Schizosaccharomyces pombe*. *Yeast* Chichester Engl. *14*, 943–951.
- Bailis, J.M., and Forsburg, S.L. (2007). From DNA Replication to Genome Instability in *Schizosaccharomyces Pombe*: Pathways to Cancer. In *Yeast as a Tool in Cancer Research*, J.L. Nitiss, and J. Heitman, eds. (Springer Netherlands), pp. 1–35.
- Bannister, A.J., and Kouzarides, T. (2011). Regulation of chromatin by histone modifications. *Cell Res.* *21*, 381–395.
- Barbarotto, E., Schmittgen, T.D., and Calin, G.A. (2008). MicroRNAs and cancer: profile, profile, profile. *Int. J. Cancer J. Int. Cancer* *122*, 969–977.
- Barciszewska, M.Z., Szymański, M., Erdmann, V.A., and Barciszewski, J. (2001). Structure and functions of 5S rRNA. *Acta Biochim. Pol.* *48*, 191–198.
- Barciszewski, J. (2009). Noncoding RNAs in Biology and Disease. eLS 1–8.
- Barrandon, C., Spiluttini, B., and Bensaude, O. (2008). Non-coding RNAs regulating the transcriptional machinery. *Biol. Cell Auspices Eur. Cell Biol. Organ.* *100*, 83–95.
- Barry, W.T., Nobel, A.B., and Wright, F.A. (2005). Significance analysis of functional categories in gene expression studies: a structured permutation approach. *Bioinforma. Oxf. Engl.* *21*, 1943–1949.
- Bartel, D.P. (2009). MicroRNAs: target recognition and regulatory functions. *Cell* *136*, 215–233.
- Baudin, A., Ozier-Kalogeropoulos, O., Denouel, A., Lacroute, F., and Cullin, C. (1993). A simple and efficient method for direct gene deletion in *Saccharomyces cerevisiae*. *Nucleic Acids Res.* *21*, 3329–3330.
- Bejerano, G., Pheasant, M., Makunin, I., Stephen, S., Kent, W.J., Mattick, J.S., and Haussler, D. (2004). Ultraconserved elements in the human genome. *Science* *304*, 1321–1325.
- Belancio, V.P., Deininger, P.L., and Roy-Engel, A.M. (2009). LINE dancing in the human genome: transposable elements and disease. *Genome Med.* *1*, 97.
- Bell, A.W., Ward, M.A., Blackstock, W.P., Freeman, H.N., Choudhary, J.S., Lewis, A.P., Chotai, D., Fazel, A., Gushue, J.N., Paiement, J., et al. (2001). Proteomics characterization of abundant Golgi membrane proteins. *J. Biol. Chem.* *276*, 5152–5165.
- Berezikov, E., Guryev, V., van de Belt, J., Wienholds, E., Plasterk, R.H.A., and Cuppen, E. (2005). Phylogenetic shadowing and computational identification of human microRNA genes. *Cell* *120*, 21–24.
- Berger, S.L. (2002). Histone modifications in transcriptional regulation. *Curr. Opin. Genet. Dev.* *12*, 142–148.

- Berger, M.F., Levin, J.Z., Vijayendran, K., Sivachenko, A., Adiconis, X., Maguire, J., Johnson, L.A., Robinson, J., Verhaak, R.G., Sougnez, C., et al. (2010). Integrative analysis of the melanoma transcriptome. *Genome Res.* *20*, 413–427.
- Bernhart, S.H., Hofacker, I.L., Will, S., Gruber, A.R., and Stadler, P.F. (2008). RNAalifold: improved consensus structure prediction for RNA alignments. *BMC Bioinformatics* *9*, 474.
- Berretta, R., and Moscato, P. (2010). Cancer biomarker discovery: the entropic hallmark. *PLoS One* *5*, e12262.
- Bexiga, M.G., and Simpson, J.C. (2013). Human diseases associated with form and function of the Golgi complex. *Int. J. Mol. Sci.* *14*, 18670–18681.
- Bien, C.M., and Espenshade, P.J. (2010). Sterol Regulatory Element Binding Proteins in Fungi: Hypoxic Transcription Factors Linked to Pathogenesis. *Eukaryot. Cell* *9*, 352–359.
- Blagoveshchenskaya, A., Cheong, F.Y., Rohde, H.M., Glover, G., Knödler, A., Nicolson, T., Boehmelt, G., and Mayinger, P. (2008). Integration of Golgi trafficking and growth factor signaling by the lipid phosphatase SAC1. *J. Cell Biol.* *180*, 803–812.
- Bowen, N.J., Jordan, I.K., Epstein, J.A., Wood, V., and Levin, H.L. (2003). Retrotransposons and their recognition of pol II promoters: a comprehensive survey of the transposable elements from the complete genome sequence of *Schizosaccharomyces pombe*. *Genome Res.* *13*, 1984–1997.
- Brenner, S., Jacob, F., and Meselson, M. (1961). An Unstable Intermediate Carrying Information from Genes to Ribosomes for Protein Synthesis. *Publ. Online* 13 May 1961 Doi101038190576a0 *190*, 576–581.
- Brosnan, C.A., and Voinnet, O. (2009). The long and the short of noncoding RNAs. *Curr. Opin. Cell Biol.* *21*, 416–425.
- Brown, M.S., and Goldstein, J.L. (1997). The SREBP Pathway: Regulation of Cholesterol Metabolism by Proteolysis of a Membrane-Bound Transcription Factor. *Cell* *89*, 331–340.
- Brown, C.J., Lain, S., Verma, C.S., Fersht, A.R., and Lane, D.P. (2009). Awakening guardian angels: drugging the p53 pathway. *Nat. Rev. Cancer* *9*, 862–873.
- Buschman, M.D., Rahajeng, J., and Field, S.J. (2015). GOLPH3 Links the Golgi, DNA Damage, and Cancer. *Cancer Res.* *75*, 624–627.
- Calin, G.A., Sevignani, C., Dumitru, C.D., Hyslop, T., Noch, E., Yendamuri, S., Shimizu, M., Rattan, S., Bullrich, F., Negrini, M., et al. (2004). Human microRNA genes are frequently located at fragile sites and genomic regions involved in cancers. *Proc. Natl. Acad. Sci. U. S. A.* *101*, 2999–3004.
- Calin, G.A., Liu, C. gong, Ferracin, M., Hyslop, T., Spizzo, R., Sevignani, C., Fabbri, M., Cimmino, A., Lee, E.J., Wojcik, S.E., et al. (2007). Ultraconserved Regions Encoding ncRNAs Are Altered in Human Leukemias and Carcinomas. *Cancer Cell* *12*, 215–229.
- Caspari, T., and Hilditch, V. (2015). Two Distinct Cdc2 Pools Regulate Cell Cycle Progression and the DNA Damage Response in the Fission Yeast *S.pombe*. *PLoS ONE* *10*.

- Ceballos, M., and Vioque, A. (2007). tRNase Z. *Protein Pept. Lett.* *14*, 137–145.
- Chan, P.P., and Lowe, T.M. (2009). GtRNadb: a database of transfer RNA genes detected in genomic sequence. *Nucleic Acids Res.* *37*, D93–D97.
- Chen, D., Toone, W.M., Mata, J., Lyne, R., Burns, G., Kivinen, K., Brazma, A., Jones, N., and Bähler, J. (2003). Global transcriptional responses of fission yeast to environmental stress. *Mol. Biol. Cell* *14*, 214–229.
- Chen, M., Xiao, J., Zhang, Z., Liu, J., Wu, J., and Yu, J. (2013). Identification of Human HK Genes and Gene Expression Regulation Study in Cancer from Transcriptomics Data Analysis. *PLoS ONE* *8*.
- Ciccio, A., and Elledge, S.J. (2010). The DNA damage response: making it safe to play with knives. *Mol. Cell* *40*, 179–204.
- Cloonan, N., Forrest, A.R.R., Kolle, G., Gardiner, B.B.A., Faulkner, G.J., Brown, M.K., Taylor, D.F., Steptoe, A.L., Wani, S., Bethel, G., et al. (2008). Stem cell transcriptome profiling via massive-scale mRNA sequencing. *Nat. Methods* *5*, 613–619.
- Coller, H.A., Sang, L., and Roberts, J.M. (2006). A new description of cellular quiescence. *PLoS Biol.* *4*, e83.
- Conklin, J.F., Goldman, A., and Lopez, A.J. (2005). Stabilization and analysis of intron lariats in vivo. *Methods San Diego Calif* *37*, 368–375.
- Connerly, P.L., Esaki, M., Montegna, E.A., Strongin, D.E., Levi, S., Soderholm, J., and Glick, B.S. (2005). Sec16 is a determinant of transitional ER organization. *Curr. Biol. CB* *15*, 1439–1447.
- Conrad, M., Schothorst, J., Kankipati, H.N., Van Zeebroeck, G., Rubio-Teixeira, M., and Thevelein, J.M. (2014). Nutrient sensing and signaling in the yeast *Saccharomyces cerevisiae*. *FEMS Microbiol. Rev.* *38*, 254–299.
- Darty, K., Denise, A., and Ponty, Y. (2009). VARNA: Interactive drawing and editing of the RNA secondary structure. *Bioinformatics* *25*, 1974–1975.
- Daulny, A., and Tansey, W.P. (2009). Damage control: DNA repair, transcription, and the ubiquitin-proteasome system. *DNA Repair* *8*, 444–448.
- Davidson, M.K., Shandilya, H.K., Hirota, K., Ohta, K., and Wahls, W.P. (2004). Atf1-Pcr1-M26 complex links stress-activated MAPK and cAMP-dependent protein kinase pathways via chromatin remodeling of *cgs2+*. *J. Biol. Chem.* *279*, 50857–50863.
- Denais, C., Dent, C.L., Southgate, L., Hoyle, J., Dafou, D., Trembath, R.C., and Machado, R.D. (2011). Dymeclin, the gene underlying Dyggve-Melchior-Clausen syndrome, encodes a protein integral to extracellular matrix and golgi organization and is associated with protein secretion pathways critical in bone development. *Hum. Mutat.* *32*, 231–239.
- Dimitrov, A., Paupe, V., Gueudry, C., Sibarita, J.-B., Raposo, G., Vielemeyer, O., Gilbert, T., Csaba, Z., Attie-Bitach, T., Cormier-Daire, V., et al. (2009). The gene responsible for Dyggve-Melchior-Clausen syndrome encodes a novel peripheral membrane protein dynamically associated with the Golgi apparatus. *Hum. Mol. Genet.* *18*, 440–453.

- Ding, F., Li, H.H., Zhang, S., Solomon, N.M., Camper, S.A., Cohen, P., and Francke, U. (2008). SnoRNA Snord116 (Pwcr1/MBII-85) deletion causes growth deficiency and hyperphagia in mice. *PLoS One* 3, e1709.
- Dinger, M.E., Amaral, P.P., Mercer, T.R., Pang, K.C., Bruce, S.J., Gardiner, B.B., Askarian-Amiri, M.E., Ru, K., Soldà, G., Simons, C., et al. (2008). Long noncoding RNAs in mouse embryonic stem cell pluripotency and differentiation. *Genome Res.* 18, 1433–1445.
- Dippold, H.C., Ng, M.M., Farber-Katz, S.E., Lee, S.-K., Kerr, M.L., Peterman, M.C., Sim, R., Wiharto, P.A., Galbraith, K.A., Madhavarapu, S., et al. (2009). GOLPH3 bridges phosphatidylinositol-4-phosphate and actomyosin to stretch and shape the Golgi to promote budding. *Cell* 139, 337–351.
- Djupedal, I., Kos-Braun, I.C., Mosher, R.A., Söderholm, N., Simmer, F., Hardcastle, T.J., Fender, A., Heidrich, N., Kagansky, A., Bayne, E., et al. (2009). Analysis of small RNA in fission yeast; centromeric siRNAs are potentially generated through a structured RNA. *EMBO J.* 28, 3832–3844.
- Eddy, S.R. (2001). Non-coding RNA genes and the modern RNA world. *Nat. Rev. Genet.* 2, 919–929.
- Efroni, S., Duttagupta, R., Cheng, J., Deghani, H., Hoepfner, D.J., Dash, C., Bazett-Jones, D.P., Le Grice, S., McKay, R.D.G., Buetow, K.H., et al. (2008). Global transcription in pluripotent embryonic stem cells. *Cell Stem Cell* 2, 437–447.
- Espenshade, P.J. (2006). SREBPs: sterol-regulated transcription factors. *J. Cell Sci.* 119, 973–976.
- Espenshade, P.J., and Hughes, A.L. (2007). Regulation of sterol synthesis in eukaryotes. *Annu. Rev. Genet.* 41, 401–427.
- Farazi, T.A., Juranek, S.A., and Tuschl, T. (2008). The growing catalog of small RNAs and their association with distinct Argonaute/Piwi family members. *Dev. Camb. Engl.* 135, 1201–1214.
- Farber-Katz, S.E., Dippold, H.C., Buschman, M.D., Peterman, M.C., Xing, M., Noakes, C.J., Tat, J., Ng, M.M., Rahajeng, J., Cowan, D.M., et al. (2014). DNA damage triggers Golgi dispersal via DNA-PK and GOLPH3. *Cell* 156, 413–427.
- Fayet-Lebaron, E., Atzorn, V., Henry, Y., and Kiss, T. (2009). 18S rRNA processing requires base pairings of snR30 H/ACA snoRNA to eukaryote-specific 18S sequences. *Embo J.* 28, 1260–1270.
- Felsenfeld, G., Davies, D.R., and Rich, A. (1957). FORMATION OF A THREE-STRANDED POLYNUCLEOTIDE MOLECULE. *J Am Chem Soc* 79, 2023–2024.
- Ferlay, J., Soerjomataram I, I., Dikshit, R., Eser, S., Mathers, C., Rebelo, M., Parkin, D.M., Forman D, D., and Bray, F. (2014). Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *Int. J. Cancer J. Int. Cancer* 136, E359–E386.
- Feschotte, C. (2008). Transposable elements and the evolution of regulatory networks. *Nat. Rev. Genet.* 9, 397–405.
- Fink, G.R. (1987). Pseudogenes in yeast? *Cell* 49, 5–6.
- Fire, A., Xu, S., Montgomery, M.K., Kostas, S.A., Driver, S.E., and Mello, C.C. (1998). Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* 391, 806–811.

- Forsburg, S.L. (2005). The yeasts *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*: models for cell biology research. *Gravitational Space Biol. Bull. Publ. Am. Soc. Gravitational Space Biol.* 18, 3–9.
- Franco-Zorrilla, J.M., López-Vidriero, I., Carrasco, J.L., Godoy, M., Vera, P., and Solano, R. (2014). DNA-binding specificities of plant transcription factors and their potential to define target genes. *Proc. Natl. Acad. Sci. U. S. A.* 111, 2367–2372.
- Frank, D.N., and Pace, N.R. (1998). RIBONUCLEASE P: Unity and Diversity in a tRNA Processing Ribozyme. *Annu. Rev. Biochem.* 67, 153–180.
- Gallagher, E.J., and LeRoith, D. (2011). Minireview: IGF, Insulin, and Cancer. *Endocrinology* 152, 2546–2551.
- Garber, M., Grabherr, M.G., Guttman, M., and Trapnell, C. (2011). Computational methods for transcriptome annotation and quantification using RNA-seq. *Nat. Methods* 8, 469–477.
- Gasch, A.P. (2007). Comparative genomics of the environmental stress response in ascomycete fungi. *Yeast Chichester Engl.* 24, 961–976.
- Ghouzzi, V. El, Dagoneau, N., Esther, K., Christel Thauvin-Robinet, W.C., Al-Gazali, C.P.-S.L.I., Verloes, A., Merrer, M. Le, Munnich, A., Trembath, R.C., and Cormier-Daire, V. (2003). Mutations in a novel gene *Dymeclin* (FLJ20071) are responsible for Dyggve-Melchior-Clausen syndrome. *Hum. Mol. Genet.* 12, 357–364.
- Gong, X., Li, J., Shao, W., Wu, J., Qian, H., Ren, R., Espenshade, P., and Yan, N. (2015). Structure of the WD40 domain of SCAP from fission yeast reveals the molecular basis for SREBP recognition. *Cell Res.* 25, 401–411.
- Gorodkin, J., Hofacker, I.L., Torarinsson, E., Yao, Z., Havgaard, J.H., and Ruzzo, W.L. (2010). De novo prediction of structured RNAs from genomic sequences. *Trends Biotechnol.* 28, 9–19.
- Gottesman, S. (2002). Stealth regulation: Biological circuits with small RNA switches. *Genes Dev.* 16, 2829–2842.
- Gottlieb, S., and Ruvkun, G. (1994). *daf-2*, *daf-16* and *daf-23*: genetically interacting genes controlling Dauer formation in *Caenorhabditis elegans*. *Genetics* 137, 107–120.
- Grallert, B., Nurse, P., and Patterson, T.E. (1993). A study of integrative transformation in *Schizosaccharomyces pombe*. *Mol. Gen. Genet. MGG* 238, 26–32.
- Graveley, B.R., Brooks, A.N., Carlson, J.W., Duff, M.O., Landolin, J.M., Yang, L., Artieri, C.G., van Baren, M.J., Boley, N., Booth, B.W., et al. (2011). The developmental transcriptome of *Drosophila melanogaster*. *Nature* 471, 473–479.
- Gray, J. V, Petsko, G.A., Johnston, G.C., Ringe, D., Singer, R.A., and Werner-Washburne, M. (2004). “Sleeping beauty”: quiescence in *Saccharomyces cerevisiae*. *Microbiol. Mol. Biol. Rev. MMBR* 68, 187–206.
- Griffiths-Jones, S. (2005). RALEE--RNA ALignment editor in Emacs. *Bioinforma. Oxf. Engl.* 21, 257–259.

- Griffiths-Jones, S. (2007). Annotating noncoding RNA genes. *Annu. Rev. Genomics Hum. Genet.* 8, 279–298.
- Grimm, C., and Kohli, J. (1988). Observations on integrative transformation in *Schizosaccharomyces pombe*. *Mol. Gen. Genet. MGG* 215, 87–93.
- Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N., and Altman, S. (1983). The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell* 35, 849–857.
- Gundem, G., Perez-Llamas, C., Jene-Sanz, A., Kedziarska, A., Islam, A., Deu-Pons, J., Furney, S.J., and Lopez-Bigas, N. (2010). IntOGen: integration and data mining of multidimensional oncogenomic data. *Nat. Methods* 7, 92–93.
- Gutell, R.R., Cannone, J.J., Shang, Z., Du, Y., and Serra, M.J. (2000). A story: unpaired adenosine bases in ribosomal RNAs. *J. Mol. Biol.* 304, 335–354.
- Guthrie, C., and Fink, G.R. (2004). *Guide to Yeast Genetics and Molecular and Cell Biology* (Gulf Professional Publishing).
- Guttman, M., and Rinn, J.L. (2012). Modular regulatory principles of large non-coding RNAs. *Nature* 482, 339–346.
- Ha, H., Song, J., Wang, S., Kapusta, A., Feschotte, C., Chen, K.C., and Xing, J. (2014). A comprehensive analysis of piRNAs from adult human testis and their relationship with genes and mobile elements. *BMC Genomics* 15, 545.
- Hamilton, A.J., and Baulcombe, D.C. (1999). A species of small antisense RNA in posttranscriptional gene silencing in plants. *Science* 286, 950–952.
- Hanahan, D., and Weinberg, R.A. (2000). The hallmarks of cancer. *Cell* 100, 57–70.
- Harada, H., Nagai, H., and Tsuneizumi, M. (2001). Identification of DMC1 , a novel gene in the TOC region on 17q25 . 1 that shows loss of expression in multiple human cancers. *J. Hum. Genet.* 46, 90–95.
- Harris, M.A., Clark, J., Ireland, A., Lomax, J., Ashburner, M., Foulger, R., Eilbeck, K., Lewis, S., Marshall, B., Mungall, C., et al. (2004). The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res.* 32, D258–D261.
- Hart, A. (2006). Behavior. *WormBook*.
- Hayles, J., Wood, V., Jeffery, L., Hoe, K.-L., Kim, D.-U., Park, H.-O., Salas-Pino, S., Heichinger, C., and Nurse, P. (2013). A genome-wide resource of cell cycle and cell shape genes of fission yeast. *Open Biol.* 3, 130053.
- He, L., and Hannon, G.J. (2004). MicroRNAs small RNAs with a big role in gene regulation. *Nat. Rev. Genet.* 5, 1–11.
- Helm, M. (2006). Post-transcriptional nucleotide modification and alternative folding of RNA. *Nucleic Acids Res.* 34, 721–733.

- Hentges, P., Van Driessche, B., Tafforeau, L., Vandenhaute, J., and Carr, A.M. (2005). Three novel antibiotic marker cassettes for gene disruption and marker switching in *Schizosaccharomyces pombe*. *Yeast* Chichester Engl. 22, 1013–1019.
- Herbig, A., and Nieselt, K. (2011). nocoRNAc: Characterization of non-coding RNAs in prokaryotes. *BMC Bioinformatics* 12, 40.
- Hicks, S.W., and Machamer, C.E. (2005). Golgi structure in stress sensing and apoptosis. *Biochim. Biophys. Acta* 1744, 406–414.
- Hirsch, H.A., Iliopoulos, D., Joshi, A., Zhang, Y., Jaeger, S.A., Bulyk, M., Tschlis, P.N., Shirley Liu, X., and Struhl, K. (2010). A Transcriptional Signature and Common Gene Networks Link Cancer with Lipid Metabolism and Diverse Human Diseases. *Cancer Cell* 17, 348–361.
- Ho, J.R., Chapeaublanc, E., Kirkwood, L., Nicolle, R., Benhamou, S., Lebret, T., Allory, Y., Southgate, J., Radvanyi, F., and Goud, B. (2012). Deregulation of Rab and Rab effector genes in bladder cancer. *PLoS One* 7, e39469.
- Hochstrasser, M. (2004). Ubiquitin signalling: what's in a chain? *Nat. Cell Biol.* 6, 571–572.
- Hoffman, C.S. (2005). Glucose sensing via the protein kinase A pathway in *Schizosaccharomyces pombe*. *Biochem. Soc. Trans.* 33, 257–260.
- Hooks, K.B., and Griffiths-Jones, S. (2011). Conserved RNA structures in the non-canonical Hac1/Xbp1 intron. *RNA Biol.* 8, 552–556.
- van der Horst, A., de Vries-Smits, A.M.M., Brenkman, A.B., van Triest, M.H., van den Broek, N., Colland, F., Maurice, M.M., and Burgering, B.M.T. (2006). FOXO4 transcriptional activity is regulated by monoubiquitination and USP7/HAUSP. *Nat. Cell Biol.* 8, 1064–1073.
- Horton, J.D., Goldstein, J.L., and Brown, M.S. (2002). SREBPs: activators of the complete program of cholesterol and fatty acid synthesis in the liver. *J. Clin. Invest.* 109, 1125–1131.
- Hua, X., Nohturfft, A., Goldstein, J.L., and Brown, M.S. (1996). Sterol resistance in CHO cells traced to point mutation in SREBP cleavage-activating protein. *Cell* 87, 415–426.
- Hua, X., Yu, L., Pan, W., Huang, X., Liao, Z., Xian, Q., Fang, L., and Shen, H. (2012). Increased expression of Golgi phosphoprotein-3 is associated with tumor aggressiveness and poor prognosis of prostate cancer. *Diagn. Pathol.* 7, 127.
- Huang, Z., Huang, D., Ni, S., Peng, Z., Sheng, W., and Du, X. (2010). Plasma microRNAs are promising novel biomarkers for early detection of colorectal cancer. *Int. J. Cancer J. Int. Cancer* 127, 118–126.
- Hughes, B.T., and Espenshade, P.J. (2008). Oxygen-regulated degradation of fission yeast SREBP by Ofd1, a prolyl hydroxylase family member. *EMBO J.* 27, 1491–1501.
- Hughes, A.L., Todd, B.L., and Espenshade, P.J. (2005). SREBP pathway responds to sterols and functions as an oxygen sensor in fission yeast. *Cell* 120, 831–842.
- Huntzinger, E., and Izaurralde, E. (2011). Gene silencing by microRNAs: contributions of translational repression and mRNA decay. *Nat. Rev. Genet.* 12, 99–110.

- Hüttenhofer, A., and Vogel, J. (2006). Experimental approaches to identify non-coding RNAs. *Nucleic Acids Res.* *34*, 635–646.
- Ikeda, F., and Dikic, I. (2008). Atypical ubiquitin chains: new molecular signals. “Protein Modifications: Beyond the Usual Suspects” review series. *EMBO Rep.* *9*, 536–542.
- Ilyushik, E., Pryce, D.W., Walerych, D., Riddell, T., Wakeman, J.A., McInerny, C.J., and McFarlane, R.J. (2005). Psc3 cohesin of *Schizosaccharomyces pombe*: cell cycle analysis and identification of three distinct isoforms. *Biol. Chem.* *386*, 613–621.
- Ishiguro, A., Kassavetis, G.A., and Geiduschek, E.P. (2002). Essential Roles of Bdp1, a Subunit of RNA Polymerase III Initiation Factor TFIIB, in Transcription and tRNA Processing. *Mol. Cell. Biol.* *22*, 3264–3275.
- Jiang, Y., Davis, C., and Broach, J.R. (1998). Efficient transition to growth on fermentable carbon sources in *Saccharomyces cerevisiae* requires signaling through the Ras pathway. *EMBO J.* *17*, 6942–6951.
- Johnson, S.M., Grosshans, H., Shingara, J., Byrom, M., Jarvis, R., Cheng, A., Labourier, E., Reinert, K.L., Brown, D., and Slack, F.J. (2005). RAS is regulated by the let-7 microRNA family. *Cell* *120*, 635–647.
- Jones, P.A., and Laird, P.W. (1999). Cancer epigenetics comes of age. *Nat. Genet.* *21*, 163–167.
- Kalikin, L.M., Frank, T.S., Svoboda-Newman, S.M., Wetzel, J.C., Cooney, K.A., and Petty, E.M. (1997). A region of interstitial 17q25 allelic loss in ovarian tumors coincides with a defined region of loss in breast tumors. *Oncogene* *14*, 1991–1994.
- Kasschau, K.D., Fahlgren, N., Chapman, E.J., Sullivan, C.M., Cumbie, J.S., Givan, S.A., and Carrington, J.C. (2007). Genome-wide profiling and analysis of *Arabidopsis* siRNAs. *PLoS Biol.* *5*, e57.
- Katoh, K., and Standley, D.M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* *30*, 772–780.
- Käufer, N.F., Simanis, V., and Nurse, P. (1985). Fission yeast *Schizosaccharomyces pombe* correctly excises a mammalian RNA transcript intervening sequence. *Nature* *318*, 78–80.
- Kavanaugh, L.A., and Dietrich, F.S. (2009). Non-coding RNA prediction and verification in *Saccharomyces cerevisiae*. *PLoS Genet.* *5*, e1000321.
- King, M.-C. (2014). The race” to clone BRCA1. *Science* *343*, 1462–1465.
- Kishore, S., and Stamm, S. (2006). The snoRNA HBII-52 regulates alternative splicing of the serotonin receptor 2C. *Science* *311*, 230–232.
- Kohtz, J.D., and Fishell, G. (2004). Developmental regulation of EVF-1, a novel non-coding RNA transcribed upstream of the mouse *Dlx6* gene. *Gene Expr. Patterns GEP* *4*, 407–412.
- Komander, D. (2010). Mechanism, specificity and structure of the deubiquitinases. *Subcell. Biochem.* *54*, 69–87.
- Kornitzer, D., and Ciechanover, A. (2000). Modes of regulation of ubiquitin-mediated protein degradation. *J. Cell. Physiol.* *182*, 1–11.

- Kouranti, I., McLean, J., and Feoktistova, A. (2010). A global census of fission yeast deubiquitinating enzyme localization and interaction networks reveals distinct compartmentalization profiles and overlapping. *PLoS Biol.* 8.
- Kruger, K., Grabowski, P.J., Zaug, A.J., Sands, J., Gottschling, D.E., and Cech, T.R. (1982). Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of *Tetrahymena*. *Cell* 31, 147–157.
- Kumar, M.S., Lu, J., Mercer, K.L., Golub, T.R., and Jacks, T. (2007). Impaired microRNA processing enhances cellular transformation and tumorigenesis. *Nat. Genet.* 39, 673–677.
- KV Venkatesh, S.B. (2013). Quantitative Analysis of a Dynamic Cell Cycle Regulatory Model of *Schizosaccharomyces pombe*. *Curr. Synth. Syst. Biol.* 01.
- Laporte, D., Lebaudy, A., Sahin, A., Pinson, B., Ceschin, J., Daignan-Fornier, B., and Sagot, I. (2011). Metabolic status rather than cell cycle signals control quiescence entry and exit. *J. Cell Biol.* 192, 949–957.
- Lawrie, C.H., Gal, S., Dunlop, H.M., Pushkaran, B., Liggins, A.P., Pulford, K., Banham, A.H., Pezzella, F., Boulwood, J., Wainscoat, J.S., et al. (2008). Detection of elevated levels of tumour-associated microRNAs in serum of patients with diffuse large B-cell lymphoma. *Br. J. Haematol.* 141, 672–675.
- Lê, S., Josse, J., Husson, F., and others (2008). FactoMineR: an R package for multivariate analysis. *J. Stat. Softw.* 25, 1–18.
- Lee, E.Y.H.P., and Muller, W.J. (2010). Oncogenes and tumor suppressor genes. *Cold Spring Harb. Perspect. Biol.* 2, a003236.
- Lee, S.K., and Calin, G.A. (2011). Non-coding RNAs and cancer: new paradigms in oncology. *Discov. Med.* 11, 245–254.
- Lee, C.-Y.S., Stewart, E. V, Hughes, B.T., and Espenshade, P.J. (2009). Oxygen-dependent binding of Nro1 to the prolyl hydroxylase Ofd1 regulates SREBP degradation in yeast. *EMBO J.* 28, 135–143.
- Lee, R.C., Feinbaum, R.L., and Ambros, V. (1993). The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. *Cell* 75, 843–854.
- Lenssen, E., Oberholzer, U., Labarre, J., De Virgilio, C., and Collart, M.A. (2002). *Saccharomyces cerevisiae* Ccr4-Not complex contributes to the control of Msn2p-dependent transcription by the Ras/cAMP pathway. *Mol. Microbiol.* 43, 1023–1037.
- Lestrade, L., and Weber, M.J. (2006). snoRNA-LBME-db, a comprehensive database of human H/ACA and C/D box snoRNAs. *Nucleic Acids Res.* 34, D158–D162.
- Levin, D.E. (2011). Regulation of cell wall biogenesis in *Saccharomyces cerevisiae*: the cell wall integrity signaling pathway. *Genetics* 189, 1145–1175.
- Li, B., Carey, M., and Workman, J.L. (2007). The role of chromatin during transcription. *Cell* 128, 707–719.

- Li, M., Chen, D., Shiloh, A., Luo, J., Nikolaev, A.Y., Qin, J., and Gu, W. (2002). Deubiquitination of p53 by HAUSP is an important pathway for p53 stabilization. *Nature* 416, 648–653.
- Li, S.-G., Zhou, H., Luo, Y.-P., Zhang, P., and Qu, L.-H. (2005). Identification and functional analysis of 20 Box H/ACA small nucleolar RNAs (snoRNAs) from *Schizosaccharomyces pombe*. *J. Biol. Chem.* 280, 16446–16455.
- Li, T., You, H., Zhang, J., Mo, X., He, W., Chen, Y., Tang, X., Jiang, Z., Tu, R., Zeng, L., et al. (2014). Study of GOLPH3: a potential stress-inducible protein from Golgi apparatus. *Mol. Neurobiol.* 49, 1449–1459.
- Liao, Y., Smyth, G.K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinforma. Oxf. Engl.* 30, 923–930.
- Lipford, J.R., Smith, G.T., Chi, Y., and Deshaies, R.J. (2005). A putative stimulatory role for activator turnover in gene expression. *Nature* 438, 113–116.
- Lister, R., Pelizzola, M., Kida, Y.S., Hawkins, R.D., Nery, J.R., Hon, G., Antosiewicz-Bourget, J., O'Malley, R., Castanon, R., Klugman, S., et al. (2011). Hotspots of aberrant epigenomic reprogramming in human induced pluripotent stem cells. *Nature* 471, 68–73.
- Liu, S., Lin, L., Jiang, P., Wang, D., and Xing, Y. (2011). A comparison of RNA-Seq and high-density exon array for detecting differential gene expression between closely related species. *Nucleic Acids Res.* 39, 578–588.
- Livak, K.J., and Schmittgen, T.D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods San Diego Calif* 25, 402–408.
- Lobo, N.A., Shimono, Y., Qian, D., and Clarke, M.F. (2007). The biology of cancer stem cells. *Annu. Rev. Cell Dev. Biol.* 23, 675–699.
- Lopez, R. (2003). WU-Blast2 server at the European Bioinformatics Institute. *Nucleic Acids Res.* 31, 3795–3798.
- Louro, R., Nakaya, H.I., Amaral, P.P., Festa, F., Sogayar, M.C., Silva, A.M. da, Verjovski-Almeida, S., and Reis, E.M. (2007). Androgen responsive intronic non-coding RNAs. *BMC Biol.* 5, 4.
- Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15, 550.
- Lowe, T.M., and Eddy, S.R. (1999). A computational screen for methylation guide snoRNAs in yeast. *Science* 283, 1168–1171.
- Lyne, R., Burns, G., Mata, J., Penkett, C.J., Rustici, G., Chen, D., Langford, C., Vetrie, D., and Bähler, J. (2003). Whole-genome microarrays of fission yeast: characteristics, accuracy, reproducibility, and processing of array data. *BMC Genomics* 4, 27.
- Maher, C.A., Kumar-Sinha, C., Cao, X., Kalyana-Sundaram, S., Han, B., Jing, X., Sam, L., Barrette, T., Palanisamy, N., and Chinnaiyan, A.M. (2009). Transcriptome sequencing to detect gene fusions in cancer. *Nature* 458, 97–101.

- Malik, S., and Roeder, R.G. (2010). The metazoan Mediator co-activator complex as an integrative hub for transcriptional regulation. *Nat. Rev. Genet.* *11*, 761–772.
- Marioni, J.C., Mason, C.E., Mane, S.M., Stephens, M., and Gilad, Y. (2008). RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. *Genome Res.* *18*, 1509–1517.
- Mata, J., Marguerat, S., and Bähler, J. (2005). Post-transcriptional control of gene expression: a genome-wide perspective. *Trends Biochem. Sci.* *30*, 506–514.
- Mattick, J.S. (2009). The genetic signatures of noncoding RNAs. *PLoS Genet.* *5*, e1000459.
- Menendez, D., Inga, A., and Resnick, M.A. (2009). The expanding universe of p53 targets. *Nat. Rev. Cancer* *9*, 724–737.
- Menzel, P., Gorodkin, J., and Stadler, P.F. (2009). The tedious task of finding homologous noncoding RNA genes. *RNA N. Y. N* *15*, 2075–2082.
- Mercer, T.R., Dinger, M.E., and Mattick, J.S. (2009). Long non-coding RNAs: insights into functions. *Nat. Rev. Genet.* *10*, 155–159.
- Mesa, R., Luo, S., Hoover, C.M., Miller, K., Minniti, A., Inestrosa, N., and Nonet, M.L. (2011). HID-1, a new component of the peptidergic signaling pathway. *Genetics* *187*, 467–483.
- Mizuno, H., Kawahara, Y., Sakai, H., Kanamori, H., Wakimoto, H., Yamagata, H., Oono, Y., Wu, J., Ikawa, H., Itoh, T., et al. (2010). Massive parallel sequencing of mRNA in identification of unannotated salinity stress-inducible transcripts in rice (*Oryza sativa* L.). *BMC Genomics* *11*, 683.
- Montgomery, S.B., Sammeth, M., Gutierrez-Arcelus, M., Lach, R.P., Ingle, C., Nisbett, J., Guigo, R., and Dermitzakis, E.T. (2010). Transcriptome genetics using second generation sequencing in a Caucasian population. *Nature* *464*, 773–777.
- Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L., and Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods* *5*, 621–628.
- Mosesson, Y., Mills, G.B., and Yarden, Y. (2008). Derailed endocytosis: an emerging feature of cancer. *Nat. Rev. Cancer* *8*, 835–850.
- Motamedi, M.R., Verdel, A., Colmenares, S.U., Gerber, S.A., Gygi, S.P., and Moazed, D. (2004). Two RNAi complexes, RITS and RDRC, physically interact and localize to noncoding centromeric RNAs. *Cell* *119*, 789–802.
- Mourier, T., and Jeffares, D.C. (2003). Eukaryotic intron loss. *Science* *300*, 1393.
- Mowbrey, K., and Dacks, J.B. (2009). Evolution and diversity of the Golgi body. *FEBS Lett.* *583*, 3738–3745.
- Muratani, M., and Tansey, W.P. (2003). How the ubiquitin-proteasome system controls transcription. *Nat. Rev. Mol. Cell Biol.* *4*, 192–201.
- Muratani, M., Kung, C., Shokat, K.M., and Tansey, W.P. (2005). The F box protein Dsg1/Mdm30 is a transcriptional coactivator that stimulates Gal4 turnover and cotranscriptional mRNA processing. *Cell* *120*, 887–899.

- Mutz, K.-O., Heilkenbrinker, A., Lönne, M., Walter, J.-G., and Stahl, F. (2013). Transcriptome analysis using next-generation sequencing. *Curr. Opin. Biotechnol.* *24*, 22–30.
- Nakaya, H.I., Amaral, P.P., Louro, R., Lopes, A., Fachel, A.A., Moreira, Y.B., El-Jundi, T.A., da Silva, A.M., Reis, E.M., and Verjovski-Almeida, S. (2007). Genome mapping and expression analyses of human intronic noncoding RNAs reveal tissue-specific patterns and enrichment in genes related to regulation of transcription. *Genome Biol.* *8*, R43.
- Naresh, A., Saini, S., and Singh, J. (2003). Identification of Uhp1, a ubiquitinated histone-like protein, as a target/mediator of Rhp6 in mating-type silencing in fission yeast. *J. Biol. Chem.* *278*, 9185–9194.
- Narlikar, G.J., Fan, H.-Y., and Kingston, R.E. (2002). Cooperation between Complexes that Regulate Chromatin Structure and Transcription. *Cell* *108*, 475–487.
- Negrini, M., Nicoloso, M.S., and Calin, G.A. (2009). MicroRNAs and cancer—new paradigms in molecular oncology. *Curr. Opin. Cell Biol.* *21*, 470–479.
- Ng, E.K.O., Chong, W.W.S., Jin, H., Lam, E.K.Y., Shin, V.Y., Yu, J., Poon, T.C.W., Ng, S.S.M., and Sung, J.J.Y. (2009). Differential expression of microRNAs in plasma of patients with colorectal cancer: a potential marker for colorectal cancer screening. *Gut* *58*, 1375–1381.
- Ng, M.M., Dippold, H.C., Buschman, M.D., Noakes, C.J., and Field, S.J. (2013). GOLPH3L antagonizes GOLPH3 to determine Golgi morphology. *Mol. Biol. Cell* *24*, 796–808.
- Nikolov, D.B., and Burley, S.K. (1997). RNA polymerase II transcription initiation: A structural view. *Proc. Natl. Acad. Sci.* *94*, 15–22.
- Nishida, N., Yano, H., Nishida, T., Kamura, T., and Kojiro, M. (2006). Angiogenesis in cancer. *Vasc. Health Risk Manag.* *2*, 213–219.
- Nizak, C., Martin-Lluesma, S., Moutel, S., Roux, A., Kreis, T.E., Goud, B., and Perez, F. (2003). Recombinant Antibodies Against Subcellular Fractions Used to Track Endogenous Golgi Protein Dynamics in Vivo. *Traffic* *4*, 739–753.
- Nurse, P., Thuriaux, P., and Nasmyth, K. (1976). Genetic control of the cell division cycle in the fission yeast *Schizosaccharomyces pombe*. *Mol. Gen. Genet. MGG* *146*, 167–178.
- Oliva, A., Rosebrock, A., Ferrezuelo, F., Pyne, S., Chen, H., Skiena, S., Futcher, B., and Leatherwood, J. (2005). The Cell Cycle–Regulated Genes of *Schizosaccharomyces pombe*. *PLoS Biol* *3*, e225.
- Onken, B., Wiener, H., Philips, M.R., and Chang, E.C. (2006). Compartmentalized signaling of Ras in fission yeast. *Proc. Natl. Acad. Sci. U. S. A.* *103*, 9045–9050.
- Osumi, M. (2012). Visualization of yeast cells by electron microscopy. *J. Electron Microsc. (Tokyo)* *61*, 343–365.
- Ouni, I., Flick, K., and Kaiser, P. (2011). Ubiquitin and transcription. *Transcription* *2*, 135–139.
- Pancaldi, V., Schubert, F., and Bähler, J. (2010). Meta-analysis of genome regulation and expression variability across hundreds of environmental and genetic perturbations in fission yeast. *Mol. Biosyst.* *6*, 543–552.

- Pantazopoulou, A., and Peñalva, M.A. (2009). Organization and dynamics of the *Aspergillus nidulans* Golgi during apical extension and mitosis. *Mol. Biol. Cell* 20, 4335–4347.
- Papanikou, E., and Glick, B.S. (2009). The yeast Golgi apparatus: Insights and mysteries. *FEBS Lett.* 583, 3746–3751.
- Parr, E. (2012). The default state of the cell: quiescence or proliferation? *BioEssays News Rev. Mol. Cell. Dev. Biol.* 34, 36–37.
- Peng, Y., Schwarz, E.J., Lazar, M.A., Genin, A., Spinner, N.B., and Taub, R. (1997). Cloning, human chromosomal assignment, and adipose and hepatic expression of the CL-6/INSIG1 gene. *Genomics* 43, 278–284.
- Phizicky, E.M., and Hopper, A.K. (2010). tRNA biology charges to the front. *Genes Dev.* 24, 1832–1860.
- Plath, K., Mlynarczyk-Evans, S., Nusinow, D.A., and Panning, B. (2002). Xist RNA and the mechanism of X chromosome inactivation. *Annu. Rev. Genet.* 36, 233–278.
- Polo, S., Pece, S., and Di Fiore, P.P. (2004). Endocytosis and cancer. *Curr. Opin. Cell Biol.* 16, 156–161.
- Prabhala, G., Rosenberg, G.H., and Käufer, N.F. (1992). Architectural features of pre-mRNA introns in the fission yeast *Schizosaccharomyces pombe*. *Yeast Chichester Engl.* 8, 171–182.
- Pruesse, E., Quast, C., Knittel, K., Fuchs, B.M., Ludwig, W., Peplies, J., and Glöckner, F.O. (2007). SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res.* 35, 7188–7196.
- Radhakrishnan, A., Sun, L.-P., Kwon, H.J., Brown, M.S., and Goldstein, J.L. (2004). Direct binding of cholesterol to the purified membrane region of SCAP: mechanism for a sterol-sensing domain. *Mol. Cell* 15, 259–268.
- Ramakers, C., Ruijter, J.M., Deprez, R.H.L., and Moorman, A.F. (2003). Assumption-free analysis of quantitative real-time polymerase chain reaction (PCR) data. *Neurosci. Lett.* 339, 62–66.
- Ramsköld, D., Wang, E.T., Burge, C.B., and Sandberg, R. (2009). An abundance of ubiquitously expressed genes revealed by tissue transcriptome sequence data. *PLoS Comput. Biol.* 5, e1000598.
- Randau, L., Schröder, I., and Söll, D. (2008). Life without RNase P. *Nature* 453, 120–123.
- Rapin, N., Bagger, F.O., Jendholm, J., Mora-Jensen, H., Krogh, A., Kohlmann, A., Thiede, C., Borregaard, N., Bullinger, L., Winther, O., et al. (2014). Comparing cancer vs normal gene expression profiles identifies new disease entities and common transcriptional programs in AML patients. *Blood* 123, 894–904.
- Ratner, J.N., Balasubramanian, B., Corden, J., Warren, S.L., and Bregman, D.B. (1998). Ultraviolet Radiation-induced Ubiquitination and Proteasomal Degradation of the Large Subunit of RNA Polymerase II. IMPLICATIONS FOR TRANSCRIPTION-COUPLED DNA REPAIR. *J. Biol. Chem.* 273, 5184–5189.
- Rawson, R.B. (2003). The SREBP pathway — insights from insects and insects. *Nat. Rev. Mol. Cell Biol.* 4, 631–640.

- Rawson, R.B., DeBose-Boyd, R., Goldstein, J.L., and Brown, M.S. (1999). Failure to cleave sterol regulatory element-binding proteins (SREBPs) causes cholesterol auxotrophy in Chinese hamster ovary cells with genetic absence of SREBP cleavage-activating protein. *J. Biol. Chem.* *274*, 28549–28556.
- Rinn, J.L., Kertesz, M., Wang, J.K., Squazzo, S.L., Xu, X., Bruggmann, S.A., Goodnough, L.H., Helms, J.A., Farnham, P.J., Segal, E., et al. (2007). Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell* *129*, 1311–1323.
- Risk, J.M., Ruhrberg, C., Hennies, H., Mills, H.S., Di Colandrea, T., Evans, K.E., Ellis, A., Watt, F.M., Bishop, D.T., Spurr, N.K., et al. (1999). Envoplakin, a possible candidate gene for focal NEPPK/esophageal cancer (TOC): the integration of genetic and physical maps of the TOC region on 17q25. *Genomics* *59*, 234–242.
- Ritzenthaler, C., Nebenführ, A., Movafeghi, A., Stussi-Garaud, C., Behnia, L., Pimpl, P., Staehelin, L.A., and Robinson, D.G. (2002). Reevaluation of the effects of brefeldin A on plant cells using tobacco Bright Yellow 2 cells expressing Golgi-targeted green fluorescent protein and COPI antisera. *Plant Cell* *14*, 237–261.
- Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. *Nat. Biotechnol.* *29*, 24–26.
- Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* *26*, 139–140.
- Rogozin, I.B., Carmel, L., Csuros, M., and Koonin, E. V (2012). Origin and evolution of spliceosomal introns. *Biol. Direct* *7*, 11.
- Roguev, A., Bandyopadhyay, S., Zofall, M., Zhang, K., Fischer, T., Collins, S.R., Qu, H., Shales, M., Park, H.-O., Hayles, J., et al. (2008). Conservation and rewiring of functional modules revealed by an epistasis map in fission yeast. *Science* *322*, 405–410.
- Rossanese, O.W., Soderholm, J., Bevis, B.J., Sears, I.B., O'Connor, J., Williamson, E.K., and Glick, B.S. (1999). Golgi structure correlates with transitional endoplasmic reticulum organization in *Pichia pastoris* and *Saccharomyces cerevisiae*. *J. Cell Biol.* *145*, 69–81.
- Rossi, L., Bonmassar, E., and Faraoni, I. (2007). Modification of miR gene expression pattern in human colon cancer cells following exposure to 5-fluorouracil in vitro. *Pharmacol. Res. Off. J. Ital. Pharmacol. Soc.* *56*, 248–253.
- Rothstein, R.J. (1983). One-step gene disruption in yeast. *Methods Enzymol.* *101*, 202–211.
- Roux, A.E., Leroux, A., Alaamery, M.A., Hoffman, C.S., Chartrand, P., Ferbeyre, G., and Rokeach, L.A. (2009). Pro-aging effects of glucose signaling through a G protein-coupled glucose receptor in fission yeast. *PLoS Genet.* *5*, e1000408.
- Roux, A.E., Arseneault, G., Chartrand, P., Ferbeyre, G., and Rokeach, L.A. (2010). A screen for genes involved in respiration control and longevity in *Schizosaccharomyces pombe*. *Ann. N. Y. Acad. Sci.* *1197*, 19–27.

- Ruby, J.G., Jan, C., Player, C., Axtell, M.J., Lee, W., Nusbaum, C., Ge, H., and Bartel, D.P. (2006). Large-scale sequencing reveals 21U-RNAs and additional microRNAs and endogenous siRNAs in *C. elegans*. *Cell* *127*, 1193–1207.
- Ryan, C.J., Roguev, A., Patrick, K., Xu, J., Jahari, H., Tong, Z., Beltrao, P., Shales, M., Qu, H., Collins, S.R., et al. (2012). Hierarchical modularity and the evolution of genetic interactomes across species. *Mol. Cell* *46*, 691–704.
- Sahoo, T., del Gaudio, D., German, J.R., Shinawi, M., Peters, S.U., Person, R.E., Garnica, A., Cheung, S.W., and Beaudet, A.L. (2008). Prader-Willi phenotype caused by paternal deficiency for the HBII-85 C/D box small nucleolar RNA cluster. *Nat. Genet.* *40*, 719–721.
- Sajiki, K., Hatanaka, M., Nakamura, T., Takeda, K., Shimanuki, M., Yoshida, T., Hanyu, Y., Hayashi, T., Nakaseko, Y., and Yanagida, M. (2009). Genetic control of cellular quiescence in *S. pombe*. *J. Cell Sci.* *122*, 1418–1429.
- Salditt-Georgieff, M., Harpold, M.M., Wilson, M.C., and Darnell, J.E. (1981). Large heterogeneous nuclear ribonucleic acid has three times as many 5' caps as polyadenylic acid segments, and most caps do not enter polyribosomes. *Mol. Cell. Biol.* *1*, 179–187.
- Samarsky, D.A., Ferbeyre, G., Bertrand, E., Singer, R.H., Cedergren, R., and Fournier, M.J. (1999). A small nucleolar RNA:ribozyme hybrid cleaves a nucleolar RNA target in vivo with near-perfect efficiency. *Proc. Natl. Acad. Sci. U. S. A.* *96*, 6609–6614.
- Sambrook, J., Fritsch, E.F., Maniatis, T., and others (1989). *Molecular cloning* (Cold spring harbor laboratory press New York).
- Sánchez-Mir, L., Soto, T., Franco, A., Madrid, M., Viana, R.A., Vicente, J., Gacto, M., Pérez, P., and Cansado, J. (2014). Rho1 GTPase and PKC ortholog Pck1 are upstream activators of the cell integrity MAPK pathway in fission yeast. *PloS One* *9*, e88020.
- Sauer, B. (1998). Inducible Gene Targeting in Mice Using the Cre/loxSystem. *Methods* *14*, 381–392.
- Schmittgen, T.D., and Livak, K.J. (2008). Analyzing real-time PCR data by the comparative C(T) method. *Nat. Protoc.* *3*, 1101–1108.
- Schneider, C.A., Rasband, W.S., and Eliceiri, K.W. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* *9*, 671–675.
- Schnell, J.D., and Hicke, L. (2003). Non-traditional functions of ubiquitin and ubiquitin-binding proteins. *J. Biol. Chem.* *278*, 35857–35860.
- Scott, K.L., and Chin, L. (2010). Signaling from the Golgi: mechanisms and models for Golgi phosphoprotein 3-mediated oncogenesis. *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res.* *16*, 2229–2234.
- Sechi, S., Frappaolo, A., Belloni, G., Colotti, G., and Giansanti, M.G. (2015). The multiple cellular functions of the oncoprotein Golgi phosphoprotein 3. *Oncotarget* *6*, 3493–3506.
- Selcuklu, S.D., Yakicier, M.C., and Erson, A.E. (2009). An investigation of microRNAs mapping to breast cancer related genomic gain and loss regions. *Cancer Genet. Cytogenet.* *189*, 15–23.

- Seshasayee, A.S.N., Sivaraman, K., and Luscombe, N.M. (2011). An overview of prokaryotic transcription factors : a summary of function and occurrence in bacterial genomes. *Subcell. Biochem.* 52, 7–23.
- Shao, W., and Espenshade, P.J. (2014). Sterol regulatory element-binding protein (SREBP) cleavage regulates Golgi-to-endoplasmic reticulum recycling of SREBP cleavage-activating protein (SCAP). *J. Biol. Chem.* 289, 7547–7557.
- Sideri, T., Rallis, C., Bitton, D.A., Lages, B.M., Suo, F., Rodríguez-López, M., Du, L.-L., and Bähler, J. (2014). Parallel profiling of fission yeast deletion mutants for proliferation and for lifespan during long-term quiescence. *G3 Bethesda Md* 5, 145–155.
- Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., et al. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* 7, 539.
- Smith, E., and Shilatifard, A. (2010). The chromatin signaling pathway: diverse mechanisms of recruitment of histone-modifying enzymes and varied biological outcomes. *Mol. Cell* 40, 689–701.
- Smyth, G.K. (2004). Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat. Appl. Genet. Mol. Biol.* 3, Article3.
- Snyder, C.M., Mardones, G.A., Ladinsky, M.S., and Howell, K.E. (2006). GMx33 associates with the trans-Golgi matrix in a dynamic manner and sorts within tubules exiting the Golgi. *Mol. Biol. Cell* 17, 511–524.
- Sollner-Webb, B., and Mougey, E.B. (1991). News from the nucleolus: rRNA gene expression. *Trends Biochem. Sci.* 16, 58–62.
- Somesh, B.P., Reid, J., Liu, W.-F., Sjøgaard, T.M.M., Erdjument-Bromage, H., Tempst, P., and Svejstrup, J.Q. (2005). Multiple mechanisms confining RNA polymerase II ubiquitylation to polymerases undergoing transcriptional arrest. *Cell* 121, 913–923.
- Stajich, J.E., Dietrich, F.S., and Roy, S.W. (2007). Comparative genomic analysis of fungal genomes reveals intron-rich ancestors. *Genome Biol.* 8, R223.
- Stefani, G., and Slack, F.J. (2008). Small non-coding RNAs in animal development. *Nat. Rev. Mol. Cell Biol.* 9, 219–230.
- Steinberg, D. (2002). Atherogenesis in perspective: Hypercholesterolemia and inflammation as partners in crime. *Nat. Med.* 8, 1211–1217.
- Stewart, E. V, Nwosu, C.C., Tong, Z., Roguev, A., Cummins, T.D., Kim, D.-U., Hayles, J., Park, H.-O., Hoe, K.-L., Powell, D.W., et al. (2011). Yeast SREBP cleavage activation requires the Golgi Dsc E3 ligase complex. *Mol. Cell* 42, 160–171.
- Su, H., Xu, T., Ganapathy, S., Shadfian, M., Long, M., Huang, T.H.-M., Thompson, I., and Yuan, Z.-M. (2014a). Elevated snoRNA biogenesis is essential in breast cancer. *Oncogene* 33, 1348–1358.

- Su, L., Liu, X., Chai, N., Lv, L., Wang, R., Li, X., Nie, Y., Shi, Y., and Fan, D. (2014b). The transcription factor FOXO4 is down-regulated and inhibits tumor proliferation and metastasis in gastric cancer. *BMC Cancer* *14*, 378.
- Suda, Y., and Nakano, A. (2012). The Yeast Golgi Apparatus. *Traffic* *13*, 505–510.
- Sudarsan, N., Barrick, J.E., and Breaker, R.R. (2003). Metabolite-binding RNA domains are present in the genes of eukaryotes. *RNA* *9*, 644–647.
- Supek, F., Bošnjak, M., Škunca, N., and Šmuc, T. (2011). REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS One* *6*, e21800.
- Sveiczner, A., Tyson, J.J., and Novak, B. (2004). Modelling the fission yeast cell cycle. *Brief. Funct. Genomic. Proteomic.* *2*, 298–307.
- Svejstrup, J.Q. (2007). Contending with transcriptional arrest during RNAPII transcript elongation. *Trends Biochem. Sci.* *32*, 165–171.
- Tafer, H., Kehr, S., Hertel, J., Hofacker, I.L., and Stadler, P.F. (2010). RNAsnoop: efficient target prediction for H/ACA snoRNAs. *Bioinforma. Oxf. Engl.* *26*, 610–616.
- Takegawa, K., Iwaki, T., Fujita, Y., Morita, T., Hosomi, A., and Tanaka, N. (2003). Vesicle-mediated Protein Transport Pathways to the Vacuole in *Schizosaccharomyces pombe*. *Cell Struct. Funct.* *28*, 399–417.
- Tamanoi, F. (2011). Ras signaling in yeast. *Genes Cancer* *2*, 210–215.
- Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S. (2011). MEGA5: Molecular Evolutionary Genetics Analysis Using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. *Mol. Biol. Evol.* *28*, 2731–2739.
- Thorrez, L., Laudadio, I., Van Deun, K., Quintens, R., Hendrickx, N., Granvik, M., Lemaire, K., Schraenen, A., Van Lommel, L., Lehnert, S., et al. (2011). Tissue-specific disallowance of housekeeping genes: the other face of cell differentiation. *Genome Res.* *21*, 95–105.
- Tian, Q., Stepaniants, S.B., Mao, M., Weng, L., Feetham, M.C., Doyle, M.J., Yi, E.C., Dai, H., Thorsson, V., Eng, J., et al. (2004). Integrated genomic and proteomic analyses of gene expression in Mammalian cells. *Mol. Cell. Proteomics MCP* *3*, 960–969.
- Tjaden, B. (2015). De novo assembly of bacterial transcriptomes from RNA-seq data. *Genome Biol.* *16*, 1.
- Tocchini-Valentini, G.D., Fruscoloni, P., and Tocchini-Valentini, G.P. (2009). Processing of multiple-intron-containing pre-tRNA. *Proc. Natl. Acad. Sci.* *106*, 20246–20251.
- Todd, B.L., Stewart, E. V., Burg, J.S., Hughes, A.L., and Espenshade, P.J. (2006). Sterol regulatory element binding protein is a principal regulator of anaerobic gene expression in fission yeast. *Mol. Cell. Biol.* *26*, 2817–2831.
- Tollervey, D., and Kiss, T. (1997). Function and synthesis of small nucleolar RNAs. *Curr. Opin. Cell Biol.* *9*, 337–342.

- Tontonoz, P., Kim, J.B., Graves, R.A., and Spiegelman, B.M. (1993). ADD1: a novel helix-loop-helix transcription factor associated with adipocyte determination and differentiation. *Mol. Cell. Biol.* *13*, 4753–4759.
- Torres-Machorro, A.L., Hernández, R., Cevallos, A.M., and López-Villaseñor, I. (2010). Ribosomal RNA genes in eukaryotic microorganisms: witnesses of phylogeny? *FEMS Microbiol. Rev.* *34*, 59–86.
- Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: Discovering splice junctions with RNA-Seq. *Bioinformatics* *25*, 1105–1111.
- Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J., Salzberg, S.L., Wold, B.J., and Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* *28*, 511–515.
- Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D.R., Pimentel, H., Salzberg, S.L., Rinn, J.L., and Pachter, L. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protoc.* *7*, 562–578.
- Twine, N.A., Janitz, K., Wilkins, M.R., and Janitz, M. (2011). Whole transcriptome sequencing reveals gene expression and splicing differences in brain regions affected by Alzheimer’s disease. *PLoS One* *6*, e16266.
- Vaquerezas, J.M., Kummerfeld, S.K., Teichmann, S.A., and Luscombe, N.M. (2009). A census of human transcription factors: function, expression and evolution. *Nat. Rev. Genet.* *10*, 252–263.
- van der Vlag, J., and Otte, A.P. (1999). Transcriptional repression mediated by the human polycomb-group protein EED involves histone deacetylation. *Nat. Genet.* *23*, 474–478.
- Wang, C., Gong, B., Bushel, P.R., Thierry-Mieg, J., Thierry-Mieg, D., Xu, J., Fang, H., Hong, H., Shen, J., Su, Z., et al. (2014a). The concordance between RNA-seq and microarray data depends on chemical treatment and transcript abundance. *Nat. Biotechnol.* *32*, 926–932.
- Wang, E.T., Sandberg, R., Luo, S., Khrebtkova, I., Zhang, L., Mayr, C., Kingsmore, S.F., Schroth, G.P., and Burge, C.B. (2008). Alternative isoform regulation in human tissue transcriptomes. *Nature* *456*, 470–476.
- Wang, H., Wang, L., Erdjument-Bromage, H., Vidal, M., Tempst, P., Jones, R.S., and Zhang, Y. (2004). Role of histone H2A ubiquitination in Polycomb silencing. *Nature* *431*, 873–878.
- Wang, L., Zhan, Y., Song, E., Yu, Y., Jiu, Y., Du, W., Lu, J., Liu, P., Xu, P., and Xu, T. (2011). HID-1 is a peripheral membrane protein primarily associated with the medial- and trans- Golgi apparatus. *Protein Cell* *2*, 74–85.
- Wang, Z., Gerstein, M., and Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.* *10*, 57–63.
- Wang, Z., Jiang, B., Chen, L., Di, J., Cui, M., Liu, M., Ma, Y., Yang, H., Xing, J., Zhang, C., et al. (2014b). GOLPH3 predicts survival of colorectal cancer patients treated with 5-fluorouracil-based adjuvant chemotherapy. *J. Transl. Med.* *12*, 15.

- Watson, J.D., and Crick, F.H.C. (1953). Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Publ. Online* 25 April 1953 Doi101038171737a0 171, 737–738.
- Watson, A.T., Garcia, V., Bone, N., Carr, A.M., and Armstrong, J. (2008). Gene tagging and gene replacement using recombinase-mediated cassette exchange in *Schizosaccharomyces pombe*. *Gene* 407, 63–74.
- Weston, A., and Harris, C.C. (2003). Multistage Carcinogenesis. In *Holland-Frei Cancer Medicine*, D.W. Kufe, R.E. Pollock, and R.R. Weichselbaum, eds. (BC Decker),.
- White, A.C., Khuu, J.K., Dang, C.Y., Hu, J., Tran, K. V, Liu, A., Gomez, S., Zhang, Z., Yi, R., Scumpia, P., et al. (2014). Stem cell quiescence acts as a tumour suppressor in squamous tumours. *Nat. Cell Biol.* 16, 99–107.
- Will, C.L., and Lührmann, R. (2011). Spliceosome structure and function. *Cold Spring Harb. Perspect. Biol.* 3.
- Wilson, C., Venditti, R., Rega, L.R., Colanzi, A., D’Angelo, G., and De Matteis, M.A. (2011). The Golgi apparatus: an organelle with multiple complex functions. *Biochem. J.* 433, 1–9.
- Winkler, W., Nahvi, A., and Breaker, R.R. (2002). Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression. *Nature* 419, 952–956.
- Wood, V., Gwilliam, R., Rajandream, M.-A., Lyne, M., Lyne, R., Stewart, A., Sgouros, J., Peat, N., Hayles, J., Baker, S., et al. (2002). The genome sequence of *Schizosaccharomyces pombe*. *Nature* 415, 871–880.
- Wu, C.C., Taylor, R.S., Lane, D.R., Ladinsky, M.S., Weisz, J.A., and Howell, K.E. (2000). GMx33: a novel family of trans-Golgi proteins identified by proteomics. *Traffic Cph. Den.* 1, 963–975.
- Wyers, F., Rougemaille, M., Badis, G., Rousselle, J.-C., Dufour, M.-E., Boulay, J., Régnault, B., Devaux, F., Namane, A., Séraphin, B., et al. (2005). Cryptic pol II transcripts are degraded by a nuclear quality control pathway involving a new poly(A) polymerase. *Cell* 121, 725–737.
- Xie, J., Zhang, M., Zhou, T., Hua, X., Tang, L., and Wu, W. (2007). Sno/scaRNAbase: a curated database for small nucleolar RNAs and cajal body-specific RNAs. *Nucleic Acids Res.* 35, D183–D187.
- Xu, K., Cui, J., Olman, V., Yang, Q., Puett, D., and Xu, Y. (2010). A Comparative Analysis of Gene-Expression Data of Multiple Cancer Types. *PLoS ONE* 5.
- Yanagida, M. (2002). The model unicellular eukaryote, *Schizosaccharomyces pombe*. *Genome Biol.* 3, comment2003.1–comment2003.4.
- Yang, J.-H., Zhang, X.-C., Huang, Z.-P., Zhou, H., Huang, M.-B., Zhang, S., Chen, Y.-Q., and Qu, L.-H. (2006). snoSeeker: an advanced computational package for screening of guide and orphan snoRNA genes in the human genome. *Nucleic Acids Res.* 34, 5112–5123.
- Yao, T., and Ndoja, A. (2012). Regulation of gene expression by the ubiquitin-proteasome system. *Semin. Cell Dev. Biol.* 23, 523–529.

- Ye, S.Q., Lavoie, T., Usher, D.C., and Zhang, L.Q. (2002). Microarray, SAGE and their applications to cardiovascular diseases. *Cell Res.* *12*, 105–115.
- Yu, Y., Wang, L., Jiu, Y., Zhan, Y., Liu, L., Xia, Z., Song, E., Xu, P., and Xu, T. (2011). HID-1 is a novel player in the regulation of neuropeptide sorting. *Biochem. J.* *434*, 383–390.
- Zaitsevskaya-Carter, T., and Cooper, J.A. (1997). Spm1, a stress-activated MAP kinase that regulates morphogenesis in *S.pombe*. *EMBO J.* *16*, 1318–1331.
- Zeeberg, B., Feng, W., Wang, G., Wang, M., Fojo, A., Sunshine, M., Narasimhan, S., Kane, D., Reinhold, W., Lababidi, S., et al. (2003). GoMiner: a resource for biological interpretation of genomic and proteomic data. *Genome Biol.* *4*, R28.
- Zeng, Z., Lin, H., Zhao, X., Liu, G., Wang, X., Xu, R., Chen, K., Li, J., and Song, L. (2012). Overexpression of GOLPH3 promotes proliferation and tumorigenicity in breast cancer via suppression of the FOXO1 transcription factor. *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res.* *18*, 4059–4069.
- Zhang, Y. (2003). Transcriptional regulation by histone ubiquitination and deubiquitination. *Genes Dev.* *17*, 2733–2740.
- Zhang, Z., Xu, J., Koppetsch, B.S., Wang, J., Tipping, C., Ma, S., Weng, Z., Theurkauf, W.E., and Zamore, P.D. (2011). Heterotypic piRNA Ping-Pong requires qin, a protein with both E3 ligase and Tudor domains. *Mol. Cell* *44*, 572–584.
- Zhou, W., Zhu, P., Wang, J., Pascual, G., Ohgi, K.A., Lozach, J., Glass, C.K., and Rosenfeld, M.G. (2008). Histone H2A monoubiquitination represses transcription by inhibiting RNA polymerase II transcriptional elongation. *Mol. Cell* *29*, 69–80.
- Zhu, J., He, F., Song, S., Wang, J., and Yu, J. (2008). How many human genes can be defined as housekeeping with current expression data? *BMC Genomics* *9*, 172.
- Zhu, J., Jiang, X., and Chehab, F.F. (2014). FoxO4 interacts with the sterol regulatory factor SREBP2 and the hypoxia inducible factor HIF2 α at the CYP51 promoter. *J. Lipid Res.* *55*, 431–442.
- (1998). Heterologous modules for efficient and versatile PCR-based gene targeting in *Schizosaccharomyces pombe*. *Yeast* *14*, 943–951.

