VNIVERSITAT DE VALÈNCIA

Facultad de Ciencias Biológicas

Programa oficial de doctorado en Biotecnología

Functional genomic study of the response to hypoxic and oxidative stress in *Candida albicans*

Trabajo realizado en el Departamento de Bioquímica y Biología Molecular de la Universitat de València por

Ana Miguel Blanco

Para optar al grado de Doctora por la Universitat de València

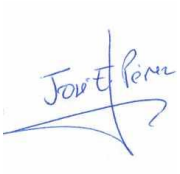
Trabajo dirigido por el

Dr. José Enrique Pérez Ortín

Abril de 2017

José Enrique Pérez Ortín, Doctor en Ciencias Biológicas y Catedrático del Departamento de Bioquímica y Biología Molecular de la Universitat de València, informa:

Ana Miguel Blanco, licenciada en Bioquímica por la Universitat de València, ha realizado bajo mi dirección el trabajo bajo el título "Functional genomic study of the response to hypoxic and oxidative stress in *Candida albicans*", el cual presenta para optar al grado de Doctora por la Universitat de València.



Valencia, Marzo de 2017

Fdo. Dr. José Enrique Pérez Ortín

One might not expect that yeast lead terribly eventful lives, yet the single-celled fungus must struggle to survive, just like everyone else.

PloS Biology, V1, I 2, P140

A la tata.

A la yaya.

Al abuelo de mi corazón.

Acknowledgements

Antes de entrar en materia, unas líneas para agradecer a todos esos que me ayudaron a hacer esto.

El primero, no por ser el orden políticamente correcto sino porque sin él esto no estaría siendo escrito, es mi director de tesis. Gracias, no sólo por haber soportado mis preguntas impertinentes de estudiante de 4º de carrera, sino por darme entonces la oportunidad de responderlas yo misma. Gracias por todo el apoyo en lo personal y en lo profesional, y gracias, sobre todo, por haber tolerado religiosamente que todos estos años haya compartido mi interés por la biología molecular con mi empeño en paralelo en salvar tortugas marinas.

Gracias a mi familia GFL, por ser la parte más grande de todo esto. Tendría que escribir otra tesis sólo para agradeceros todo lo que habéis hecho todos por mí. Gracias Fany, Daniel y María por la infinita ayuda, porque siempre he podido contar con vosotros cuando os he necesitado. Gracias Seta por ser mi inspiración en lo personal y en lo profesional, por los viajes, los festivales, y todo lo que viví contigo que me alegró la vida de mil maneras. Gracias Nati por haberme enseñado el valiosísimo concepto del "High Throughput". Gracias Fofinha por haberme dado una de las lecciones de vida más grandes que nunca nadie me ha dado. Antonio, creo que no exagero lo más mínimo si digo que cada página de esta tesis debería llevar un pie de página que diga "Gracias, Toni". Has sido el mejor amigo, compañero, hermano mayor que me pudo tocar en suerte. Esta tesis es tan tuya como mía. Thing, thank you for lighting up my life every day of this thesis with your awesomeness. I had no idea how lucky I was when you first walked into the room that day. I am incredibly grateful for every second we spent together. I wouldn't have done this without you. Gracias a todos los estudiantes que tuve, Toni, Jordi, y especialmente a Lucas que acabó motivando un capítulo de esta tesis. Y millones de gracias a todos los doctorandos visitantes que llegaron para alegrarnos las horas de laboratorio. Especialmente a todo el grupo de Sebastián Chávez: Douglas, Lidia, Lola, Gonzalo, y Xenia, que siempre fueron nuestros hermanos de pipeta en la distancia.

Gracias a todo el personal del Departamento de Bioquímica y Biología Molecular por ser esa "familia del pasillo" que te ayuda tantísimo en tantas cosas del día a día. A cada profesor y a cada doctorando con el que he compartido mi tiempo allí le estoy eternamente agradecida por el apoyo, los consejos, y el cariño con el que me rodearon cada día.

Una de las suertes más grandes que tuve con este trabajo fue que mi tesis salió de un proyecto de colaboración internacional. Gracias a ello tuve el honor de trabajar con 3 grupos de investigación distintos en Madrid, Lleida, y Alemania de los que aprendí muchísimo. Millones de gracias Jesús Pla y Enrique Herrero por tantísimos consejos y buenos momentos en meetings del Pathogenomics y congresos alrededor del mundo. Os admiro profundamente como científicos pero os admiro el doble como personas.

Gracias Elvira por ser tan auténtica y haber hecho la ciencia una de las cosas más divertidas que he hecho en mi vida. A thousand times thank you Joachim for letting me stay in your lab during that summer and making me part of it. My stay in Germany is one of the best memories I have of the last 4 years. Special thanks to Mateusz, Prashant, Mario, Marc and Christoph for so much help during my time there.

Detrás de todo el mundo científico está todo ese mundo que te empuja a seguir en esto cada día, sin ser conscientes de que lo hacen la mayor parte del tiempo. Unas de las más importantes fueron mis queridas Lligaldas y todo lo que les rodeó. Gracias hermanas por haber hecho de todos mis martes y fines de semana mi mejor escape al mundo real.

Como parte de ese equipo salió mi familia Valenciana que acabaron siendo las verdaderas supervivientes de esta tesis: el Bunker Carapace 3.0. Nadie más que vosotras sabe lo que habéis significado para mí en esta maratón. Gracias Pauli por todos y cada uno de los ron-colas que salvaron todos mis días, por ser la mejor cocinera a las 11:00 de la noche por esperar a que llegara a cenar con vosotras. Gracias Persi por haber sido mi mejor novia, psicóloga, madre master chef y el mayor personaje que he conocido y nunca conoceré. Gracias Peanut, por haber sido, ser, y porque sé que siempre serás una grandísima parte de mi vida. Veo una vida nueva, y tú siempre estás en ella.

Gracias princesa por ser la viva definición de la amistad incondicional y estar ahí en cada paso que doy. Porque eres, simple y llanamente, lo mejor que me ha pasado en la vida. Ni esto ni nada de lo que hago hubiera tenido ningún sentido sin ti.

Finalmente, infinitas gracias a toda mi familia por haberme hecho llegar hasta aquí. Especiales gracias a 'Sanca' y 'Tun' por haberme visto crecer y seguir en el día a día de mi vida 30 años después.

Esta tesis (y mi vida entera) va dedicada a mis padres. Thank you for making me who I am today and for your endless love and support. I look up to you and all I aim for in life is get to be half as great as you are. From you I learnt the most valuable lesson every scientist learns over a PhD: never giving up. I am and will always be incredibly thankful for everything you have done for me.

And thank you Rose, for getting back in my life and making it what I always wanted it to be.

Abstract

Candida albicans is the most prevalent human fungal pathogen. Despite being a harmless commensal organism, it is also an opportunistic fungus which can cause life-threatening infections in immunocompromised people. Although major virulence factors have been characterized, the exact mechanism of *Candida* pathogenesis still remains unknown. Mechanisms of adaptation under the different environments it faces during body invasion are yet poorly characterized.

In this thesis, functional genomic analyses were utilized to study the global response of *C. albicans* to two environmental insults it faces during invasion of the human body: oxidative and hypoxic stress. Several genes were found to be involved in the adaptation to these stresses, as reflected by the up or down regulation of their mRNA and protein products. Interestingly, the expression of the genes involved in the response to oxidative stress was found to be regulated mainly at the post-transcriptional level.

To explore the possible implication of non-coding RNAs in the response to these stresses, an online web application was created to enable the detection of ncRNAs which levels change under stress conditions. Two groups of 154 and 159 ncRNAs, including intergenic and antisense to ORFs, were found to be involved in the responses to hypoxia and oxidative stress respectively.

To study changes in nascent transcription under these stresses, Genomic run-on (GRO), a well-established technique used for the study of the nascentome in *Saccharomyces cerevisiae*, was implemented for its use in *C. albicans* in its most up-to-date version, BioGROseq. The GRO technique was later used for the study of changes in gene expression parameters at different growth temperatures, such as RNApol II density or translation rate, using the model yeast *S. cerevisiae* as a tool. The results showed that RNA pol II density decreases as growth temperature increases, and that this decrease in density is compensated with an increase in RNA pol II speed to maintain mRNA homeostatic levels, compensation that is regulated at the RNA pol II initiation level. Within the optimal range of growth temperatures in *S. cerevisiae* (26-34 ^{o}C) both total RNA and mRNA amounts were

found to be kept at homeostatic levels, decreasing markedly at temperatures above and below. Finally, an increase in translation rate was observed with the increase in growth temperatures. Overall, the results presented in this thesis provide clues into the regulatory mechanisms that yeast utilize to grow at different temperatures and contribute to the understanding of how *C. albicans* thrives during infection of the human body.

Resumen

Candida albicans es el patógeno humano más prevalente. A pesar de ser un organismo comensal, también es un hongo oportunista que puede causar infecciones de seria gravedad en personas inmunodeprimidas. Aunque los principales factores de virulencia han sido caracterizados, el proceso exacto de patogénesis en *Candida* todavía se desconoce. Los mecanismos de adaptación a los distintos ambientes a los que se enfrenta durante la invasión del cuerpo humano todavía no están suficientemente caracterizados.

En este trabajo se usaron análisis genómicos funcionales para el estudio de la respuesta global de *C. albicans* a dos estreses que encuentra durante la invasión del cuerpo humano: estrés oxidativo e hipóxico. En el análisis se encontraron varios genes implicados en la adaptación a estos estreses, lo cual se vio reflejado en la regulación de sus niveles de RNA mensajero y proteína. Un hallazgo interesante de este estudio fue que la regulación de los genes implicados en la respuesta a estrés oxidativo ocurre principalmente a nivel post-transcripcional.

Para estudiar la posible implicación de RNAs no codificantes en la respuesta a estos estreses, se creó una aplicación web que permitió la detección de aquellos ncRNAs cuya expresión se vio alterada durante el estrés. Dos grupos de 154 y 159 nRNAs, incluyendo intergénicos y antisentido a ORFs, se identificaron como implicados en la respuesta a estrés hipóxico y oxidativo, respectivamente.

Para el estudio de cambios en transcripción naciente durante estos estreses, la técnica de Genomic run-on (GRO), ya establecida para el estudio del nascentoma de *Saccharomyces cerevisiae*, se implementó para su uso en *C. albicans* en su versión más actualizada, el BioGROseq. La técnica de GRO fue después usada para el estudio de cambios en parámetros de estudio de la expresión génica a distintas temperaturas de crecimiento, tales como la densidad de RNApol II o la tasa de traducción, usando el organismo modelo *S. cerevisiae* como herramienta. Los resultados mostraron que la densidad de RNApol II disminuye con el incremento en la temperatura de crecimiento, y que esta disminución en densidad se compensa con un incremento en la velocidad de elongación de la RNApol II para mantener

la homeostasis de los niveles de mRNA. Esta compensación está regulada a nivel de iniciación de la RNApol.

Dentro del rango óptimo de temperaturas de crecimiento en *S. cerevisiae* (26-34 ^{o}C), tanto los niveles de RNA total como los de mRNA se mantienen en homeostasis, disminuyendo marcadamente a temperaturas inferiores o superiores. Finalmente, con el aumento en la temperatura de crecimiento se observó un aumento en la tasa de traducción.

En general, los resultados presentados en este trabajo proporcionan nuevos hallazgos sobre los mecanismos de regulación que las levaduras utilizan para su crecimiento a distintas temperaturas, y ayudan a entender cómo *C. albicans* sobrevive durante su invasión en el cuerpo humano.

Contents

RESULTS

Abbreviations

µg	micrograms
µM	micromolar
2DE	2-Dimensional gel electrophoresis
6-AU	6-Azaurazil
A	Absorbance/ Adenine
Ab	Antibody
AS	Antisense
ATP	Adenosine Tri-Phosphate
BioGRO	Biotin Genomic run-on
BioGROseq	Biotin Genomic run-on followed by sequencing
Bio-UTP	Biotin UTP
bp	Base pair
BRE	TFIIB Recognition Element
CAT	Catalase
CBC	Cap-binding protein
cDNA	Copy DNA
ChIP	Chromatin Immunoprecipitation
Ci	Curie
cm	centimeters
CPF	Cleavage polyadenylation factor
cpm	counts per minute
cps	counts per second
CPU	Central processing unit
cRNA	Copy RNA
CTD	C-terminal domain
CTP	Cytosine Tri-Phosphate
Cu	Copper
CUT	Cryptic unstable transcript
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleotide
DPE	Downstream promoter element
DRp	Protein degradation rate
DRr	mRNA Degradation rate
dT	Deoxytimine
DTT	Dithiothreitol
e.g.	*Exempli gratia* (for example)
EDTA	Ethylenediaminetetraacetic acid
EtOH	Ethylic alcohol
F	Forward (DNA strand)
G	Guanine
g	grams

GMP	Guanidine Tri-Phosphate
GRO	Genomic run-on
GRX	Glutaredoxin
GTF	General transcription factor
GTP	Guanosine Tri-Phosphate
h	hours
H_2O_2	Hydrogen peroxide
HCl	Hydrochloric acid
HRP	Horse radish peroxidase
Hsp	Heat-shock protein
i.e.	For example
Ig	Immunoglobulin
Kb	Kilobase
KCl	Potassium chloride
LiCl	Lithium Chloride
M	Molar
m	meter
MAPK	Mitogen-activated protein kinase
Mb	Megabase
mCi	Milli Curie
Mg	Magnesium
mg	milligram
$MgCl_2$	Magnesium Chloride
min	Minute
miRNA	Micro RNA
mL	Milliliter
mm	millimeters
mM	millimolar
mRNA	Messenger RNA
mRNP	Messenger ribonucleoprotein
MTE	Motif ten element
NaAc	Sodium acetate
NaCl	Sodium Chloride
NaOH	Sodium hydroxide
ncRNA	Non-coding RNA
NFR	Nucleosome-free region
ng	nanogram
NGS	Next-generation sequencing
nm	Nanometers
nt	Nucleotide
nTF	Negative transcription factor
NTP	Nucleoside triphosphate
nTR	Nascent transcription rate

O/N	Overnight
O_2	Oxygen
^oC	Degrees Celsius
OD	Optical density
OH	Radical hydroxyl
ORF	Open Reading Frame
P	Phosphate
PA	Protein amount
PAS	Polyadenylation signal
PCR	Polymerase chain reaction
PIC	Pre-initiation complex
PNK	Polynucleotide Kinase
PRX/Px	Peroxidase
pTF	Positive transcription factor
qPCR	Quantitative PCR
R	Reverse (DNA strand)
RA	mRNA amount
RAM	Random access memory
rcf	Relative centrifuge force
RNA	Ribonucleic acid
RNApol	RNA polymerase
rNTP	Ribonucleotide triphosphate
ROS	Reactive oxygen species
RPCC	RNApol ChIP-on-chip
rpm	revolutions per minute
rRNA	Ribosomal RNA
RT	Reverse transcription
RT	Room temperature
S	Sulphur/ Svedverg (sedimentation coefficient)
s	second
SAPE	Streptavidin R-Phycoerytrin conjugate
SC	Synthetic complete
SDS	Sodium dodecyl sulphate
seq	Sequencing
Ser	Serine
siRNA	Small interfering RNA
snoRNA	Small nucleolar RNA
snRNA	Small nuclear RNA
SOD	Superoxide dismutase
SR	Synthesis rate (mRNA)
srpRNA	Signal recognition particle RNA
SSC	Saline Sodium Citrate
T	Thymine

TAE	Tris Acetate EDTA
t-BOOH	Tert-butyl hydroperoxide
TBP	TATA-binding protein
TCA	Trichloroacetic acid
TE	Tris EDTA
TF	Transcription factor
TLR	Translation rate
TR	Transcription rate
tRNA	Transfer RNA
TRO	Transcription run-on
TRX	Thioredoxin
TSS	Transcription start site
TTS	Transcription termination site
U	Uridine
U	Unit
UAS	Upstream activating sequence
URS	Upstream repressing sequence
UTP	Uridine Triphosphate
UTR	Untranslated region
UV	Ultraviolet
v	volume
YPD	Yeast Peptone Dextrose
YRE	Yap1-responsive element
Zn	Zinc

Introduction

1. Organisms studied in this work

1.1. *Saccharomyces cerevisiae* as a model organism

Yeasts are fungi that grow as single cells, producing daughter cells either by budding (budding yeasts) or by binary fission (fission yeasts). They typically grow in moist environments where there is abundant supply of simple, soluble nutrients such as sugars and amino acids. For this reason they are common on leaf and fruit surfaces, on roots and in various types of food.

The yeast species *Saccharomyces cerevisiae* belongs to the *Ascomycota* phylum, and is it commonly known as baker's or brewer's yeast because it has been utilized for many years in several industrial processes, including the production of bread, beer and wine[1]. Over the years, it has also become a powerful tool from the scientific point of view for several reasons. First, many of the basic cell regulatory mechanisms are evolutionarily conserved from yeast to humans. Second, unlike the complexity and expensiveness that studies in humans imply, yeasts can be easily grown in standard laboratory conditions and proliferate rapidly, needing only approximately two hours to duplicate. Third, the genome of *S. cerevisiae* has not only been completely sequenced[2] but also widely characterized. Out of a total number of 6572 verified open reading frames (ORFs), 5095 (~78%) have been already characterized according to the reference data base for this organism[3]. The facility of culture, the similarity with the molecular biology of higher eukaryotes and the vast knowledge that is currently available for this organism make *S. cerevisiae* an ideal model organism for the study of eukaryotic regulatory networks, as well as for the study of other less known yeast species.

1.2. *Candida albicans: a human fungal pathogen*

Humans are colonized by millions of residential microbes belonging to the three domains of cellular life Bacteria, Archaea and Eukarya, and the domain of non-cellular life Viruses[4]. All these microorganisms together make up the so-called human flora. A large number of these organisms perform many functions that contribute in several ways to the overall health of the human host. However, some of these commensal organisms can cause disease at high numbers. Such microorganisms include bacteroides, coliform bacteria (i.e. *E. coli*), and yeasts. The yeast genus *Candida* is one of the most common residents of the human mucous membranes, such as the gut, the urinary tract and the vagina[5]. Although others are also widely found, *Candida albicans* is the most common type among *Candida* species[6]. In healthy individuals, populations of *C. albicans* are kept in check by competing friendly bacteria and the human immune system. However, immune deficiencies and alterations of the microflora that disrupt the normal balance of microorganisms allow *C. albicans* to increase its numbers and become the majority. The overgrowth of this microorganism leads to a wide variety of infections within the host which are typically divided in two broad categories: mucocutaneous and systemic. Mucocutaneous infections affect vaginal, oral-pharyngeal, esophageal and gastrointestinal mucosae[7]. They are superficial, generally localized, and respond to antifungal treatments in most cases. Systemic infections, on the contrary, involve the spread of *C. albicans* via the bloodstream and the colonization of multiple organs, such as the liver or the heart. The presence of *Candida* in blood is a condition known as "Candidemia", and the invasion and colonization of organs is known as "disseminated candidiasis"[8]. Severely immunocompromised individuals, such as those with neutrophenia, burns, or those under prolonged therapy with broad-spectrum antibiotics are particularly susceptible to these infections[9]. Unlike superficial infections, systemic infections are not only very difficult to eliminate, but also life-threatening. Once acquired, these infections have a mortality rate up to 75 %[10] and a high rate of re-infection[11]. Although infections with non-*albicans Candida* species are increasing in prevalence, *C. albicans* is still responsible for 66% of all cases[6]. One of the main reasons these infections are difficult to

eliminate is that in most cases, treatment is hampered by antifungal resistance. *C. albicans* is capable of developing resistance against many of the currently available drugs using various mechanisms, such as targeting mutations or changes in the expression of genes that code for efflux pumps that are responsible to remove antifungal drugs[12]. Given its prevalence and the health social problem it implies, characterizing the molecular bases of drug resistance, finding new cellular targets and further understanding the overall mechanism of *C. albicans* pathogenesis are important goals to achieve nowadays.

***1.3. C. albicans and S. cerevisiae*: similarities and differences**

S. cerevisiae and *C. albicans* evolved from the same *Sacharomycotina* ancestor and were estimated to diverge from each other about 180 million years ago[13,14] (Figure 1A). During the period of independent evolution, inversions of small segments of DNA were the major cause of gene rearrangement and the significant loss of synteny between both species[15]. Since the two species exhibit different lifestyles, with *S. cerevisiae* being generally a free-living yeast and *C. albicans* being a commensal organism that inhabits the human body, they use different ways of using glucose as a carbon source. Most *Ascomycota* fungi, including *C. albicans*, are respiratory and rely on oxidative phosphorylation processes to produce energy. However, *S. cerevisiae* prefers to ferment glucose, even when oxygen is available. This phenomenon is known to as the "Crabtree effect", and has been proposed to provide a general strategy to increase energy production enabling rapid glucose consumption in yeast, which is a trait that could have been evolutionarily selected to "starve" competitors in nature[16]. Based on this concept, *S. cerevisiae* is defined as a "Crabtree positive" yeast, whereas *C. albicans* is a "Crabtree negative" yeast.

In nature, *S. cerevisiae* reproduces sexually, switching readily between two mating types: haploid 'a' cells that mate with haploid 'α' cells to form diploids. Under nutrient-poor conditions, diploids can be induced to undergo meiosis and sporulation, forming four haploid spores, two of each mating type (Figure 1B). *C. albicans* is capable of undergoing either parasexual or sexual reproduction[17], being its mating types termed *white* and

opaque, corresponding to *S. cerevisiae*'s '*a*' and '*α*' respectively. The possibility of switching between these two reproductive modes has been suggested to allow the pathogen to generate clonal populations in order to thrive in their well-adapted environmental niches via asexual reproduction, while being able to reproduce genetically diverse offspring via sexual reproduction if needed in response to novel environmental pressures[18].

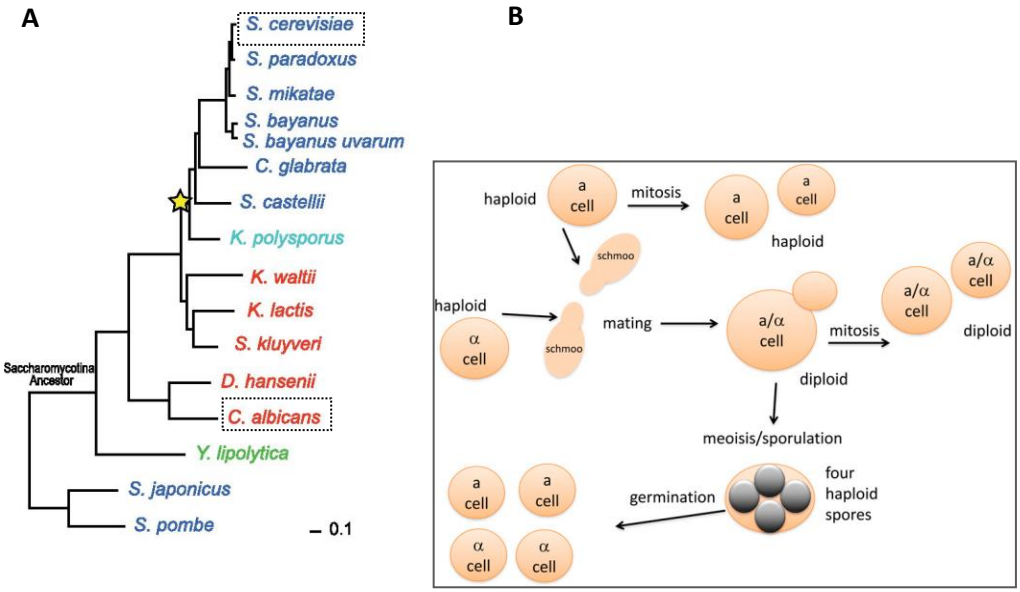


Figure 1. A) Phylogenetic tree indicating the two Ascomycota species in this study. Dark blue: respiro-fermentative; red: respiratory; green: obligate respiratory; light blue: intermediate between respiro-fermentative and respiratory. Star: Whole Genome Duplication event.(Adapted from Thompson *et al.* 2013[19]). **B) Simplified life cycle diagram of laboratory budding yeast.** Haploid yeast cells can be one of two mating types: *MAT*a (a cell) or *MAT*α (α cell). These cells can undergo mitotic cell division through budding, producing daughter cells. The two cell types release pheromones, initiating the formation of shmoos and subsequent mating, ultimately resulting in a stable diploid *MAT*a/*MAT*α (a/α cell). Diploid cells also divide mitotically by budding to produce genetically identical daughter cells. Under nitrogen poor conditions, diploids are induced to undergo meiosis, forming four haploid spores, which can germinate into two *MAT*a cells and two *MAT*α cells. (Taken from Duina *et al.* 2014[20].)

Whereas *S. cerevisiae* lab strains mostly exist in the yeast (blastopore) form, *C. albicans* has the ability to grow in a variety of morphological forms. These range from the yeast form (blastopore, similar to that of *S. cerevisiae*, figure 2A, left panel) to true hyphae with parallel-sided walls[21-23](Figure 2A, right panel). In between these two extremes, the fungus can exhibit a variety of growth forms that are collectively referred to as pseudohyphae

(Figure 2A, middle panel). This form is also found in wild *S. cerevisiae* strains in poor nitrogen media[24]. In these pseudohyphal forms, the daughter bud elongates and the daughter cell remains attached to the mother cell after septum formation. As a result, filaments composed of elongated cells with constrictions at the septa are formed. The ability to switch between yeast, hyphal and pseudohyphal morphologies is considered to be necessary for virulence, since strains that lack this ability are rendered avirulent[25]. The ability to form hyphae has been proposed to promote tissue penetration during the early stages of infection, whereas the yeast form is more suitable for dissemination in the bloodstream[26] (Figure 2B).

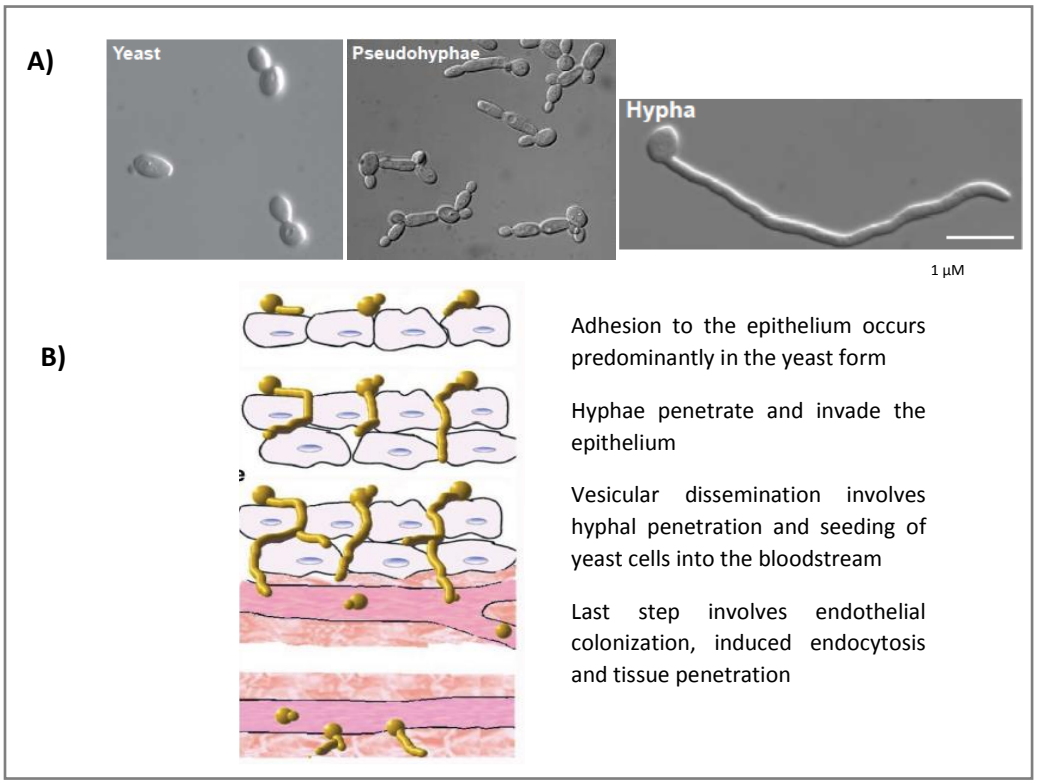


Figure 2. A) The three different morphological states of *C. albicans*. *C. albicans* can grow in a variety of morphological forms, ranging from the yeast form (blastopore, left) to true hyphae with parallel-sided walls (right). In between these two extremes, the fungus exhibits intermediate growth forms known as pseudohyphae (middle). (Adapted from Sudbery *et al* 2004[27]). **B) *C. albicans* and its morphological transitions during tissue invasion.** The ability to switch between morphological states is considered necessary to promote virulence and facilitates tissue penetration during early stages of infection. (Adapted from Gow *et al* 2012[26]).

S. cerevisiae lab strains mostly exists in the haploid form, and its genetic information is distributed in 16 distinct chromosomes with a total number of 13.4 megabases (Mb). *C. albicans* on the contrary is considered to be an obligate diploid, and has a nuclear genome of about 16 Mb, approximately 33% larger than that of *S. cerevisiae*. Its genome is distributed in 8 pairs of homologous chromosomes which are numbered from 1 (largest) to 7 (smallest), with the one carrying ribosomal DNA termed R[28]. An initial striking observation of the *C. albicans* genome was that it contained a large proportion of ORF products that were non-homologous to known proteins compared to *S. cerevisiae*[29]. Therefore, the genome of *C. albicans* seems to contain a significantly high proportion of genes specific to this organism, which is related to its pathogenic nature and versatile way of living. In concordance with this, a relatively larger number of proteins is assigned to functional categories related to interactions with the environment and morphogenesis in *C. albicans* than in *S. cerevisiae*, compared to proteins involved in metabolism, proliferation or subcellular compartmentalization[29]. The larger size of the *C. albicans* genome is also partly explained by the greater number of retrotransposon families, as well as by the presence of families of proteases or lipases that are needed for virulence which are lacking in *S. cerevisiae*.

Despite the clinical relevance of this pathogen, 73% of its genome still remains uncharacterized, compared to almost the same percentage of ORFs that are characterized in *S. cerevisiae* (Figure 3), and although both species are evolutionarily divergent, many *C. albicans* proteins have orthologues in *S. cerevisiae*. For this reason, many researchers nowadays use *S. cerevisiae* as a yeast model organism, against which *C. albicans* is compared[30-32].

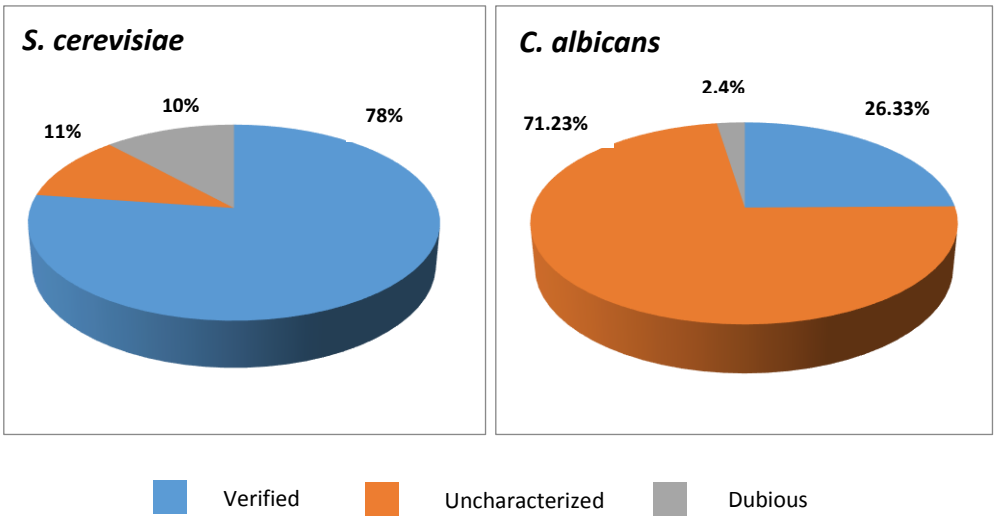


Figure 3. Graphical view of protein coding genes in the yeast species utilized in this study. The percentages of ORFs corresponding to each category are shown in different colors (bottom legend). In the haploid state, a total number of 6604 and 6198 ORFs are represented in *S. cerevisiae* and *C. albicans* respectively. These data were retrieved from the correspondent databases of each species, *Saccharomyces* Genome Database (SGD) and *Candida* Genome Database (CGD) left and right respectively. Taken from genome inventory 25 January 2016.

2. Transcription in Eukaryotes

Transcription is the first step of gene expression, in which a particular DNA segment is copied into an RNA molecule by the enzyme RNA polymerase (RNApol). During transcription, RNApol slides along the DNA region reading its sequence and producing a complementary RNA strand, known as the "primary transcript". The stretch of DNA transcribed into an RNA molecule is therefore defined as a "transcription unit" and encodes at least one gene. If the gene transcribed is a protein-coding gene, the produced transcript is then referred to as messenger RNA (mRNA), as it contains the information needed for the production of a protein that is achieved through the process of translation. Alternatively, the transcribed gene may encode for non-coding RNAs, which are RNA molecules that perform certain functions within the cell despite not being translated into a protein, such as micro RNAs (miRNA), ribosomal RNA (rRNA), or transfer RNA (tRNA).

In eukaryotes, the process of transcription is performed in the nucleus by three different RNA polymerases (RNApol I, II and III), each of which transcribe a different set of RNA

molecules. RNApol I performs its function in the nucleolus transcribing large ribosomal RNA molecules (rRNA), namely 25S, 18S and 5.8S rRNA. RNApol II transcribes mRNAs, most small nuclear RNAs (snRNAs), small interfering RNA (siRNAs) and miRNA. Finally, RNApol III transcribes transfer RNA (tRNA), other small RNAs (including the small 5S rRNA), snRNA U6, the RNA contained in the signal recognition particle (SRP RNA) and other stable short RNAs. According to the relative abundance of the genes they transcribe, the independent contributions of RNA pol I, II and III to the overall level of transcription within the cell are approximately 50-70%, 20-40% and 10% respectively.

All three RNA polymerases are structurally similar, sharing the presence of subunits that are common to all of them, while containing type-specific enzymatic subunits that are unique of each polymerase [33].

2.1. Transcription by the RNApol II: elements involved

During the RNApol II transcription cycle, several elements come into play. The presence and structural organization of these elements will regulate how transcription is achieved, and determine the outcome of the transcriptional process.

2.1.1. DNA and *cis* elements: the transcription template

A typical eukaryotic RNA pol II gene contains several sequence elements that define the particular DNA coordinates where transcription will occur. The coding region (that is, the length of DNA that will be read by the RNApol to be copied into an RNA molecule) is preceded by a generally upstream sequence element called the “promoter”. This region serves as a reference site of transcription for the RNApol, which recognizes this element and binds to it to prepare for transcription. The promoter is generally located around 200 bp upstream of the transcription start site (TSS, denoted as *+1*), and contains several dispersed sequence elements that serve as scaffold for the recognition and binding of other proteins needed to assist the RNApol II. In particular, the minimum set of these sequence

elements required to yield a basal level of transcription is known as the "core promoter" [34], and can contain 5 main sequence elements: The *TATA box*, the *Initiator* element (Inr), the *Downstream promoter element* (DPE), the *Motif-ten element* (MTE) and the *TFIIB recognition element* (BRE) (Figure 4C).

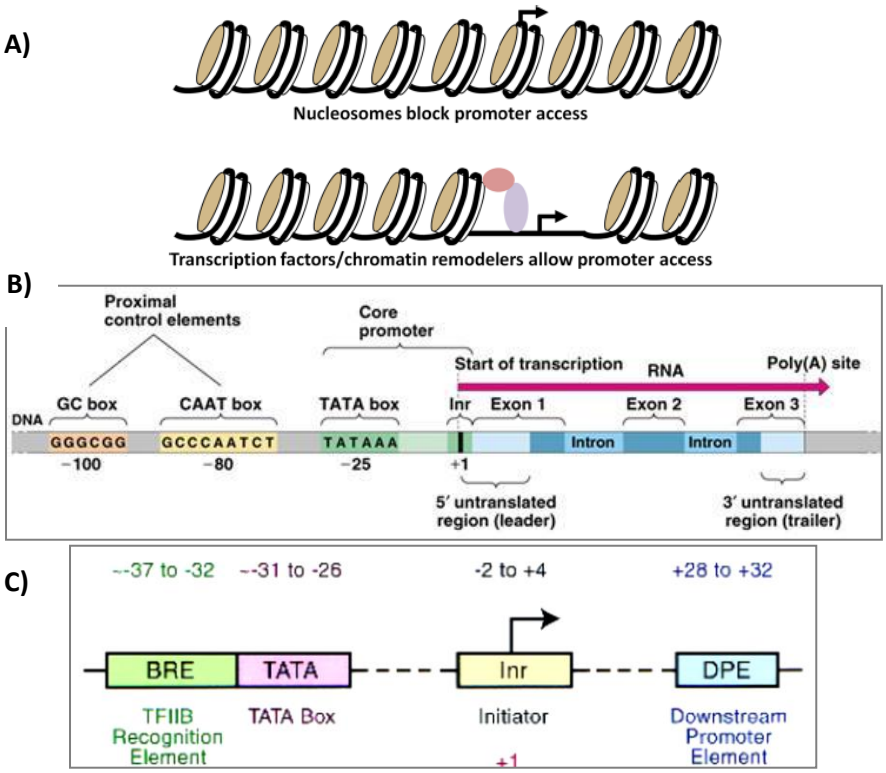


Figure 4. **DNA elements in a typical eukaryotic RNA pol II gene**. **A) Schematic view of chromatin structure.** DNA (black string) is wrapped around individual sets of histones called nucleosomes (brown circles). Nucleosome occupancy inhibits promoter sequence recognition. The occupancy of transcription factors (in purple) and chromatin remodelers (in pink) correlates with the nucleosome free region (NFR) leaving the promoter DNA region exposed. (Taken from Tsukiyama 2002[35]). **B) Typical sequence elements present in a canonical RNApol II promoter region.** The presence and disposition of the particular elements within the proximal promoter is species and gene specific. **C) The RNA polymerase II core promoter**. Amplified view of the core promoter region (Adapted from Butler and Kadonaga 2002[36]).

In addition, eukaryotic genes can contain extra regulatory sequence elements called the "proximal promoter" and "enhancer" (UAS: upstream activating sequence in yeast) or "silencer" (URS: upstream repressing sequence in yeast) regions that are located further upstream of the core promoter (Figure 4B). Although the particular disposition of all the

above mentioned elements along the gene sequence is common to most genes, their presence or absence is gene-specific, and determines the level of regulation each gene is subjected to: the higher the number of regulatory elements, the higher the level of regulation of the gene. For instance, in spite of being supposedly a main component of the core promoter, 80% of yeast genes are known to be *TATA-like*[37], meaning they lack the canonical TATA box element and contain a similar structure instead[38]. One such example is house-keeping genes, which, as they do not need to be subjected to the same level of regulation as other inducible genes, lack both the TATA box element and the DPE element. Once the RNApol recognizes the promoter it will then slide along the DNA template proofreading it and producing the specific complementary RNA strand. The process will finish when the RNApol reaches a sequence element called the poly (A) site, where a poly (A) tail of variable length will be added to the newly synthesized RNA molecule (Figure 4B). Finally there is an additional regulatory factor that determines the availability of the DNA template to be transcribed. In the eukaryotic context, DNA is not as easily accessible to the RNApol as they are in prokaryotes. Instead, eukaryotic genomes are highly compacted in a complex structural conformation denoted as "chromatin", where DNA is wrapped around individual and regularly spaced sets of 8 proteins called histones (Figure 4A), which is known as a histone octamer, or nucleosome. Despite initially believed to be regularly positioned all along the DNA in an unspecific manner, it has now been defined that nucleosomes occupy specific preferred positions along the genome [39]. High-resolution genome-wide studies have revealed a common disposition pattern: nucleosomes are not only absent in most enhancer, promoter, and transcription termination regions, but also occupy different positions depending on whether the DNA region is intergenic or gene-containing [40-42]. In particular, both immediately before the TSS and immediately after the polyadenylation site there are regions that 150-200 bp in length which are known to be nucleosome-free, hence denoted as "Nucleosome free regions" (NFRs) (Figure 5). In yeast, nucleosomes flaking the TSS upstream and downstream (known as nucleosome -1 and nucleosome +1 respectively) are known to be strongly positioned, whereas the position of the subsequent nucleosomes located further downstream of the gene body becomes gradually more diffused [43,44].

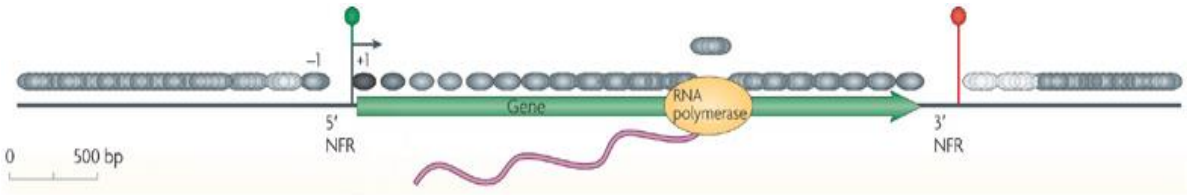


Figure 5. Nucleosomal landscape of yeast genes. The consensus distribution of nucleosomes around all RNA pol II yeast genes is shown in grey ovals, aligned by the beginning and end of every gene. The arrow under the green circle near the 5' nucleosome-free region (NFR) represents the TSS, and the red circle indicates transcriptional termination within the 3' NFR. Adapted from Jiang and Pugh (2009)[45].

These nucleosomes the DNA is coiled around have been reported to interfere with many DNA transactions[46,47], thus serving as general gene repressors. This assures the inactivity of all the many thousands of genes in eukaryotic cells except those which transcription is brought about by specific positive regulatory mechanisms. Upon gene activation, chromatin of the promoter is transformed from a static to a dynamic state[48] where nucleosomes are rapidly removed and promoter DNA is made transiently available for interaction with the transcription machinery.

2.1.2. The RNApol II transcription machinery

Due to the loss of nucleosomes from transcriptionally active promoters, the RNA pol II transcription pre-initiation complex (PIC) is assembled on the naked DNA molecule.

RNApol II is a highly conserved multiprotein complex composed of 12 subunits denoted Rpb1 to 12[49], with Rpb1 and 2 being the largest of all of them as well as the most conserved across species (Figure 6). The association of these two subunits forms the catalytic core of the enzyme, which contains an Mg^{2+} ion as a cofactor that is essential for its enzymatic activity. It is within this catalytic core where the RNA strand is synthesized. Except for Rpb4 and Rpb9, all other ten subunits are essential[50,51] and constitute the holoenzyme. Attached to the holoenzyme, Rpb4 and 7 together form a subcomplex that can dissociate from the catalytic core[49,52-54].

A feature that is unique to the RNA pol II enzyme is the presence of several tandem repeats of the heptapeptide YSPTSPS in the C-terminal domain (CTD) of its largest subunit, Rpb1. This amino acid sequence is repeated 26 times in *S. cerevisiae* and 52 times in humans, and the number of repetitions is directly correlated with the complexity of the genome. The CTD does not play a structural role however it is essential to the enzyme, as it acts as a dynamic scaffold for the interaction of the RNApol II with other proteins that are needed for both the transcription and the processing of the newly synthesized mRNAs. The CTD is strategically located next to the channel through which the RNA strand exits the enzymatic core[55], and it undergoes several post-translational modifications all along the transcriptional cycle, such as glycosylation and isomerization of prolines and phosphorylation of the serines located in positions 2, 5 and 7 (Ser2, Ser5 and Ser7 respectively)[55]. All these modifications take place sequentially as the RNApol II slides along the DNA to produce the new RNA strand, in which the particular pattern of modifications of the CTD is a reflection of the transcriptional state of the enzyme, as well as the position it occupies along the gene body. For instance, when RNA pol II is recruited to the promoter to form the PIC for the first time to initiate transcription the CTD is hypophosphorylated, whereas RNApol in active elongation is sequentially phosphorylated at Ser5 and Ser2 as the enzyme slides along the template.

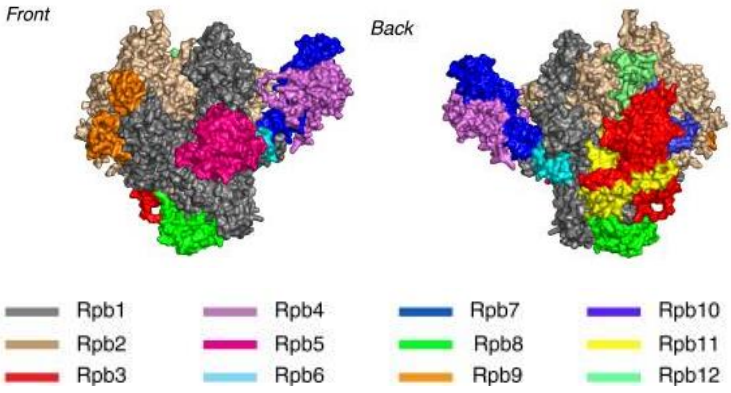


Figure 6. Structure of *S. cerevisiae* RNApol II. Surface representation of front (*Left*) and back (*right*) views are shown. Individual subunits are colored as indicated below (Adapted from Spahr *et al* 2009[56]).

RNA pol II is capable of unwinding DNA, synthesizing RNA, and rewinding DNA. However, RNApol II alone is incapable of recognizing the promoters and initiating transcription, where for these functions, the participation of other factors is required. The 12-subunit polymerase is assisted all throughout the transcriptional process by a set of 8 **general transcription factors** (GTFs), denoted TFIIA,B, D ,E ,F ,H ,J and S that are responsible for core promoter recognition and for unwinding the promoter DNA[57]. These GTFs are also accompanied by a 20-subunit protein complex called the Mediator, which transduces regulatory information from activator and repressor proteins to the RNApol II[58-60](Figure 7). The **transcription factors** (TFs), are the elements that connect other cellular processes to the process of transcription itself. Activated via other events such as their own phosphorylation, proteolysis, interaction with other proteins or location switch from the cytoplasm to the nucleus, TFs are capable of activating or repressing transcription by binding to *cis* elements placed at the proximal promoter. Those that are known to activate transcription are known as positive TFs (pTFs), and those that repress it are known as negative TFs (nTFs). All these TFs can either be more or less gene-specific, and can act either close to or far away from the TSS[61].

The **Mediator** is a 20-subunit protein complex that interacts directly with both activator proteins and RNApol II, transducing regulatory information between them[62]. It forms a tight

complex with an activator and an enhancer, and it subsequently binds RNApol II and the GTFs at the promoter to stimulate the initiation of transcription[63]. Mediator is not only the basis for regulated transcription, but it is required for transcription on almost all RNApol II promoters, being for this reason no less essential for transcription than RNApol II itself[64,65]. In addition, the Mediator complex is important not only for positive but also for negative regulation of transcription. Although it is commonly referred to as a coactivator, the mediator also acts as a corepressor and a GTF.

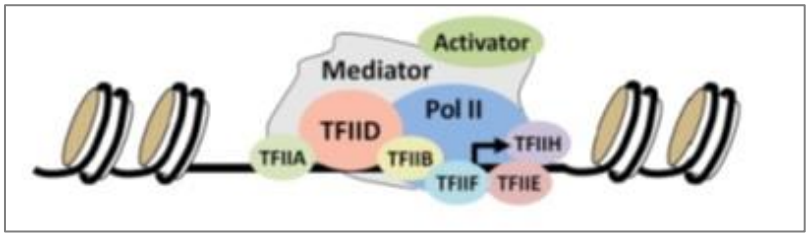


Figure 7. The RNA pol II transcription machinery. Once transcription factors and chromatin remodelers allow access to the promoter region, general transcription factors (GTFs), the RNA pol II and the Mediator form the pre-initiation complex (PIC) on the exposed region. (Adapted from Guo and Price 2013[66]).

2.2. The transcription cycle

The transcriptional process is generally divided into 3 main steps: 1) Initiation, where the RNApol binds to the promoter and recruits the transcription machinery that is needed to start transcription; 2) Elongation, where RNApol slides along the DNA synthesizing RNA; and 3) Termination, when RNA synthesis is completed. Upon completion, RNApol is released from the template and will either start a new transcription cycle or is degraded[67].

2.2.1. Initiation

As mentioned before, RNA pol II is unable to bind the core promoter alone, and requires assistance by a particular plethora of transcription initiation factors (GTFs) to do so. For this reason, transcription only starts when RNA pol binds DNA upon successful assembly of the PIC [68]. *In vitro* biochemical reconstitution studies have revealed the order of assembly of these factors [69].The first factor to locate and bind the promoter is TFIID, one of which

subunits is the TATA-box binding protein (TBP) that specifically recognizes and binds the TATA-box sequence element present in the core promoter. The assembly of the TBP onto the promoter region can also occur via a Spt-Ada-Gcn5-Acetyltransferase complex termed SAGA. TFIID and SAGA are compositionally and functionally related complexes[70] and in principle, a given promoter can utilize either one or the other[71-73] . However, the SAGA pathway seems to be tailored towards TATA-containing promoters, whereas the TFIID pathway plays a greater role at TATA-less promoters[37]. TFIIA and B are recruited next, and help stabilize the interactions of TBP with the assembling PIC complex. The bound TFIIB then recruits RNApol II and TFIIF. TFIIE binds next the complex and recruits TFIIH, whose ATPase and helicase activities unwind the DNA duplex around the TSS to form the transcription bubble[74]. The binding of TFIIH completes the formation of the PIC. Once the transcription bubble is formed, the single DNA strand is inserted into the active site of the RNApol II, and the polymerization of the new RNA strand begins.

The initial DNA/RNA hybrid is often not long enough to form a stable elongation complex, therefore results in several rounds of synthesis of short RNAs that all together are commonly referred to as the "abortive initiation phase"[75]. When the RNA strand reaches a minimum length of 6 nucleotides, initiation factors are released, RNApol II loses contact with the promoter, and a stable elongation complex escapes the promoter region to start advancing along the template[68,76].

Once RNApol II is successfully engaged in elongation, some GTFs, such as TFIID, A and B as well as the Mediator complex remain in the promoter[77]. This allows the rapid reassembly of new RNApol II molecules to initiate the subsequent rounds of transcription on the sample template. However many other factors, such as chromatin remodelers needed to remove nucleosomes, travel with the RNApol II along the gene body in order to continue assisting the enzyme during elongation.

2.2.2. Elongation

During elongation, the polymerase uses NTP as an energy source by incorporating it into the growing RNA strand to move along DNA, advancing one bp at a time[78]. The elongation

process does not occur in an entirely processive way. Instead, if DNA damage is found, or when incorrect incorporation of a given NTP occurs, RNApol II can pause and backtrack to correct it. This phenomenon of backward movement of the polymerase on DNA followed by the return of the enzyme to the original position is termed “backtracking”, and it is the focus of many transcriptional studies that are now suggesting this event to be an additional regulatory mechanism for the transcription of many genes[79]. Despite the importance of this phenomenon, not all pauses are necessarily associated with backtracking. Some of them are speculated to be a result of molecular rearrangements of the elongation complex that render the elongation incompetent. These pauses are known as “ubiquitious” or “elemental” to distinguish them from backtracking events.

As RNApol II backtracks, the entire elongation bubble shifts, and the 3’ end of RNA loses its register with the active center of the polymerase and inhibits NTP incorporation. When this occurs, the polymerase randomly slides back and forth on the DNA until the 3’ end of the transcript realigns with the active center to allow RNApol to resume elongation. Many transcription factors assist the polymerase, such as TFIIS which enhances transcriptional fidelity by stimulating the cleavage of misincorporated nucleotides, assisting RNApol II to recover faster from backtracking[80-82], and the transcription factor Spt4/5, that stabilizes the elongation complex increasing its processivity[83].

Overall, when these pauses are not considered for calculations, the resulting mean elongation velocity of RNApol II at saturated concentrations of NTPs is 15-23 bp/s[84]. This velocity can vary slightly depending on the DNA template[85], as well as on other factors that may aid or oppose the enzyme, such as changes in cell growth temperature[86].

As mentioned before, the DNA template in eukaryotes is wrapped in nucleosomes, which act as mechanical barriers to the advancing polymerase. For this reason, the pause durations and hence the densities of RNApol II molecules increase upon encountering a nucleosome[84]. The advancement of the enzyme through the barrier depends on the fine interplay between the enzyme dynamics (including pausing and translocation) and the fast dynamics of nucleosome fluctuations that are regulated by specialized machineries that modify or remodel nucleosomes. Such remodelers include the elongation complex[87],

Gcn5[88,89] and Set2[90]. It is still unclear, however, whether nucleosome modifications directly affect polymerase elongation or indirectly regulate transcription through recruitment of other factors. While these factors help remove nucleosomes, other factors facilitate nucleosome reassembly to restore the chromatin structure that is left behind by the advancing polymerase. Some of these factors include the FACT complex[91] and Spt6[92]. By facilitating nucleosome restoration, they also prevent the cell from producing cryptic transcripts, which are difficult to detect due to their small amount and lack of correspondence with predicted canonical genes.

2.2.3. Termination

Termination is the last step of transcription that occurs when RNApol II ceases RNA synthesis and both the enzyme and the nascent RNA are released from the DNA template. This process, rigorously defined as the disruption of the DNA-RNApol II-RNA ternary complex, does not necessarily occur at the end of the gene body, but can also happen either upstream or within the ORF. In fact, for most eukaryotic genes termination often occurs close to the TSS by RNApol molecules that are transiently paused next to the promoter, and this phenomenon termed as “premature termination”.

The particular termination site will vary depending on the type of transcript RNApol II is transcribing, as such that when transcribing protein-coding genes to produce the correspondent mRNA, RNApol II termination generally occurs downstream of the specific DNA sequences that dictate when the nascent RNA is processed for completion of transcription. These sequences are termed “polyadenylation signals” (PAS) or “poly(A) sites”, and are generally conserved across most eukaryotic species. In particular, cleavage and polyadenylation sites are known to occur 10 to 30 nt downstream from a conserved hexanucleotide, AAUAAA, and 30 nt upstream of a less conserved U- or GU-rich region. In budding yeast, however, these *cis* elements display considerable variation[93]. In addition, recent studies have revealed that many transcripts contain several alternative cleavage and polyadenylation sites in yeast, which generates different transcript isoforms with both different coding potential and 3’UTRs[94,95]. In all cases, RNApol II termination is facilitated

when *cis*-acting elements in the nascent transcript are recognized by RNA-binding proteins that co-transcriptionally associate with the CTD. For this reason, the type of CTD phosphorylation influences the mechanism of termination.

RNApol II termination mechanisms are elicited through two main different pathways: the poly(A)-dependent and the Sen1-dependent termination pathways. Based on both the distance that the enzyme has progressed from the TSS and the associated change in CTD phosphorylation status, yeast RNA pol II selectively uses one or the other termination pathway[96]. Generally speaking, poly(A)-dependent termination predominates at distances >1kb from a TSS (typical for mRNA-encoding genes), whereas Sen1-dependent termination predominates at distances <1kb from a TSS (typical for genes encoding snRNAs and snoRNAs, and also for cryptic transcripts)[96-98].

2.2.3.1. Termination in protein-coding genes: Poly(A)-dependent termination

RNApol II termination occurring downstream of most protein-coding genes is functionally coupled with an mRNA maturation event in which the 3' end of the nascent transcript undergoes cleavage and polyadenylation[99,100]. In this sense, the 3' end processing can be broken down into two steps: first, the poly(A) site is transcribed, reducing the rate of RNApol II elongation and causing its pausing downstream of the poly(A) site[101-103], which is followed by endoribonucleolytic cleavage of the nascent transcript; and second, the upstream cleavage product is polyadenylated, whereas the downstream cleavage product is degraded. The 3'end processing reaction will initiate when *cis*-acting elements in the poly(A) site of the nascent transcript are recognized by RNA-binding factors that associate with RNApol II. This 3' end processing machinery is highly complex, comprising more than 20 polypeptides in yeast[104] , including the *cleavage polyadenylation factor* (CPF), that is recruited to the elongation complex by its interaction with the RNApol II body, and the *cleavage factor IA* (CFIA) that interacts with the enzyme via the CTD (Figure 8A). When the AAUAAA sequence is recognized by the CPF, this factor induces pausing of the enzyme. Additionally, when CFIA binds to the downstream GU-rich processing signal, CPF binds to CFIA, releases its hold on the enzyme body and accompanies CFIA to the CTD, which leads

to CPF-mediated cleavage and release of the paused RNApol II. Efficient release of RNApol II downstream of the poly(A) site requires not only these two factors but also the 5'-3' exoribonuclease RNA trafficking protein 1 (Rat1), which, together with its activating cofactor Rat1-interacting 1 (Rai1) and its CTD-interacting partner regulator of Ty1 transposition 103 (Rtt103), acts as a 'torpedo' to promote polymerase eviction from the DNA template[105] [106,107].

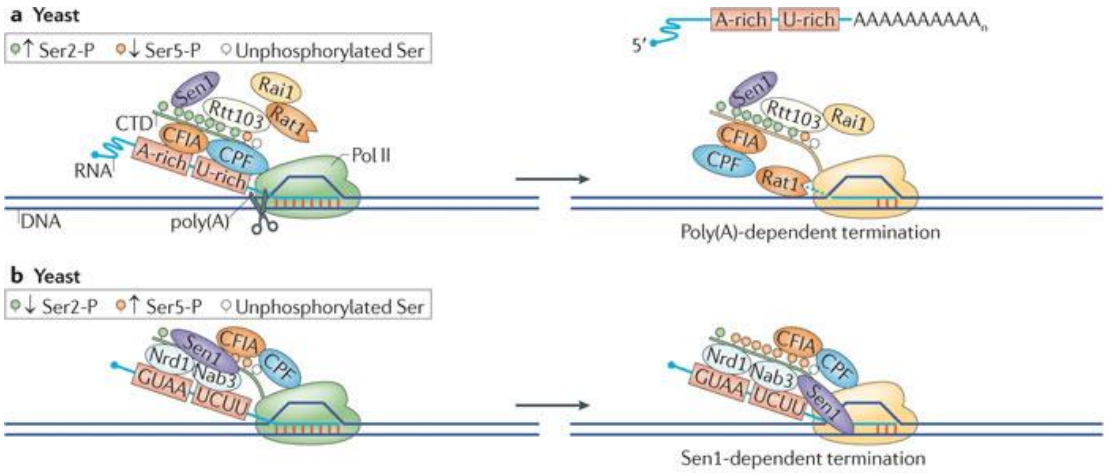


Figure 8. Factors involved in poly (A)-dependent and Sen1-dependent termination. Termination factors in yeast are shown on the DNA template, being the known interactions between RNA, RNA pol II and other factors indicated by direct contacts. RNA pol II CTD phosphorylation dynamics are displayed in green and orange dots as described in the legends above, with Ser2-P being higher than Ser5-P in regions of poly(A)-dependent termination, and the reverse pattern being observed in regions of Sen1-dependent termination. (Adapted from Kuehner *et al.* 2011[108])

2.2.3.2. Termination in non-coding RNAs: Sen1- dependent termination

An alternative RNApol II termination pathway occurring in most non-coding RNAs is the Sen1-dependent pathway. Unlike mRNAs, the 3' ends of yeast snRNAs and snoRNAs are generated by endoribonucleolytic cleavage and/or exoribonucleolytic trimming by the nuclear *exosome-TRAMP* complex, and do not have a poly(A) tail in their mature form[109]. A distinct set of core factors is required for recognition and transduction of the transcription termination signal, including the RNA-binding protein Nrd1, the nuclear polyadenylated

RNA-binding 3 (Nab3) and the putative RNA and DNA helicase Sen1[97,110-112], which alone is also needed for 3'-end processing of other non-coding RNAs (ncRNAs)[113,114]. In addition to targeting snRNAs and snoRNAs, the Sen1-dependent pathway is required for the termination of different types of ncRNAs, including intergenic and antisense (AS) cryptic transcripts such as those termed *cryptic unstable transcripts* (CUTs)[115,116]. Similarly to the poly(A)-dependent mechanism, the interaction of these factors with the CTD and with their RNA-binding sites could provoke conformational changes in the transcription complex. However, instead of a collision with Rat1 dislodging the enzyme, Sen1 is proposed to promote termination by unwinding the DNA-RNA hybrid in the active site[111]. In addition, Sen1 alone is also known to contribute to the poly(A)-dependent termination pathway. Although its role is yet unknown, it has been speculated to make RNA more accessible to Rat1 by removing proteins or secondary RNA structure[117]. (Figure 8B)

2.3. Coupling transcription and mRNA processing

The ultimate goal of transcription is the production of a functional messenger ribonucleoprotein (mRNP), which involves the packaging of a mature mRNA molecule with proteins to make a particle competent for nuclear export and translation in the cytoplasm. The maturation of mRNA precursors occurs simultaneously with their synthesis by RNApol II, and involves attachment of a 7-methylguanosine cap to the 5' end, intron excision, and formation of the 3' end by cleavage and addition of the poly (A) tail as previously discussed. This processing does not only occur simultaneously with transcription but is also mechanistically coupled to it, which means that synthesis and processing are interdependent. As mentioned before, the CTD of RNApol II Rpb1 is dynamically modified during transcription, and in particular it undergoes reversible phosphorylation events at multiple positions (reviewed in [118]). The timing of these modifications will regulate RNA processing events by recruiting the factors needed, that will be recognized at each transcriptional stage by the specific modification patterns in the CTD.

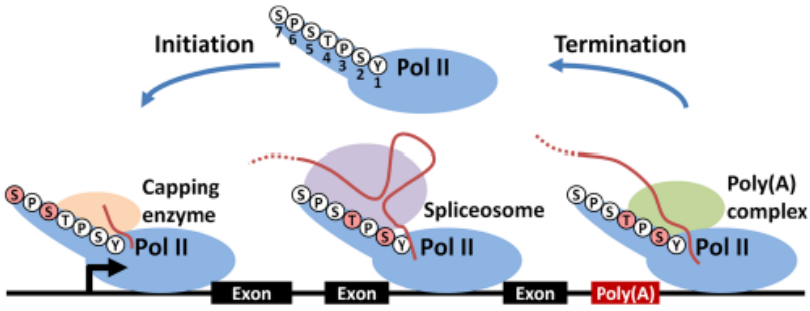


Figure 9. **Coupling of RNA processing with transcription and the influence of CTD phosporylation.** RNA pol II at different stages of elongation across an average gene. Seven amino acid repeat in the CTD is illustrated so that white circles are unphosphorylated and red circles are phosphorylated. Changes in the phosphorylation of the CTD and their influence on the indicated machineries are shown. (Taken from Guo and Price 2013[66])

2.3.1. mRNA capping

The first step in mRNA processing is the attachment of a 7-methylguanosine cap to the 5´ end, commonly known as 5' capping. This mRNA modification takes place during the initial stages of elongation, when the nascent transcript is 20 to 30 nt long[119,120]. The process requires 3 different enzymatic activities (phosphatase, guanylyltransferase and methyltransferase) that are catalyzed by separate proteins in yeast. First, the capping enzyme hydrolyzes the 5' triphosphate to a diphosphate and adds GMP to the first nucleotide via 5'-5' triphosphate linkage. Then, a methyltransferase methylates the N7 position of the transferred GMP generating the m^7G cap. This cap protects mRNA from degradation by 5'-3' exonucleases and is required for efficient initiation of translation once the mRNA is in the cytoplasm[121]. The Ser5-phosphorylated CTD has a crucial role in co-transcriptional capping by directly interacting with both the methyltransferase and the mRNA-capping enzyme, which are also activated by the CTD[122-124]. In addition to co-transcriptional capping, a cytoplasmic capping activity has been reported[125,126] suggesting the possibility of an additional post-transcriptional capping mechanism. The presence of the m^7G cap on is recognized by the *nuclear cap binding complex* (CBC) and promotes co-

transcriptional processing of the nascent mRNA by splicing and polyadenylation machineries[127].

2.3.2. mRNA splicing

Introns are DNA sequences within a gene that are removed while the final mature mRNA product of a gene is being generated, whereas exons are the sequences that remain for further translation. In this sense, the process by which introns are removed and exons are then joined together is known as "mRNA splicing".

Most mammalian genes contain multiple introns, as such that splicing is critically regulated to determine the coding sequence in the mRNA. Majority of yeast genes, however, do not contain introns. Only about 280 genes (5% of all yeast genes) are intron-containing genes in *S. cerevisiae* and alternative splicing has been reported for very few cases[128]. Introns are removed by a complex called the *spliceosome*, which is composed of 5 snRPs (U1, U2, U4, U5, and U6) and associated proteins exceeding a total number of 100 factors involved[129]. Just like mRNA capping, splicing occurs co-transcriptionally in yeast[130-132]. This happens via a mechanism termed *Terminal exon pausing*, in which RNApol II pauses at the end of intronic regions for the time that is needed for the recruitment of the spliceosome and its assembly on the nascent RNA chain[133]. This was further confirmed by a recent study that showed that RNApol II elongation rate is inversely correlated with splicing efficiency[134].

In the case of the *C. albicans* genome, there are 215 ORFs containing at least one intron. Four of these have two introns, one gene (encoding the Hxt4p transporter) has three, and the *SIN3* gene has four. A total of 43 (20.2%) of these genes encode ribosomal proteins, 63 (29.6%) encode products with enzymatic activity, and 26 (12.2%) encode trans-membrane proteins involved in small molecule transport[135]. A significant proportion of these introns are located in the 5' end of ORFs, with 32% of introns being located within the first 10% of the coding sequences. Seventy out of 215 intron-containing ORFs have reciprocal best matches with *S. cerevisiae* genes that also contain introns[135]. Among these 70 ORFs, 25 introns (35.7%) share the same position and the same phase. This suggests that these

commonly positioned introns descended from a common ancestor, as it had previously been suggested by other authors[136].

2.4. Non- coding RNAs and antisense transcription

Transcription is known to be a pervasive phenomenon: a very large amount of the DNA contained within a cell is transcribed into RNA molecules. As mentioned before, these molecules can be either coding or non-coding RNA (ncRNA), meaning they will or will not be translated into a protein. Non-coding RNAs are frequent and widespread among all kingdoms of life[137], and are known to be involved in a large array of gene expression regulatory processes in both prokaryotic and eukaryotic cells, both at the transcriptional and the post-transcriptional level. According to their length, these ncRNAs are classified as small ncRNAs (sncRNAs), being less than 200 nt long (i.e. tRNAs, microRNAs, small interfering RNA), or long non-coding RNAs, being more than 200 nt long (lncRNAs; i.e. rRNA, intergenic, antisense, and intronic). Similar to most mRNAs transcribed by RNApol II, lncRNAs are both capped and polyadenylated[138-140]. While sncRNAs are already accepted to be fundamental players in gene regulation [141], the function of the vast majority of lncRNAs is just now beginning to be understood.

Antisense transcripts (AS transcripts) are a class of lncRNA molecules that are transcribed from the strand opposite to that of the sense transcript of a given gene, encompassing both protein-coding and non-coding genes. Within this context, the originally annotated transcript is referred to as the “sense transcript”, and the one identified on the opposite strand is referred to as the “antisense transcript”. Although AS transcripts are known to significantly vary in abundance, length, and location relative to the gene body, they all seem to share the feature of being expressed at very low levels in respect to their corresponding sense transcript. For example in humans, AS transcripts have been reported to be, on average, more than 10-fold lower in abundance than sense expression [142].

Antisense transcription has been identified in many organisms, including bacteria, protozoa, plants, invertebrates and mammals [137].Several species of yeast have recently been reported

to produce antisense transcripts, including *Schizosaccharomyces pombe* [143-145], *S. cerevisiae* [146,147], *S. paradoxus*, *S. mikatae*, *S. bayanus*, *S. castellii*, and *Kluyveromyces lactis* [148]. Some of the AS transcripts described in these studies have demonstrated to have a function, suggesting that at least part of the total antisense transcriptional events are regulated and functionally significant. However, most of AS transcripts are cryptic (without function) and are produced either by fail in surveillance mechanisms[149] or by incorrect TSS selection[150]. The functionality of AS transcripts in relation to stress responses is a matter of debate (see reference 335 for discussion). Despite being generally much less studied, widespread occurrence of antisense transcription has also been reported in *C. albicans* [151], and some studies have recently suggested a regulatory role for some of the detected AS transcripts [152]. However, the underlying mechanisms that regulate the appearance of antisense transcription in this organism still remain elusive.

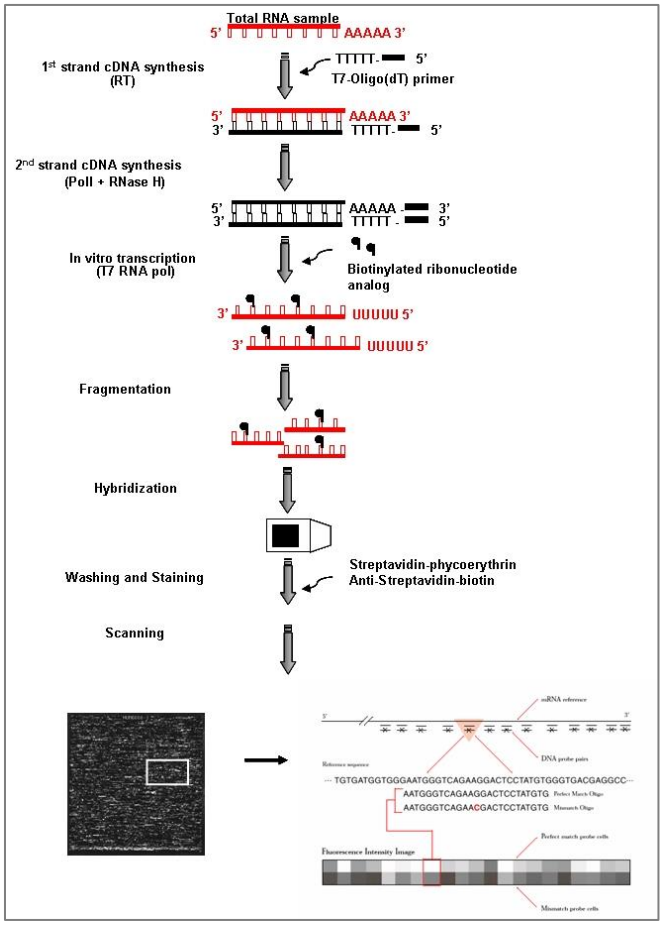
3. Next-generation technologies for transcriptomic studies: Microarrays and RNAseq

3.1. DNA tiling microarray technology

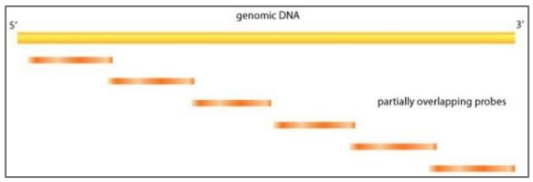
A DNA microarray (also known as DNA chip) is a collection of microscopic DNA spots attached to a solid surface. Each DNA spot contains picomoles of a specific DNA sequence known as "probe" or "oligo", that is usually a short fragment of a gene that is hybridized to a DNA or cRNA sample (target). This target sample is generally either a fluorophore or chemiluminescence-labeled, as such probe-target hybridization can be detected by the appearance of a signal in the correspondent probe spot (Figure 10). The intensity of this signal will determine the relative abundance of this nucleic acid sequence in the target sample.

Since an array can contain tens of thousands of probes, a microarray experiment can address relative nucleic acid amount for thousands of genes in parallel. There are different types of microarrays depending on the nature and coverage of the probes. Some only contain probes for ORFs, restricting the information obtained to previously described genomic features. One of the most commonly used type of DNA microarrays is called "tiling array", which contains a set of overlapping oligonucleotide probes that represent a subset or all of the genome at very high resolution. In a typical microarray experiment, the sample is labeled and hybridized to the specific probes under high stringency conditions that prevent unspecific hybridization from occurring. Hybridized sample is then washed, stained and scanned for signal detection. Several manufactures are providing researchers with tiling microarray platforms nowadays. The Microarrays used in this study were strand-specific tiling arrays provided by the Affymetrix company, which rely on chemiluminescent labeling for the detection of the hybridized sample (Figure 10).

A)



B)



C)

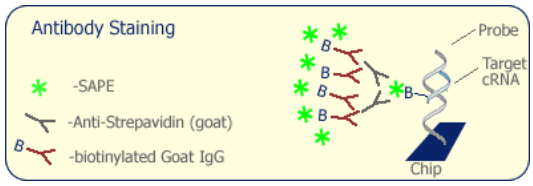


Figure 10. Schematic pipeline for labeling and detection of samples in *Affymetrix* tiling array platforms. (A) Total RNA samples are labeled using Affymetrix's whole transcript (WT) sense target labeling assay, through which biotinylated ribonucleotide analogs are incorporated to the cDNA produced from the original RNA sample. (B) Schematic disposition of the probes on a tiling microarray. **(C)** Staining is then performed by incubation of the labeled sample in a mixture containing Streptavidin, R-Phycoerythrin Conjugate (SAPE). Phycoerythrin also known as R-Phycoerythrin (RPE) is an intensely bright phycobiliprotein with maximum absorption at approximately 496, 546 and 565 nm, and a fluorescence emission maximum of ~578 nm.

3.2. RNA-Seq technology

Microarray technology has been successfully used for decades and has proven to be a powerful tool for the study of transcriptomes. However, the recent development of novel high-throughput DNA sequencing methods has provided new methods for both mapping and quantifying transcriptomes at higher resolution. One of these deep-sequencing based methods based is termed RNA sequencing technology (RNA-Seq). In a typical RNA-Seq experiment, a population of RNA molecules (either total or fractionated, such as poly(A) RNA) is converted to a library of cDNA fragments with adaptors attached to one or both ends. Each molecule, with or without amplification, is then sequenced in a high-throughput manner to obtain short sequences from one end (single-end sequencing) or both ends (pair-end sequencing).The reads are typically 30–400 bp long, depending on the DNA-sequencing technology used. Following sequencing, the resulting reads are either aligned to a reference genome or reference transcripts, or assembled *de novo* without the genomic sequence to produce a genome-scale transcription map that consists of both the transcriptional structure and/or level of expression for each gene (Figure 11).

3.3. Microarrays *versus* RNA-seq technology

Since its inception, RNA-seq has been compared to microarray technology in terms of how they generate transcriptome information. Both follow a similar path to answering a biological question, starting from experimental design, followed by data acquisition, and finally data analysis and interpretation. They share several similarities, such as high reproducibility, the lack of need to perform technical replicates of the experiment, and as every technique, the existence of certain biases intrinsic to each of them. However, there are a few key differences between the two technologies, which will determine the platform of choice when designing a given experiment. First, using microarray technology limits the researcher to detecting transcripts that correspond to existing genomic sequence information. RNA-seq experiments on the other hand serve for both the investigation of known transcripts and the exploration of new ones.

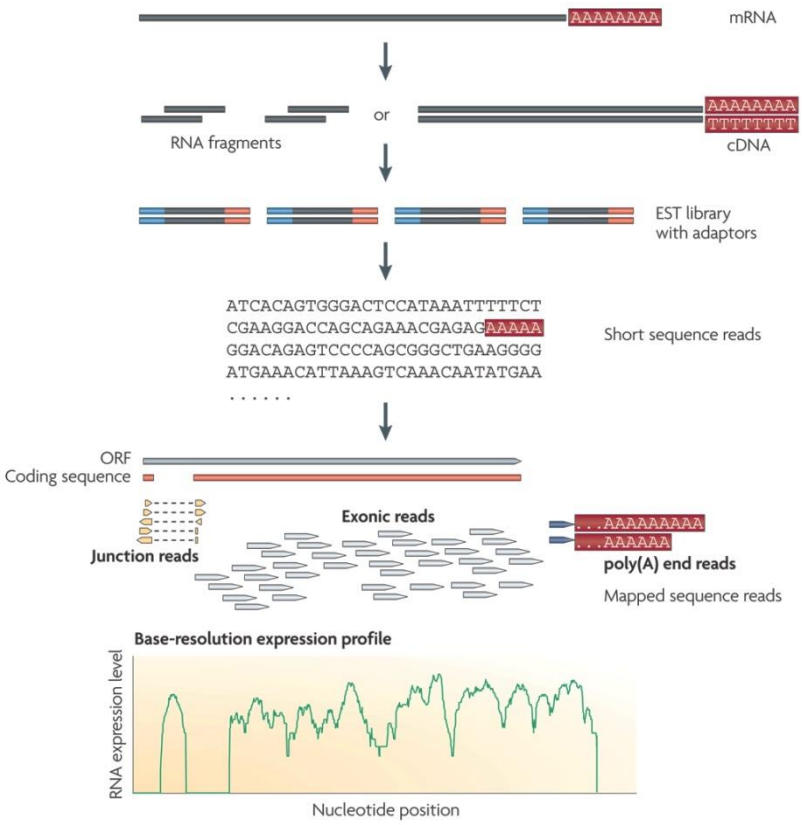


Figure 11. **Pipeline of a typical RNA-Seq experiment.** Firstly, long RNAs are first converted into a library of cDNA fragments through either RNA fragmentation or DNA fragmentation. Sequencing adaptors (in blue) are subsequently added to each cDNA fragment and a short sequence is obtained from each cDNA using high-throughput sequencing technology. The resulting sequence reads are aligned with the reference genome or transcriptome, and classified as three types: exonic reads, junction reads and poly(A) end-reads. These three types are used to generate a base-resolution expression profile for each gene, as illustrated at the bottom; a yeast ORF with one intron is shown. Taken from Wang *et al.* (2009)[153].

Even with organisms lacking a reference genome, *de novo* transcriptome assembly and differential expression analysis can be performed. Therefore, RNA-seq is ideal for discovery-based experiments. In addition, sequencing data can be reanalyzed when new sequence information is available, whereas a microarray template would have to be re-designed to investigate novel genes. Second, unlike microarray experiments, RNA-seq delivers low background signal. This is because DNA sequences can be unambiguously mapped to unique regions of the genome. As a result, noise in the experiment can be eliminated during analysis. Hybridization issues seen with microarrays (i.e. cross-hybridization) are also eliminated in RNA-seq experiments which implies an extra signal-to-noise

advantage. Finally, RNA-seq has the ability to quantify a large and tunable dynamic range of expression levels, which is dependent on the depth of sequencing, meaning that in theory, by sequencing deep enough, one can get the same dynamic range as the number of actual RNA molecules present in the sample. This allows RNA-seq to provide absolute values of abundance, rather than the relative ones that can be achieved when using microarrays. Microarray technology on the contrary has a relatively fixed dynamic range, showing saturation at high levels, while suffering loss of signal at the low levels, thus failing to detect transcripts that are very low in abundance. However, the two methods agree fairly well for genes with medium levels of expression, showing a correlation coefficient of 0.5 for this group of genes[153].

However, in most cases, the cost of performing a microarray-based experiment (100-200€/sample) is lower than an RNA-seq based experiment (300-1500 €/sample). Furthermore, the data analysis for this technology requires extensive bioinformatics skills and yields very large file sizes that are costly to store, consuming high amounts of computer resources (CPU and RAM) and making data sharing more complicated, whereas microarray data analysis is simpler and straightforward, and does not require specific computing infrastructure. Hence depending on the specific goal of the project, microarray technology will sometimes be sufficient to address the matter of interest. In many other instances, the best strategy may be a combination of both RNA-seq and microarray approaches, in order to maximize the effectiveness of each, while overcoming specific limitations of each of the technologies.

4. Nascent transcription

4.1. Concept of transcription rate (TR) and mRNA amount, and the importance of TR determination.

In order to characterize the transcriptional landscape of a given organism under a particular environmental condition, most studies focus on the determination of mRNA amount (RA), which refers to the number of fully processed (mature) mRNA molecules that are released in the cytoplasm for each gene at a given time (Figure 12). RA determination is useful as it reflects the effective transcriptional input: it measures levels of mRNAs that will later be

translated into proteins if needed, and thus are likely to have a significant impact on cell physiology. For chemical reactions the correct parameter to be used is mRNA concentration ([RA]) instead of RA. This should be taken into account when dealing with cell samples having different cytoplasmic volume[154].

The measured [RA] of a given gene, however, depends on the equilibrium between its respective mRNA synthesis and degradation rates (SR and DR respectively), as well as external factors such as particular conditions of the study and sample processing. For this reason, studies that exclusively focus on mature mRNA levels fail to address how these factors lead to the observed RA abundance[155]. The SR, understood as the rate of appearance of mature mRNA molecules in the cytoplasm[154], is generally a reflection of the transcriptional activity. However, the very first step of transcription, influenced by transcription factors via cell signaling under a specific physiological event, occurs at the nucleus, where new RNA molecules are produced independent of their subsequent functional roles. For this reason, a more robust indicator of how a given physiological event affects transcription at its very basal level is the measurement of not only mature mRNA molecules, but all nascent transcripts that are produced in response to the stimulus. In a steady state condition (when [RA] is constant over time) the following equation applies (where k_d states for the kinetic degradation constant of the mRNA):

$$SR = DR = k_d\,[RA] \qquad \text{[eq. 1]}$$

To meet this need, researchers focus on a different parameter involved in the flow of gene expression, which is the nascent transcription rate (nTR). nTR refers to the rate of production of new mRNA molecules onto the chromatin, and is defined as the number of these that are produced by the RNApol per unit of time acting on a single gene copy (molecules/min/gene copy). This parameter has been used for years as an estimator of how actively a gene is being transcribed, based on the idea that the number of nascent RNA molecules on a given gene is directly proportional to the transcriptional activity acting on it.

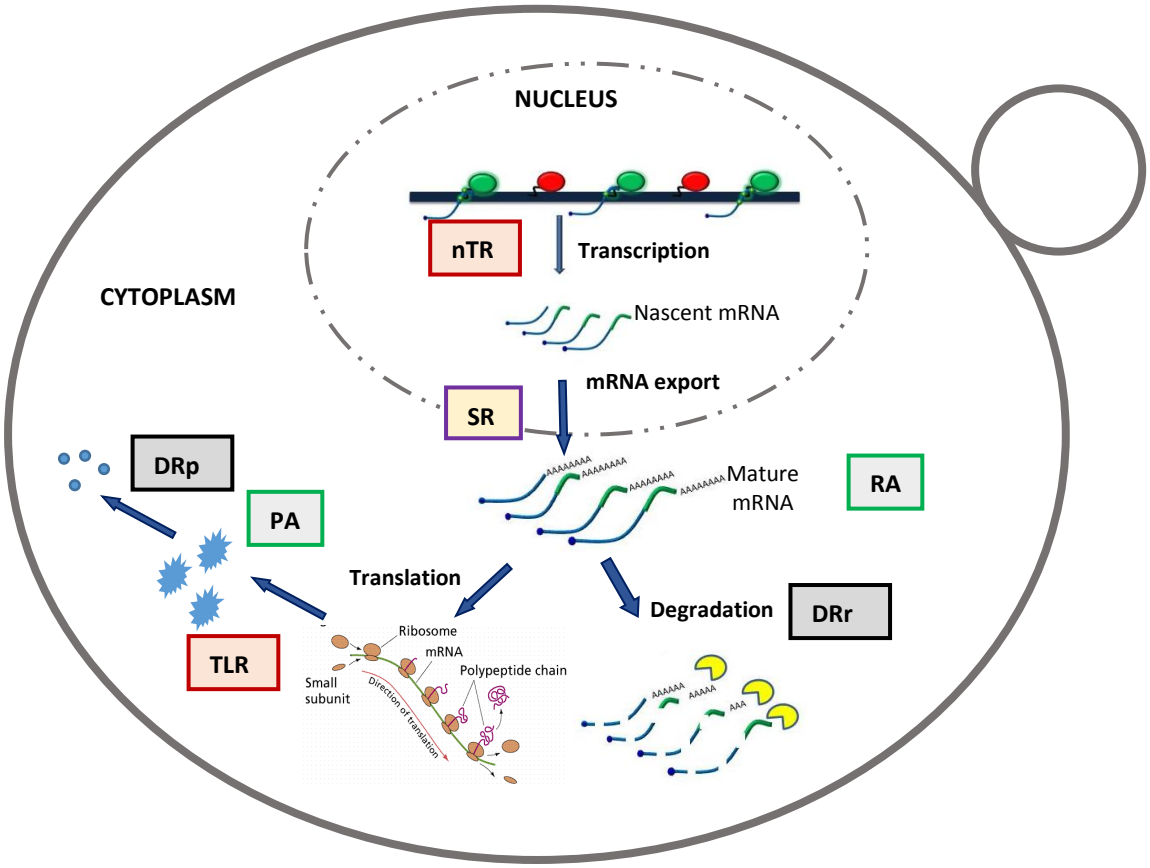


Figure 12. Parameters used for the estimation of mRNA and protein abundances in a yeast cell. The nascent transcription rate (nTR, red box) refers to the rate of production of new mRNA molecules within the nucleus. The nascent RNA molecules are then capped and polyadenylated for maturation and released into the cytoplasm. The mRNA amount (RA, green box) refers to the number of fully processed (mature) mRNA molecules that are released in the cytoplasm for each gene at a given time. The rate at which these molecules are released in the cytoplasm is referred to as the mRNA synthesis rate (SR, purple box). In the cytoplasm, mature mRNA molecules will then be either translated into proteins or degraded to account for the cell needs. The rate at which mRNA molecules are degraded in the cytoplasm is referred to as the degradation rate (DR, grey box). The rate at which protein molecules are produced is referred to as the translation rate (TLR, red box). The total amount of protein molecules present in the cytoplasm at a given time is referred to as protein amount (PA, green box). The rate at which proteins are degraded is known as protein degradation rate, herein referred to as DRp (grey box).

Similarly, the abundance of RNApol molecules on the gene also serves as an indirect estimator. As a consequence, nTR values are generally determined either measuring the number of RNApol molecules sitting on a specific gene, or the number of nascent RNA molecules that arise from it at a given time.

At a given average speed, the transcriptional activity on a gene is directly proportional to the total number of elongating RNApol molecules present on it. However, because different genes have different lengths, this number must be corrected by the gene length, thus the relevant parameter for nTR estimation will not really be the number of RNApol molecules, but their density in molecules/kb[156]. This density will depend on several factors such as the number of transcription initiation events at the gene promoter, changes of RNApol speed during transcription, or the number of abortive elongations observed. All these three factors are regulated processes[68] and must be taken into account for nTR calculations.

nTR determinations are of great utility for many researchers nowadays as they help gain insight into the underlying regulatory mechanisms that govern the observed RA levels in the cell under a given condition. A good example of how this tool can be utilized to address cellular processes was described in a recent publication by members of our laboratory of a genome-wide study that, by setting apart the independent contributions of RA, nTR and DR, proved an mRNA degradation factor, Xrn1, to be involved in mRNA synthesis. This study proved the process of gene expression to be circular, in which all independent steps (i.e. mRNA synthesis and decay) are tightly interconnected[157].

4.2. Genome-wide methods to evaluate transcription rates in yeast.

As mentioned before, nTRs are generally measured estimating either the density of RNApol molecules present on a gene or the number of nascent RNA molecules that are produced from it. To do so, several techniques have been implemented in yeast:

4.2.1. Genomic-run on (GRO)

"Genomic Run-on"(GRO) is a technique developed in our research group[158] that was implemented a decade ago for the quantification of nTRs in *S. cerevisiae*. It is a scale-up

approach of the "Transcription Run-on" technique (TRO)[156] that was previously implemented by other researchers to determine nTRs in different cell types. GRO technique is based on the estimation of nascent RNA abundance, via the labeling of the RNA molecule as it is being transcribed with radioactively-labeled nucleotide analogs. A typical GRO assay basically consists of 3 steps (Figure 13):

- *Permeabilization of the yeast cells*. Cell permeabilization is achieved by adding the detergent Sarkosyl, which provokes the following: 1) creates pores in the membrane allowing the cellular content to flow out, thus depleting pre-existent nucleotide pools and stopping all physiological processes. When this happens, polymerases stop transcription on the specific sites they were elongating at the moment the detergent was added.2) Sarkosyl molecules disrupt chromatin structure while leaving the DNA-RNApol- RNA ternary complex intact. This facilitates RNApol elongation on a nucleosome-free template.

- *Nucleotide pulse (run-on reaction)*. Nascent RNA is radioactively labeled as it is being created by subjecting the permeabilized cells to a nucleotide pulse in the presence of radioactive UTP. The addition of this new set of nucleotides allows polymerases to resume transcription to a certain extent (50 nt on average[159]), yielding short labeled RNA molecules.

-*Isolation of the labeled RNA and hybridization to specific probes*. For nascent RNA detection and abundance measurements RNA is purified and hybridized on macroarray that contains specific probes for thousands of RNApol II genes across the *S. cerevisiae* genome. The signal intensity observed for each gene is a reflection of the density of active polymerases on that gene. Further on, assuming a constant RNApol elongation rate along the gene, the correspondent nTR for this gene can be calculated[155].

nTR may be considered a proxy of SR if corrected by cell volume[154]. Thus, nTR determination, along with the correspondent [RA] measurement in the same sample, allows

indirect estimations of mRNA stability for the studied genes (see equation [1] and Figure 13).

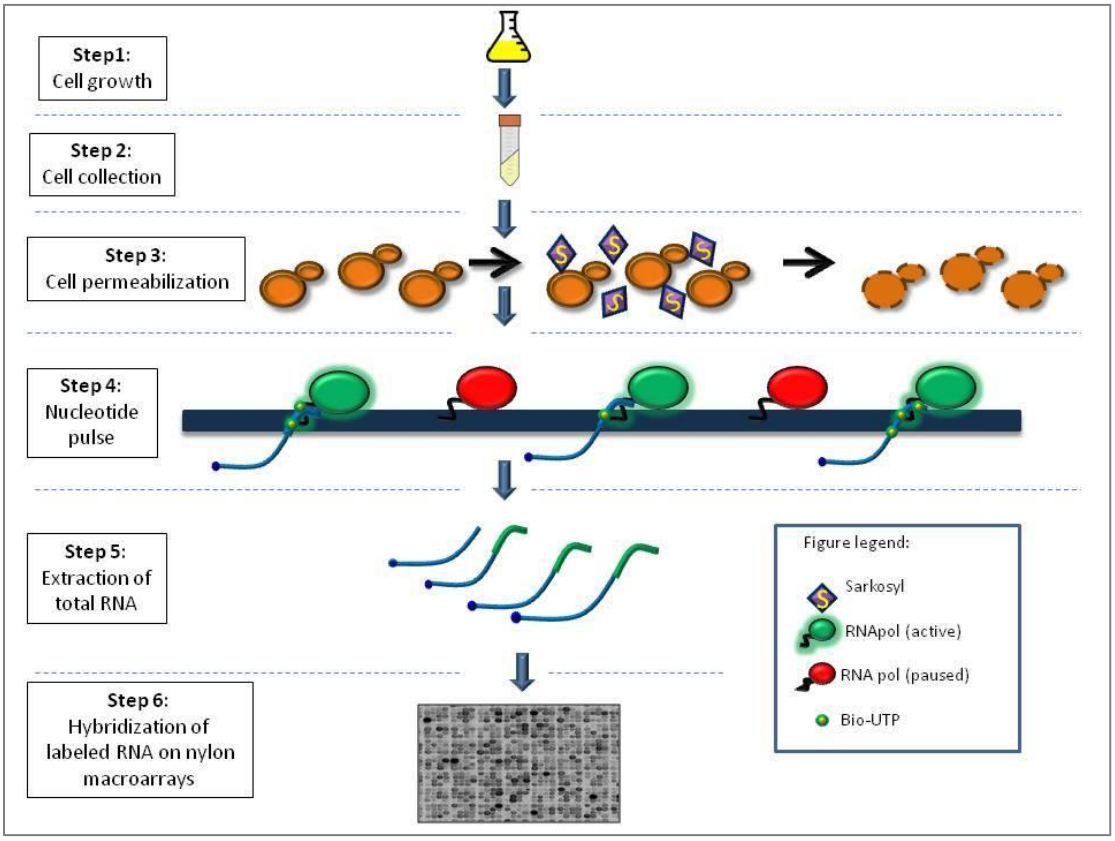


Figure 13. **Experimental pipeline of a typical GRO experiment.** Cells are grown in the desired conditions up to exponential growth phase, to then be harvested and made permeable by the addition of the detergent Sarkosyl. RNA polymerases are then subjected to a nucleotide pulse containing a radioactively labeled nucleotide analog that allows transcription resumption. Labeled RNA is then isolated and hybridized on a nylon macroarray containing oligonucleotide probes for the majority of genes of the *S. cerevisiae* genome.

4.2.2. RNApol II ChIP-on-chip (RPCC)

RPCC is a chromatin immunoprecipitation (ChIP)-based technique that uses specific antibodies against RNApol II subunits. This technique relies on the fact that, upon the addition of the crosslinking agent formaldehyde, RNApol molecules crosslink with DNA sequences that are transiently bound to them during transcription[160,161]. By isolating and amplifying these sequences via polymerase chain reaction (PCR) this technique can be

scaled at the genomic level through further labeling and hybridization of the amplified DNA sequences on tiling arrays. The GRO technique allows the tracking of transcriptionally active polymerases, whereas RPCC allows the tracking of polymerases that are bound to DNA, either active or paused. This difference allows the comparison of GRO with RPCC data which can lead to the discovery of elongation-based biological phenomena, while allowing detection and correction of technique-specific biases for each methodology[162,163]. RPCC assays can be virtually performed using any antibody that specifically recognizes RNApol II. However, the quality of the results will depend on the efficiency and specificity of the antibody, and hence the antibody of choice should be carefully selected prior to the experiment. Prior studies have successfully used antibodies against either a tagged version of RNApol II or against the different phosphorylation states of the CTD, but antibodies against other RNApol II subunits may also be used[97,164].

4.3. Genomic run-on (GRO): past and present

The radioactive approach of GRO technique has been successfully used over the years in the model yeast species *S. cerevisiae*, yielding a wide set of publications that address nTR determination under different experimental conditions[157,165,166]. Due to the implementation of this technique, a complete data set of transcription rates of different strains under various biological conditions is now available for this organism[162]. However, this technique is based on the use of radioactive-labeled samples that are hybridized on DNA macroarrays, which, despite being useful for standard transcriptomic studies, limits the resolution of this technique to the study of expression changes of known genes without resolution inside the transcribed region.

With the arrival of next-generation sequencing technologies, the technique was made compatible with current DNA microarray platforms by substituting the radioactive UTP analog (either ^{33}P or ^{32}P) for Biotin-UTP (Bio-UTP) in a new version of the technique that is based on chemiluminescent detection termed “BioGRO”[167] (Antonio Jordán-Pla, doctoral thesis 2013). By hybridizing the Biotin-labeled sample on high-resolution strand-specific tiling arrays, this new technique allows the study of transcription in much greater detail, enabling the monitoring of all regions of the genome, both protein-coding and non protein-

coding, as well as permitting the study of antisense transcription as the nature of the data is strand-specific[159]. Despite being suitable for studies at higher resolution, chemiluminescent labeling has a main drawback compared to the radioactive approach which is the significant loss of sensitivity. When total RNA is isolated from a GRO experiment, two main RNA populations are retrieved: newly synthesized nascent RNA and mature RNA that was already present within the cell at the moment of the extraction. Although the former accounts for a small minority of the total population of RNA molecules (between 0.05 and 0.34 % in yeast according to previous studies[132,168]), mature RNA, which is a contaminant for this purpose, remains to be the large majority. Total RNA isolated from a classical GRO experiment containing these two RNA populations can be directly hybridized to achieve good overall sensitivity. However, this level of sensitivity cannot be achieved in a BioGRO assay by the same procedures. To compensate for the lack of sensitivity, two main things were modified in the BioGRO protocol (Figure 15): first, an RNase A treatment is performed prior to the nucleotide pulse; and second, total RNA is size-selected after isolation. The RNase treatment aims to first deplete contaminant pre-existing RNA molecules within the cell and then adjust the proper RNase concentration, where RNase A trims the nascent RNA chain that is already hanging from the polymerase prior to the nucleotide pulse without affecting run-on efficiency[169]. By doing this, when the pulse is given and polymerases resume transcription, the 5' end position of the nascent RNA will reflect the precise location of the polymerase at the moment of Sarkosyl treatment. Additionally, the size selection step further enriches the total RNA sample in nascent RNA molecules, which have been estimated to be short (50 nt long on average[159]).

In summary, this new version of the GRO technique improves the classical version in several ways: first, it avoids the use of radioactivity; second, it allows the study of transcription in a strand-specific manner at much higher resolution; and third, it increases mapping resolution.

Despite these advantages, BioGRO still suffers from lack of sensitivity: with the current approach it has only been possible to obtain reliable TR data for around 809 genes that are included in the population of genes that are transcribed at the highest level (high TR genes)

which yield good hybridization signal intensity[159]. This addresses the need to broaden the scope of this technique, in order to obtain TR data for the whole gene set of this organism instead of being restricted to a relatively representative number of the population.

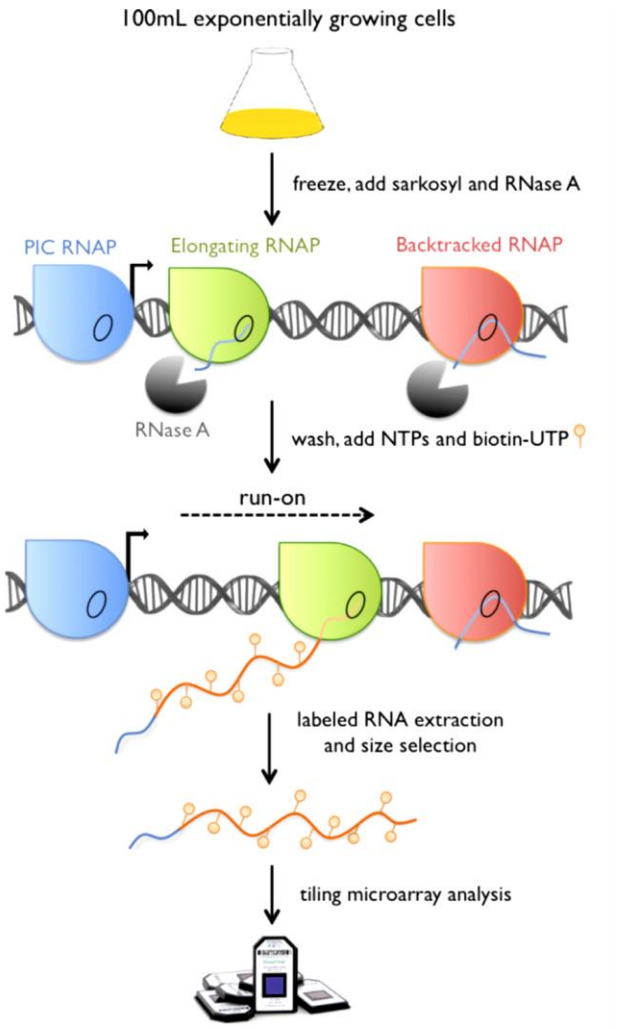


Figure 14. **BioGRO workflow**. Cells are grown in the desired conditions to exponential growth phase and harvested to then be made permeable by the addition of Sarkosyl and subjected to an RNAse treatment to deplete pre-existent mature RNAs within the cell. Transcription is resumed by the addition of a nucleotide pulse containing Bio-UTP, an analog that allows labeling and further detection of the RNA sample by hybridizing it on Affymetrix tiling microarrays. Taken from Jordán-Pla *et al* 2015[159].

5. Changes in gene expression as a response to changing extracellular environments in yeast

When studying the pathogenic mechanisms that underlie a given human pathogen, the way it adapts and responds to survive both changing environments and the response of host defenses are important issues to address. Invasive pathogens face many different microenvironments when spreading throughout the body, interacting with the host in different niches[170] in which the environment is continually changing and the organism must adapt to survive. In this sense, organisms that can adjust to changes in the environment are likely to exhibit a greater degree of adaptiveness than those that cannot. Environmental changes, irrespective of their source, cause a variety of "stresses" or "shocks" that the pathogen must face repeatedly, and to which its genome must respond in a programmed manner. How the genome perceives and transduces environmental stimuli to subsequently regulate the expression of pertinent genes is the focus of many research studies nowadays. Sensing mechanisms first alert the cell to imminent danger, and then trigger the orderly sequence of events that will mitigate it. Many, though not all signals, are perceived at the cell surface by plasma membrane receptors. Activation of such receptors by mechanisms such as ligand binding may lead to alterations in other cellular components, ultimately resulting in alterations in cell shape, ion conductivity, gene activity, and other cellular functions[171]. The sensing of the signal and its transduction to the transcriptional machineries catalyzes transcription initiation on those DNA sequences that allow the cell to activate its defenses against the environmental insult, expressing an array of genes that are needed for the response and for the trigger of the required metabolic changes.

Studies of changes in gene expression in *C. albicans* under different infection scenarios (i.e. mucosal versus systemic) have reported that only very few genes are strongly induced in different conditions, suggesting that the transcriptome of this organism within the human host is mostly niche-specific[172]. However, genome-wide expression profiling revealed that this organism also seems to display a core stress response (similar to that observed in other yeast species such as *S. cerevisiae* or *S. pombe*) in which a fraction of the genome is

expressed in a stereotypical manner in response to different stress conditions, regardless of their nature[31].

It is commonly understood that most changes in gene expression occurring during a stress response occur at the transcriptional level[173]. For this reason, the next sections mainly focus on transcription. However, although minor, post-transcriptional regulation has also been described, either related to mRNA stability[155,168] or regulation of translation[174]. The contribution of translation to these changes is described in the last sections.

5.1. Changes in gene expression in response to hypoxic stress in *C. albicans*.

5.1.1. Hypoxia: concept and occurrence in yeast

Molecular oxygen (O_2) plays an essential role in several metabolic pathways, such as the biosynthesis of sterols, unsaturated fatty acids, NAD, and porphyrins[175-177]. It also acts as an electron acceptor in the generation of chemical energy via mitochondrial respiration. For this reason, the level of available oxygen is a critical factor in determining overall cellular metabolism in eukaryotes. Moreover, as most eukaryotic human fungal pathogens are considered obligate aerobes, oxygen availability during fungal pathogenesis can play a critical role in the outcome of infection.

Oxygen availability is usually described to range from 0 to 21% O_2. In a given niche, the environment is usually described as anaerobic or "anoxic" when there is complete absence of oxygen (0% O_2); "hypoxic" when available oxygen is reduced compared to atmospheric levels; and "normoxic" when oxygen availability reaches normal atmospheric levels (21% O_2)[178]. The oxygen levels a given pathogen is exposed to will differ depending on the site of infection, for example, oxygen will be abundant on skin, exposed mucosal layers or blood, but sparse in the gut and deep tissues. Even though the exact oxygen concentrations at sites of infection have not been measured directly *in vivo*, and this level will vary with anatomical location, it is generally accepted that hypoxia occurs at sites of infection, generating, as a consequence, significant environmental stress on most pathogen cells[179-182]. In healthy tissues in the human body, oxygen levels of 2.5% to 9% are considered normoxic, while oxygen levels of ≤1% (described in tumors and wounds) are typically considered

hypoxic[180,183-187]. In the case of *C. albicans*, its normal anatomical location is the human gastrointestinal tract, which contains significant regions of hypoxia[188,189]. As a normal commensal organism living in this part of the body, it is optimally adapted to live in both aerobic and hypoxic environments, oxidizing carbohydrates through respiration in aerobic conditions and fermenting them under hypoxic conditions[190]. In the context of infection, along with its inherent capability to switch from normoxic to hypoxic growth mode, numerous host components may become available for the pathogen facilitating hypoxic growth. For example, *C. glabrata* has been reported to replace or at least supplement its ergosterol (essential cell membrane lipidic component in fungi) with cholesterol recruited from host cells. This property has been proposed to be responsible for its elevated resistance to azoles, drugs that inhibit sterol biosynthesis[191].

In order to understand how *C. albicans* adapts to oxygen limitation, the molecular mechanisms associated with the ability to adapt to hypoxia have been investigated through global transcriptomic studies. Three major analyses of the *C. albicans* transcriptomic response to hypoxia have been published[192-194]. These studies show concordant results, despite differences in experimental design, such as fungal strain used and the way the specific oxygen levels were applied. All these studies reported that, upon exposure to low oxygen levels, *C. albicans* upregulates genes involved in iron metabolism, heme biosynthesis, fatty acid metabolism, ergosterol biosynthesis, glycolysis and fermentation, cell wall and membrane structure, and hypha-specific transcripts[193-195]. In contrast, expression of transcripts involved in oxidative respiration, such as those from the tricarboxylic acid cycle and mitochondrial respiration chain along with general ATP synthesis, were down-regulated [193-195]. These genes are collectively referred to as “hypoxic genes”. (Figure 15).

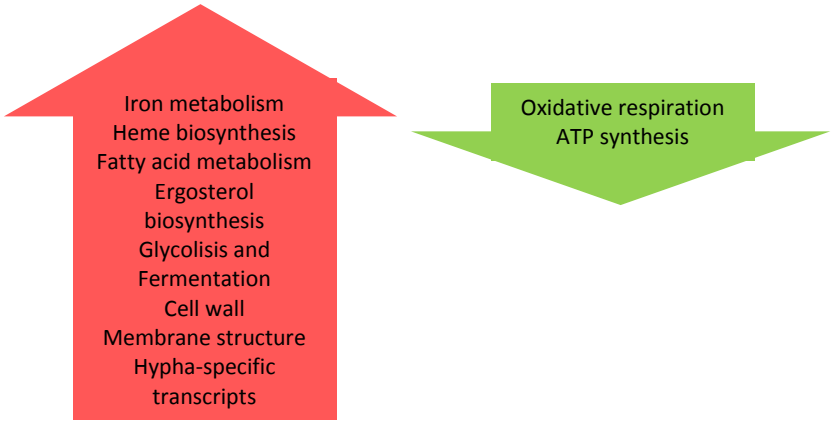


Figure 15. Gene categories up and down-regulated under hypoxia in *C. albicans* as reported by four different global transcriptomic studies. Up-regulated genes are shown in red and down-regulated genes are shown in green.

5.1.2. Sensing of oxygen levels in yeast and regulation of genes involved in the hypoxic response

Sensing of oxygen levels in yeast occurs through the perception of the intracellular levels of specific molecules that are only synthesized in the presence of oxygen, such as heme, unsaturated fatty acids and ergosterol (Figure 16).

- Oxygen is required for the **biosynthesis of heme**, in which its presence activates the transcriptional regulator Hap1[196]. The function of Hap1 is to induce the expression of genes involved in respiration and oxidative stress responses while promoting the biosynthesis of the transcriptional repressors Rox1 and Mot3 to mediate the repression of hypoxic genes through the action of the Tup1-Ssn6 complex[197]. Hap1 activity is controlled *in vivo* by heme (and not by its precursors) being Hap1 activated by the presence of heme even in anoxic cells[196]. Conversely, in hypoxic conditions, the lack of heme biosynthesis prevents the synthesis of both Rox1 and Mot3 repressors, allowing transcription of hypoxic genes. Even though both Hap1 and Rox1 are important regulators they only partially govern hypoxic responses, as they only regulate the expression of less than half of all hypoxia-activated and only a few of the hypoxia-repressed genes[198].

- Oxygen is also required in the process of **fatty acid desaturation**. In the absence of oxygen, the lack of desaturation results in a decrease in membrane fluidity, which can trigger partial

proteolysis and release of membrane-bound precursors of the Mga2 and Spt23 transcription factors, that induce the expression of hypoxic genes such as OLE1, a fatty acid desaturase[199,200].

- The process of **Ergosterol biosynthesis** also requires oxygen. Ergosterol is a sterol found in the cell membranes of fungi, serving many of the same functions that cholesterol serves in animal cells. This molecule is essential for the viability of both budding and fission yeast, which is the reason why the enzymes needed for its synthesis have become important targets for drug discovery. In fission yeast as well as in mammalian cells, sterol depletion leads to the export of Sre1 (SRBP) precursor from the ER membrane to the Golgi, wherein it is cleaved to permit nuclear import and activation of hypoxic genes[201]. Budding yeast does not contain a homologous Sre1 system, instead, oxygen sensing via sterol levels occurs in part by the Upc2 transcription factor[202]. In addition to all these, some studies reported alternative oxygen sensing mechanisms that relate to respiratory chain activity, including the production of oxygen radicals and nitric oxide[203,204].

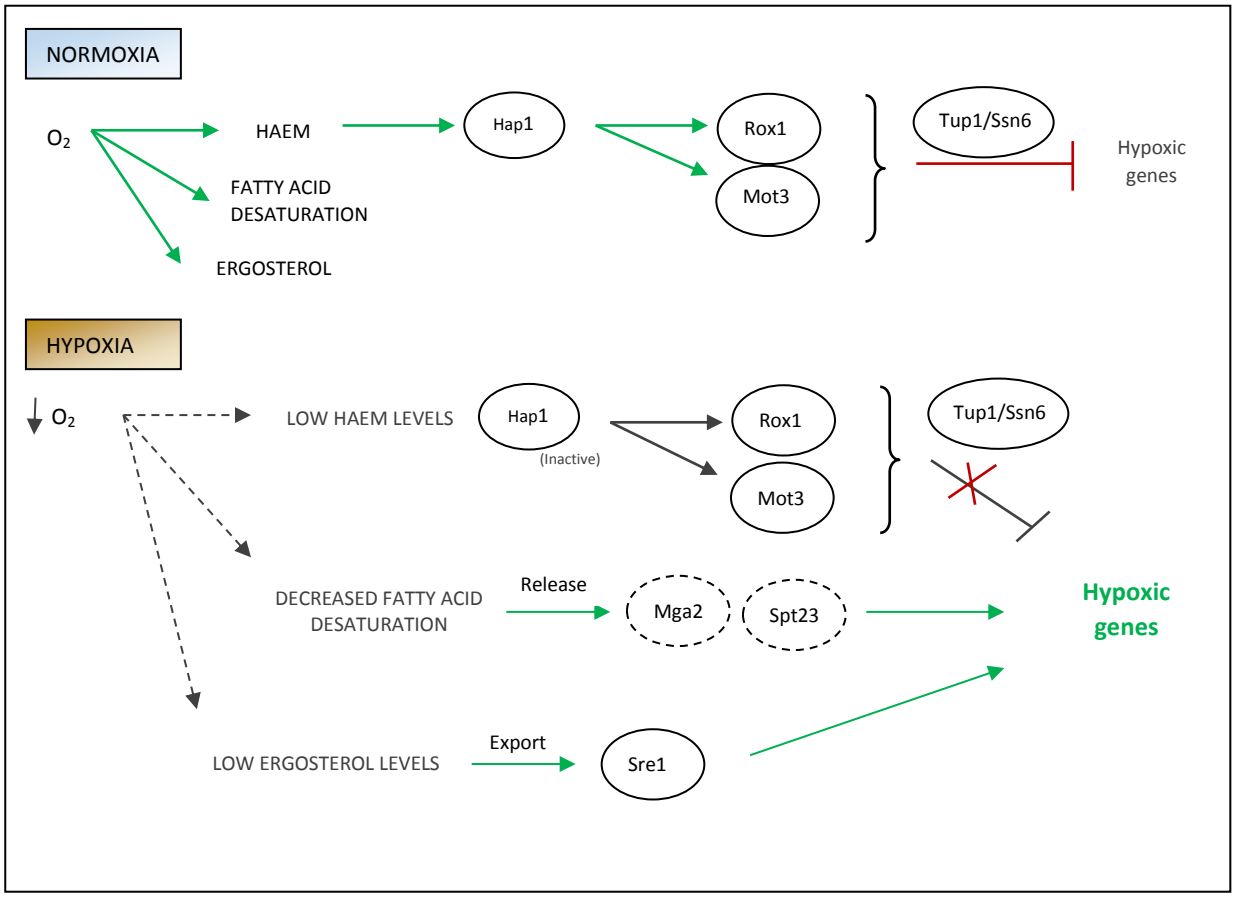


Figure 16. Activation and repression of hypoxic genes in response to the intracellular levels of molecules for which synthesis oxygen is required in S. cerevisiae. In normoxic conditions, normal levels of heme, ergosterol and fatty acid desaturation maintain hypoxic genes in a repressed expression state. In hypoxic conditions, decreased levels of these molecules trigger the activation of the transcription factors implicated in the expression of hypoxic genes.

Compared to fission and budding yeast, the molecular mechanisms that allow *C. albicans* to adapt to hypoxic environments still remain largely unknown. Major oxygen sensing pathways established in budding yeast do not apply to *C. albicans*, because it does not contain functional homologues of neither Hap1 nor Rox1 hypoxic regulators, and the Tup1/Ssn6 complex does not repress the expression of hypoxic genes[205,206]. Instead, other hypoxic regulators have been reported, that do not correspond to the *S. cerevisiae* homologues identified in hypoxic responses in the model organism.

The transcription factor **Efg1** has been confirmed as major regulator of the hypoxic response for several reasons. First, it is required to allow hypoxic regulation of about half of all genes that are normally regulated by hypoxia, as observed in a study where the authors reported that hypoxic up-regulation or down-regulation of genes (e.g. CAT1 encoding catalase and RIP1 encoding cytochrome c reductase respectively) are abolished in an *efg1* mutant. In accordance with this, the *efg1* mutant strain lacks effective repression of several genes under hypoxia, including OLE1. Finally, Efg1 prevents inadequate hypoxic regulation of numerous genes that are not normally up- or down-regulated in hypoxia[193]. The role of Efg1 in hypoxic responses was initially suggested upon the observation that *efg1* mutants, that are blocked in hypha formation in normoxic conditions, were able to filament under microaerobic conditions if grown on or within agar at temperatures ranging from 25 to 35 ^{o}C[207,208].Efg1 is a member of the APSES group of transcriptional regulators in *Ascomycetes* that control sphere-filament interconversions, and it is a downstream target of the protein kinase A pathway that leads to hypha formation. Closely associated with Efg1, **Flo8**, another TF involved in this pathway, leads to hypha formation and regulates a subset of Efg1-regulated genes. Genes repressing hyphal filamentation in hypoxia are apparently among genes co-regulated by Efg1 and Flo8, as *flo8* mutants showed enhanced filamentation during embedded growth in agar[209]. This type of growth along requires the TF **Czf1**, that has opposing functions to Efg1 as it prevents its repressor activity[210].

Another important player involved in the hypoxic response in *C. albicans* is the TF **Ace2**, which is the homologue of the Swi5/Ace2 TF in *S. cerevisiae* that drives M-G1 cell cycle transition. In normoxia, Ace2 has similar functions to Efg1, however its roles in hypoxia differ. Similar to *efg1* mutants, strains lacking Ace2 downregulate glycolytic genes and upregulate genes involved in respiration[211]. However, morphogenetic changes occurring in response to hypoxia are different between these two strains, as *ace2* mutants are unable to filament[211], while *efg1* mutants grow in hyperfilamentous form. This capability makes Ace2 a positive regulator of biofilm formation in normoxia, but not in hypoxia[212].

As previously mentioned, ergosterol levels serve as measure of oxygen availability. However, homologues of the SREBP sterol-sensing system have not been identified in *C. albicans*. Instead, **Upc2**, a homologue of the Upc2 zinc finger TF in budding yeast was identified to play this role in this organism[213]. Upc2 contains a C-terminal membrane anchor that is cleaved off upon ergosterol depletion, releasing the factor for subsequent nuclear import and activation of ERG genes.

5.2. Changes in gene expression in response to oxidative stress in *C. albicans*

5.2.1. ROS, concept of oxidative stress and molecular defense against ROS

In its ground state (normal configuration, O_2), molecular oxygen is relatively unreactive. However, during normal metabolic activity, or as a consequence of various environmental perturbations, O_2 is capable of giving rise to reactive excited states such as free radicals and derivatives[214,215]. These oxygen states are generally referred to as reactive oxygen species (ROS), which include the superoxide radical (O_2^{*-}), hydrogen peroxide (H_2O_2) and the hydroxyl radical (OH^*) (Table 1). Unless abated, all these species are extremely reactive, interacting with proteins, lipids and nucleic acids inflicting extensive molecular damage (summarized in Table 2). To minimize the damaging effects of ROS, aerobic organisms evolved both non-enzymatic and enzymatic antioxidant defenses (Table 3). Non-enzymatic defenses include compounds of intrinsic antioxidant properties, such as Vitamins C and E, glutathione, and β-carotene. Enzymatic defenses include superoxide dismutases (SOD), catalases (CAT) and peroxidases, whose activities that protect the cell by directly scavenging superoxide radicals and hydrogen peroxide to convert them into less reactive species. In particular, SODs deal with the first product of the univalent reduction of O_2, converting it to H_2O_2, which must then be destroyed by CAT and/or peroxidases (Figure 17). Thus, the SOD and CAT serve, in tandem, as front-line antioxidant defenses.

Name	Notation	Some comments and basic sources
Molecular oxygen (triplet ground state)	O_2; $^3\Sigma$	Common form of dioxygen gas
Singlet oxygen (1st excited singlet state)	$^1O_2$; $^1\Delta$	Photoinhibition; UV irradiation; PS II e^- transfer reactions (chloroplasts)
Superoxide anion	$O_2^{\bullet-}$	Formed in many photooxidation reactions (flavoprotein, redox cycling); Mehler reaction in chloroplasts; mitochondrial e^- transfer reactions; glyoxysomal photorespiration; peroxisomal activity; nitrogen fixation; reactions of O_3 and $OH^\bullet$ in apoplastic space; defense against pathogens; oxidation of xenobiotics
Hydrogen peroxide	H_2O_2	Formed from $O_2^{\bullet-}$ by dismutation; photorespiration; ß-oxidation; proton-induced decomposition of $O_2^{\bullet-}$; defense against pathogens
Hydroxyl radical	$OH^\bullet$	Decomposition of O_3 in apoplastic space; defense against pathogens; reactions of H_2O_2 with $O_2^{\bullet-}$ (Haber-Weiss); reactions of H_2O_2 with Fe^{2+} (Fenton); highly reactive with all macromolecules
Perhydroxyl radical	$O_2H^\bullet$	Protonated form of $O_2^{\bullet-}$; reactions of O_3 and $OH^\bullet$ in apoplastic space

Table 1. Reactive oxygen species of interest in oxidative stress. Taken from Scandalios 2005[214].

Oxidative damage to lipids
• Occurs via several mechanisms of ROS reacting with fatty acids in the membrane lipid bilayer, leading to membrane leakage and cell death.
Oxidative damage to proteins
• Site-specific amino acid modifications (specific amino acids differ in their susceptibility to ROS attack)
• Fragmentation of the peptide chain
• Aggregation of cross-linked reaction products
• Altered electrical charge
• Increased susceptibility to proteolysis
• Oxidation of Fe-S centers by $O_2^{\bullet-}$ destroys enzymatic function
• Oxidation of specific amino acids "marks" proteins for degradation by specific proteases
• Oxidation of specific amino acids (e.g., Try) leads to cross-linking
Oxidative damage to DNA
• DNA deletions, mutations, translocations
• Base degradation, single-strand breakage
• Cross-linking of DNA to proteins

Table 2. Examples of ROS damage to lipids, proteins and DNA. Adapted from Scandalios 2005[214].

ROS are produced in all aerobic organisms and exist in the cell in normal conditions to maintain equilibrium with antioxidant molecules. Oxidative stress occurs when this critical balance is disrupted due to depletion of antioxidants, or when the formation of ROS increases beyond the ability of the defenses to cope. For this reason, a rapid and clear indicator of oxidative stress is the increase of endogenous ROS levels that is generally accompanied by induction of antioxidant defenses.

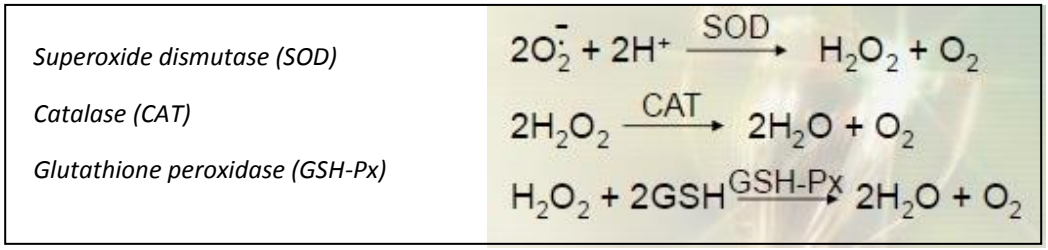


Figure 17. Mechanism of intracellular antioxidant enzymes. Intracellular Superoxide dismutases (SODs), Catalases (CATs) and Peroxidases (Px) protect the cell by scavenging superoxide radicals and hydrogen peroxide and converting them into less reactive oxygen species. Chemical reactions involved are shown for each enzymatic activity.

Non-enzymatic antioxidant molecules	
Antioxidant molecule	Subcellular location
Ascorbate (vitamin C)	Plastid; apoplast; cytosol; vacuole
β-Carotene	Plastid
Glutathione, reduced (GSH)	Plastid; mitochondrion; cytosol
Polyamines (e.g., putrescine, spermine)	Nucleus; plastid; mitochondrion; cytosol
α-Tocopherol (vitamin E)	Cell and plastid membranes
Zeaxanthin	Chloroplast
Antioxidant enzymes	
Enzyme	Subcellular location
Ascorbate peroxidase	Plastid stroma and membranes
Peroxidases (non-specific)	Cytosol; cell wall-bound
Catalase	Glyoxysome; peroxisome; cytosol; mitochondria
Superoxide dismutase (SOD)	Cytosol (Cu/ZnSOD); plastid (Cu/ZnSOD; FeSOD); mitochondrion (MnSOD); peroxisome
Dehydroascorbate reductase	Cytosol; plastid
Glutathione reductase	Mitochondrion; cytosol; plastid
Monodehydroascorbate reductase	Plastid stroma
Glutathione S-transferases	Cytosol; microsomal

Table 3. **Enzymatic and non-enzymatic antioxidant defenses.** Adapted from Scandalios 2005[214].

In the particular context of a pathogen, the sources of ROS can be either internal (emerging from virtually all intracellular organelles and compartments as a consequence of normal metabolic activity) or external. External oxidative bursts are naturally generated by activated macrophages and neutrophils during the course of infection, as part of their arsenal to neutralize microorganisms[216-221]. In this sense, the responsive antioxidant defense mechanisms of the microbe will play a crucial role in pathogenesis, since its success as a pathogen is partly based upon its resistance to oxidative stress and other environmental insults. In the case of *C. albicans,* its oxidative stress response seems to be a niche-specific phenomenon during the establishment of a systemic infection[222], and it is accepted that this response is at least partially necessary to establish a disseminated infection[223,224]. This idea is further supported by the fact that the major causative agents of candidiasis*, C. albicans* and *C. glabrata*, display overall greater resistance to oxidative stress compared to other pathogenic *Candida* species[225]. Moreover, other data obtained from *C. albicans* mutants revealed the existence of a correlation between resistance to oxidants *in vitro*, survival in phagocytes and virulence[223].

5.2.2. Response mechanisms to ROS by Candida *species*

To mitigate and repair the damage caused by free radicals, *Candida* species have developed antioxidant mechanisms involving the presence superoxide dismutases (SODs), catalases (CATs), thioredoxins (TRXs), glutaredoxins (GRXs), and peroxidases (PRXs).

Superoxide dismutases are antioxidant enzymes that catalyze the dismutation of superoxide radicals to hydrogen peroxide[226], which in the *C. albicans* genome includes a family of six SOD genes, which help neutralize the superoxide generated during the respiratory burst in several anatomical sites of the human body[227]. Four of these SODs are copper zinc (CuZn)-dependent (Sod1p, Sod4p, Sod5p, and Sod6p), and the other two are manganese-dependent (Sod2p and Sod3p)[223]. These enzymes are located in different organelles: Sod1p and Sod3p are found in the cytoplasm, Sod2p in the mitochondria, and

the rest on the cell surface. Among these, Sod1p, Sod4p Sod5p have been shown to be involved in pathogenesis of *C. albicans*[218,226,228].

Catalases catalyze the decomposition of hydrogen peroxide to water and oxygen. Catalases are largely, but not exclusively, localized in peroxisomes, where in many H_2O_2-producing enzymes reside. *C. albicans* has a single catalase (Cta1p)[229] that confers this organism resistance to this compound. Disrupting of the *CTA1* gene generates a mutant strain that is sensitive to H_2O_2 and exhibits attenuated virulence[230].

Peroxidases catalyze the reduction of H_2O_2 to water. Several peroxidase activities are found in the *C. albicans* genome, located in different parts of the cell. These include the mitochondrial cytochrome-c peroxidase *CCP1*[231], or others such as the thioredoxin peroxidase PRX1[232], or the hydroperoxide peroxidase *TSA1*[233]. In *C. albicans*, the location of the Tsa1p depends on growth conditions. Accordingly, an immunocytochemical study showed that this enzyme is located in the cytoplasm in the yeast morphotype while it is translocated into the nucleolus
during transition to hyphae[234]. Interestingly, in mutants lacking *TSA1*, the concentration of H_2O_2 increases by an order of magnitude, suggesting that the function of Tsa1p is not redundant with other enzymes that neutralize this peroxide. Tsa1p has also been proven indispensable in the yeast to hyphae transition under stress[234].

Glutaredoxins are cytosolic enzymes that protect and repair the thiol groups of proteins, and are responsible for reducing the disulfide glutathione linkages[235]. *C. albicans* has four genes encoding Glutaredoxins, namely *GRX1*, *GRX2*, *GRX3*, and *GRX5*. Chaves *et al.* showed that a mutant lacking the *GRX2* gene presented a deficiency in the formation of hyphae and was more susceptible to killing by polymorphonuclear macrophages[236]. Also, the virulence of the *grx2Δ* strain was attenuated after systemic infection in a murine model and was susceptible to menadione and resistant to diamide[235,236]. These observations indicate that Grx2p participates in the detoxification of ROS.

Thioredoxins are specific NADPH-dependent oxidoreductases involved in reduction of disulfide bridges[237]. *TRX1* encodes the major thioredoxin of *C. albicans* involved in the response to reactive oxygen species[238]. Cells lacking *TRX1* have impaired resistance to hydrogen peroxide, as well as significantly attenuated virulence in a murine model of systemic infection[238].

Although *Candida* species have developed efficient enzymatic mechanisms to counteract the ROS, other proteins have recently been identified that also protect these pathogens from a stress such as **heat-shock proteins** and **enolases**. It was recently determined that heat-shock proteins are induced during the oxidative stress response in *C. albicans*[239]. The same occurs for enolases in other *Candida* species, namely *C. glabrata* and *C. krusei*. In addition, it was shown that enolase and the heat-shock proteins Hsp90 and Hsp70 in *C. albicans* are the major immunodominant proteins detected in patients with invasive candidiasis[240,241]. These observations suggest that these immunomodulatory proteins are also a primary ROS-protecting factor in these species.

Furthermore, other mechanisms that allow *C. albicans* to rapidly adapt to ROS inside the phagosome have been identified. These include changes in the expression of proteins involved in different metabolic pathways such as the downregulation of the carbon metabolism, and the upregulation of lipid, fatty acid, glyoxylate, and tricarboxylic acid cycles, which indicates that yeast shifts to a starvation mode[242].

5.2.3. MAPK pathways and TFs involved in the response to oxidative stress in C. albicans

Mitogen-activated protein kinase pathways (MAPKs) are signal transduction pathways present in eukaryotic cells that are triggered upon external stimuli and allow cells to sense changes in the external environment. They comprise a conserved module of three protein kinases: the MAP kinase (MAPK), the MAP kinase kinase (MAPKK), and the MAP kinase kinase kinase (MAPKKK)[243-245], and additional upstream elements such as transmembrane sensors, scaffold proteins and downstream transcription factors.

In *C. albicans*, two MAPKs become activated under oxidative stress conditions: the Hog1 and the Mkc1 pathways[246,247] (Figure 18). The Hog1-mediated pathway is considered as a general stress response cascade, involved in morphogenesis and cell wall integrity, and is essential for the response to osmotic and oxidative stress (reviewed in[248]). The Mkc1-mediated pathway is mainly implicated in cell wall reconstruction and morphogenesis, but other stimuli have been reported to activate this cascade. It is activated under oxidative stress, however the response mediated by its activation and the functional consequences are still being investigated.

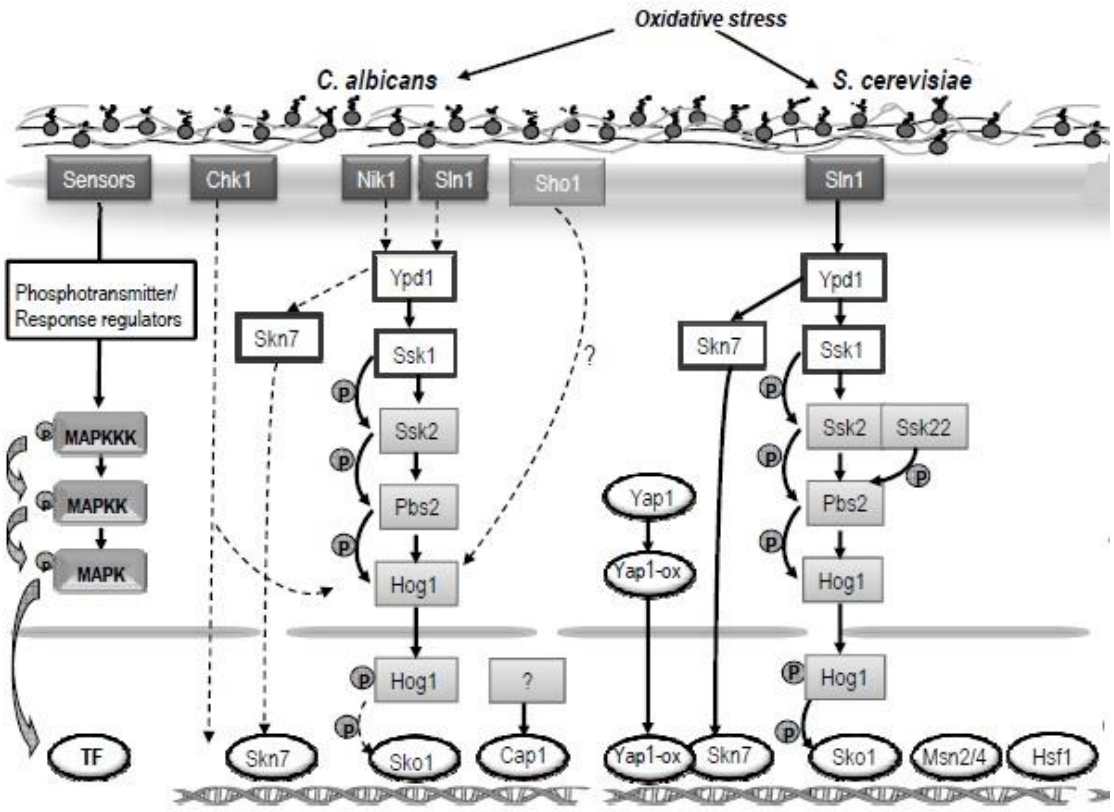


Figure 18. Structure of the HOG pathway in response to oxidative stress in *C. albicans* and *S. cerevisiae.* The stimuli leading to activation/repression of the pathways are shown either as stimulating (->) or inhibitory (--|) arrows. The solid arrows indicate direct relations or interactions and dotted lines show putative relations between the elements implicated. The phosphorylated elements are indicated with a P. (Adapted from de Dios *et al.* 2010[248])

Most of the elements in the *C. albicans* HOG pathway have been isolated by functional complementation of *S. cerevisiae* mutants of by sequence homology with this yeast model. Inactivation of Hog1 and key upstream regulators of this pathway have shown to confer oxidative stress sensitivity[238,246,249-252]. Oxidative stress signals appear to be transduced to Hog1 via the histidine kinases Sln1, Chk1[253], and Nik1. Upon activation of the pathway by autophosphorylation of its histidine residue, the phosphate is transferred to the Ssk1 response regulator[249,254] via Ypd1, a histidine-containing phosphotransfer protein (HPt)[255]. Ssk1 directly binds and regulates Ssk2 MAPKKK[256], which then phosphorylates the MAPKK Pbs2[257] that will transfer the phosphate to Hog1[258]. Phosphorylation of Hog1 provokes its translocation from the cytoplasm to the nucleus, where it then activates transcription factors that regulate the expression of genes involved in oxidative stress response.
Remarkably, *C. albicans hog1* mutants display an increased sensitivity to most of the agents producing reactive oxygen species such as menadione, hydrogen peroxide, potassium superoxide, t-butyl hydroperoxide, diamide and UV light[246,247,250]. This fact implies that Hog1 plays a crucial role in the response to oxidative stress independently of the nature of the oxidative agent involved.

The most studied transcription factor involved in the response to oxidative stress is Cap1, the orthologous gene of *S. cerevisiae* Yap1[259]. Cap1 contains redox-sensitive cysteine residues near its carboxy terminus that become oxidized following oxidative stress. This leads to the Hog1-independent nuclear accumulation of Cap1 and the activation of its target genes via Yap1-responsive elements (YRE) in their promoters[31,260,261]. Cap1 targets include genes involved in the detoxification of oxidative stress (e.g. catalase and superoxide dismutase *CAT1* and *SOD1*), glutathione synthesis (e.g. gamma-glutamylcysteine synthetase *GCS1*), redox homeostasis and oxidative damage repair (e.g. glutathione reductase and thioredoxin *GLR1* and *TRX1*). Together, these functions detoxify ROS and mediate cellular adaptation to stress. Consequently, the inactivation of Cap1 attenuates the induction of these genes, rendering *C. albicans* sensitive to oxidative stress[31,259]. The redox status of Cap1, and hence oxidative stress adaptation, is modulated by the redox regulator

thioredoxin (Trx1)[238]. Since it has been demonstrated that the nuclear accumulation of Cap1 is not dependent on Hog1 and most oxidative stress-induced transcripts are induced in a Hog1-independent fashion[31], the downstream molecular mechanisms that underlie Hog1-mediated oxidative stress resistance remain an area of active research.

Another transcription factor discovered to be involved in the oxidative stress response is Skn7p, which deletion mutant proved susceptible to treatment with H_2O_2[262]. An additional transcriptional regulator is Crr1, which contributes to oxidative stress resistance in *C. albicans*, but is not required for Hog1 activation in response to H_2O_2[263].

6. Correlation between protein and mRNA abundance in yeast

Yeast physiology constitutes an intricately coordinated system wherein an array of regulatory mechanisms allows response and adaptability to various internal and external stimuli. To do so, fundamental biological processes control the information flow from genes to proteins as their correspondent functional output. The total complement of mRNA within a cell at any given moment constitutes its transcriptome. The transcriptome forms the template for protein synthesis, resulting in the corresponding protein complement, known as the proteome. In this sense, global analyses of mRNA and protein abundance are referred to as transcriptomic and proteomic studies respectively, and comparative transcriptomic-proteomic studies are those that interrogate simultaneously the abundance of both. These studies are useful as they allow researchers to tease apart the abundance of a transcript from its actual use.

To date, several studies have reported simultaneous measurements of the concentrations of mRNAs and proteins from various organisms, including mammalian cells, worms, flies, yeast and a few species of bacteria[264-266]. In general, in both bacteria and eukaryotes, the cellular concentrations of proteins seem to correlate with the abundances of their corresponding mRNAs, but not strongly. They often show a squared Pearson correlation coefficient of approximately 0.40, which implies that only ≈40 % of the variation in protein concentration can be explained by correlating mRNA abundances[264,266]. The reason for the

remaining ~60 % of the variation still needs to be addressed[174]. This lack of strict correlation has provided clues for new research topics that are just now beginning to elucidate the underlying phenomena that explain this discordance.

7. Translation

7.1. Definition and importance

Despite the importance and the central role of RNA molecules in the flow of gene expression, proteins are the ultimate workhorses of the cell, carrying out all the functions necessary for life. For this reason, gene expression involves the conversion of genetic information, stored in the nucleus as DNA, into functionality, represented mainly as proteins. The process of protein synthesis is known as translation and takes place in the cytoplasm in eukaryotic cells. For translation to happen, protein-coding RNAs are exported from the nucleus to the cytoplasm where the translation machinery resides. This machinery consists of specialized particles called ribosomes, which are made mainly of RNA (rRNA) and other proteins that associate with it. During translation, mRNAs are decoded by ribosomes to produce a specific amino acid chain, or polypeptide, which later folds into an active protein to perform its functions in the cell. In eukaryotic cells, the ribosome is composed of two subunits: the large (60S) subunit and the small (40S) subunit. S, which stands for Svedberg unit, is a measure of sedimentation velocity and therefore, refers to the mass of each subunit. While each subunit exists separately in the cytoplasm, the two join together on the mRNA molecule to start the translation process. Translation is a cyclic process, where the small (40S) and large (60S) subunits first associate with mRNA to form the ribosomal complex or ribosome (80S). Ribosomes then move along the mRNA during elongation, to later dissociate again into the two subunits upon termination. Over the elongation period of one ribosome, further ribosomes can initiate translation on the same mRNA molecule, resulting in the formation of structures known as polysomes (Figure 19).

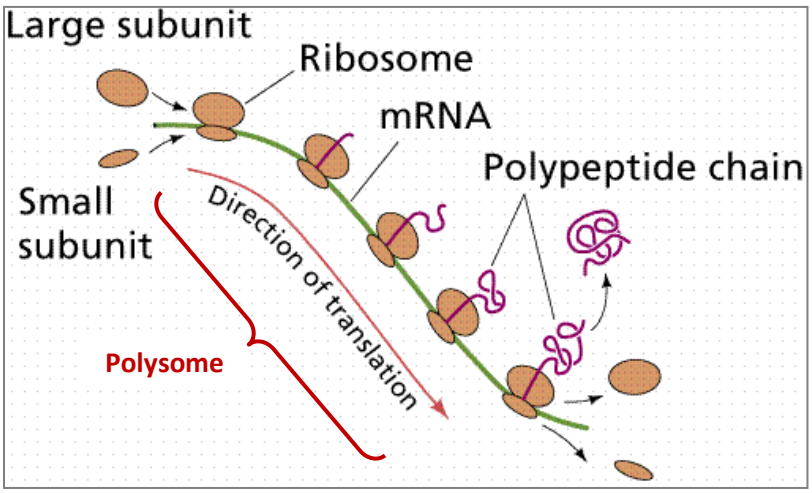


Figure 19. Polysome structures form during protein synthesis. During translational elongation, a ribosome moves along the mRNA synthesizing the new polypeptide chain to then dissociate into the 40S and 60S subunits upon termination. During this process, further ribosomes can initiate translation on the same mRNA, forming a structure known as "polysome".

In spite of being the goal of gene expression and performing most of the structural and catalytic functions in the cell, proteins are in general poorly regulated at the synthesis or degradation levels[267]. The reason for this is that protein synthesis is a lot more costly for the cell than regulating the amount of transcripts available for translation[268]. In this sense, majority of the changes the cell undergoes upon any environmental change mainly have an impact on the cell at the transcriptional level. Still, the factors that will ultimately govern and dictate changes in cell physiology will be the amount and the type of proteins available to make the change happen.

7.2. Global approaches for the study of translation

To date, the vast majority of studies devoted to monitor translation at a global scale have relied on two different methods, namely metabolic labeling followed by 2D gel electrophoresis and mass spectrometry[269-271] . In this work, three different methods have been used for the study of translation.

7.2.1. Two-dimensional gel electrophoresis (2DE)

To address global protein abundance within a cell, several high-throughput methods have been developed over the years (reviewed in [272]). The most common implementation for proteome analysis is two-dimensional gel electrophoresis (isoelectric focusing-SDS-polyacrylamide gel electrophoresis, 2DE). 2DE first permits the separation, visualization, and quantitation of thousands of proteins reproducibly on a gel[273,274]. Later on analytical techniques add the possibility of establishing the identities of the separated proteins (reviewed in [275]). Current protein analytical technology is based on the mass spectrometric generation of peptide fragment patterns that are idiotypic for the sequence of a protein. Protein identity is then established by correlating such fragment patterns with sequence databases[276]. In parallel, next-generation advances now allow for routine large-scale quantification of RNA abundances in any organism (discussed in section 3 of this work). By combining transcriptomic and proteomic approaches, quantitative protein and mRNA expression measurements of selected genes can be correlated. In this study, 2DE was used to address global protein abundance upon oxidative stress, described in Chapter 1 of the results section.

7.2.2. Polysome profiling

Whereas polysomes undergo active translation resulting in protein synthesis, mRNAs that are associated with a single ribosome (also called monosomes) are known to be translationally inactive. For this reason, the ribosome load on a given transcript becomes an indicator of how much it is being translated[277]. Each polysomal complex can contain from two to more than ten ribosomes, being this reflected in its total mass. This population of polysomes within a cell can then be sized-fractionated based on the ribosome load on the mRNAs it contains. This can be carried out using sucrose-gradient fractionation (also called polysome profiling), a technique that allows the separation of free ribonucleoprotein particles (ribosome-free mRNA) from mRNAs bound to an increasing number of ribosomes (polysome-bound mRNA). To do so, polysomes are isolated from the cells and loaded to a density gradient which is prepared with increasing concentrations of sucrose. Upon centrifugation of the gradient, polysomes separate all along the gradient fractions based on their individual density. By collecting the separated fractions individually and isolating the specific population of mRNAs that are bound to each of them, highly-translated mRNAs can be distinguished from poorly translated mRNAs (Figure 20).

The ribosome load on a population of mRNAs depends primarily on the rates of translation initiation and elongation. Both processes are strongly influenced by a large number of factors, such as the availability of ribosomal units, charged tRNAs, energy in the form of ATP and GTP and active translation factors. As a result, the translational process is very sensitive to the physiological status of the cell, and thus polysome profiles serve as an indicator of cell growth and overall cell status. For example, the ratio of polysomes to monosomes decreases markedly as cells enter late exponential growth phase, or during the initial phase of a stress response[278-280]. In this sense, polysome profiling can be used as a tool to address translational changes upon environmental stresses. Indeed, this has been demonstrated in model systems as diverse as the response to Ras and AKT signaling[281], dehydration of *Arabidopsis*[282], mating pheromone in yeast[283], inflammatory signals in neutrophils[284], and

analysis of breast cancer cell lines[285]. In this work, polysome profiling was used to validate the proteomic results obtained upon subjecting *C. albicans* cells to oxidative stress.

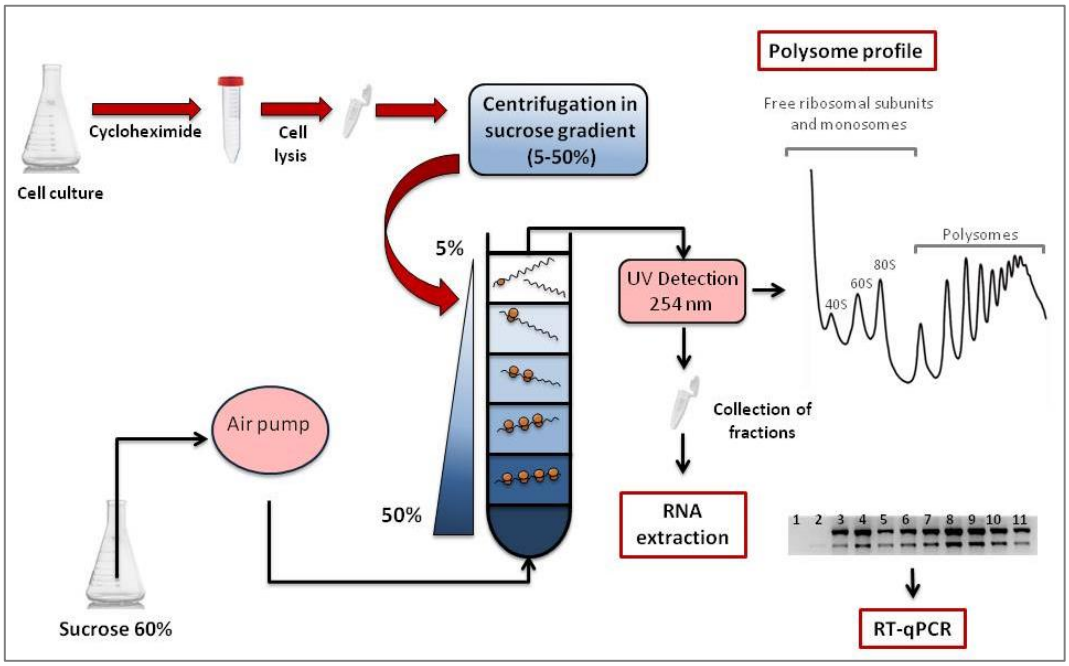


Figure 20. Schematic pipeline of a polysome fractionation experiment. For a polysome profiling experiment, translation in exponentially growing cells is stopped by the addition of the antibiotic cycloheximide to the culture, followed by cell lysis and density-based fractionation of the ribonucleoprotein particles in a sucrose gradient. The RNA present in each fraction is detected through A_{260} nm measurements, which abundance in each fraction is reflected in the so called "polysome profile" (Top right). The detected polysome-bound mRNA can be later on isolated and used for downstream purposes, such as RT-qPCR analyses.

7.2.3. Metabolic labelling: In vivo *^{35}S-Methionine incorporation*

Metabolic labelling techniques are used to study the biosynthesis, processing, intracellular transport, secretion, degradation, and physicochemical properties of proteins[286]. These techniques are based on the addition of a pulse of a labelled amino acid (e.g. ^{35}S-Methionine or ^{35}S-Cysteine). As the cells grow in the presence of the labelled amino acid, this is incorporated into the proteins that are being synthesized during the incubation time. The ultimate aim of this assay is to obtain a quantitative *in vivo* measurement of total protein synthesis, which is measured as the amount of radioactivity detected in proteins after the pulse is given ("*Incorporated*"), in respect to the total amount of radioactivity added to the

cells ("*Total*"). To do so, after a short incubation in the presence of the labelled amino acid, all protein content is precipitated using Trichloroacetic acid (TCA), followed by the measurement of the amount of radioactivity in a scintillation counter[287] (Figure 21) . In this study, metabolic labelling using ^{35}S-Methionine was used to address global protein synthesis at different growth temperatures in *S. cerevisiae*, described in Chapter 3 of the results section.

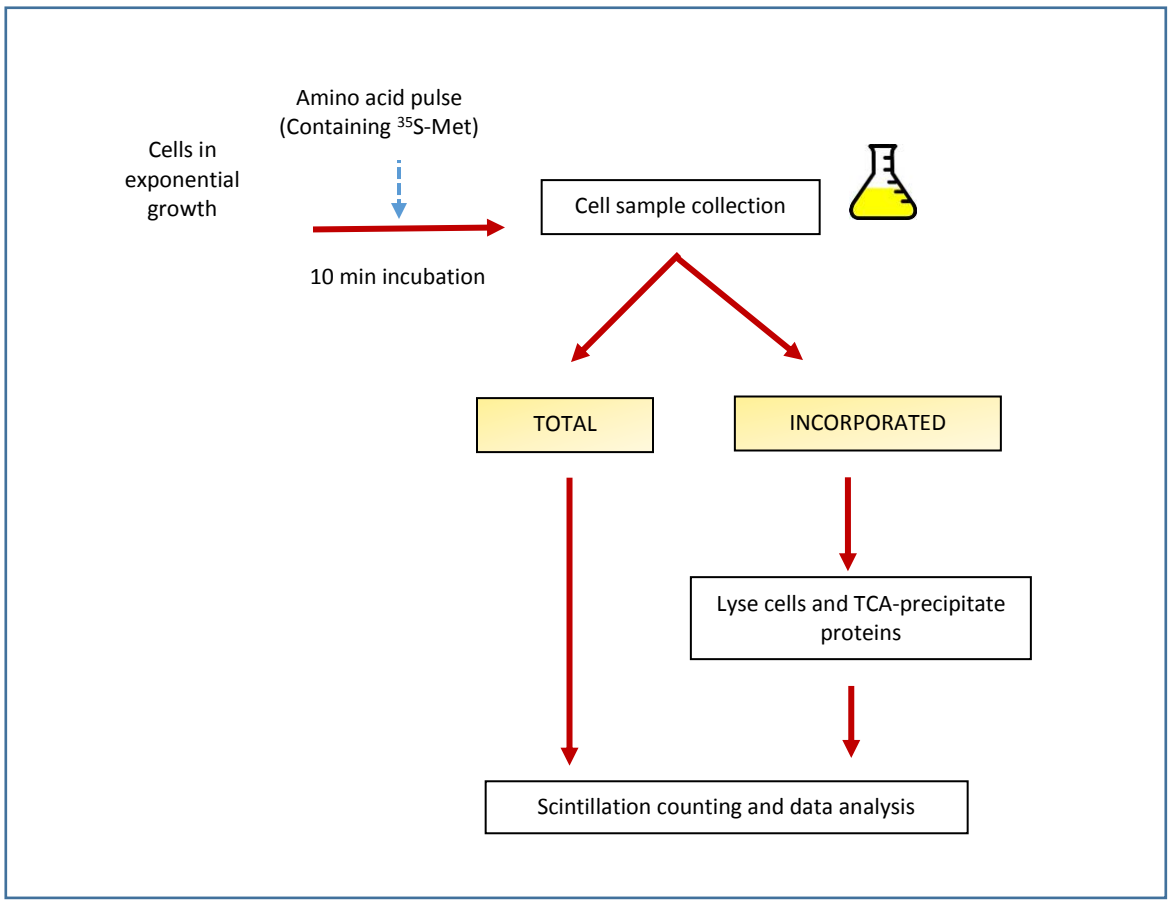


Figure 21. Schematic pipeline of a metabolic labelling experiment using ^{35}S-Methionine. Cells in exponential growth are pulsed with radioactively labelled Methionine (^{35}S-Met), and incubated for 10 min to allow the incorporation of the amino acid into newly synthesized proteins. After the incubation, 2 cell samples are collected, one to account for the total amount of radioactivity added to the culture and one to account for the radioactivity incorporated into proteins. The "Total" sample is directly dried onto glass fiber paper. For the "Incorporated" sample cells are first lysed, and then proteins are precipitated with TCA onto a glass fiber paper that is then thoroughly washed to discard any remaining of cell debris. Both samples are then subjected to scintillation counting for data analysis

Objectives

Despite its clinical relevance and compared to other model organisms, basic aspects of the physiology of the pathogenic yeast *Candida albicans* still remain elusive. Biological processes still poorly characterized in this organism include the existence of antisense transcription or other ncRNAs involved in the response to stress, or the general dynamics of its nascentome. Other unexplored fields include the relative contributions of transcription and translation to a stress response, or the impact of changes in growth temperature on its gene expression parameters.

The main aim of this research was to better characterize the transcriptional and translational response of *C. albicans* under two environmental stresses that it faces during infection within the host (hypoxic and oxidative), as well as to implement tools for the study of its nascentome. To do so, the model species *Saccharomyces cerevisiae* was utilized for the implementation of protocols.

Specific aims to achieve were as follows:

1. Determine the relative contribution of transcription and translation to the response of *C. albicans* to oxidative stress.

2. Develop a bioinformatic tool that allows the strand-specific detection of differentially expressed ncRNAs in the *C. albicans* transcriptome under hypoxic and oxidative stress.

3. Implement the protocol of BioGRO, a technique used for the high-resolution study of the nascentome, for its use in *C. albicans*.

4. Study changes in transcription and translation occurring during growth at different physiological temperatures in *S. cerevisiae*.

Materials and methods

1. Materials of especial relevance

1.1. Antibodies

Streptavidin-HRP (Pierce). Streptavidin coupled to the enzyme peroxidase (Horseradish peroxidase), utilized for the purification and selective detection of RNA molecules containing biotinilated uridines.

1.2. Transcriptional drugs

6-Azaurazil (6-Aza-2,4-dihydroxypyrimidine, 6-AU) (≥98%, Sigma). 6-AU 6 is an inhibitor of enzymes that are involved in purine and pyrimidine biosynthesis, which leads to alterations in nucleotide pool levels *in vivo*. Subsequently, the depletion of nucleotide levels by 6-azauracil can diminish transcription elongation.

Thiolutin (Pfizer). Thiolutin is a sulfur-containing antibiotic that serves as an inhibitor of bacterial and yeast RNA polymerases. *In vitro*, it inhibits the RNA synthesis directed by all three yeast RNA polymerases. In this work, Thiolutin was used for the transcriptional shut-off done to determine mRNA half-lives.

1.3. Translational drug

Cycloheximide (Sigma). Cycloheximide is an antibiotic that is used for the inhibition of translation elongation in eukaryotes, resulting in cell growth arrest and cell death. In this study cycloheximide was used to stop translation at a specific time point during polysome profiling assays.

1.4. Modified nucleotide

Bio-11-UTP (Biotin-11-Uridine-5'-triphosphate) (Ambion). Biotinylated UTP is used as a substrate in *in vitro* transcription reactions with a variety of RNA polymerases. In this work it has been used for RNA labeling instead of radioactively labeling of RNA. Biotin-bound streptavidin can be detected directly, using fluorescently labeled streptavidin, or indirectly, using a streptavidin-conjugated enzyme assays.

1.5. Radioactive Methionine

^{35}S-Methionine (175 Ci/mmol, 11mCi/mL; 8.4 µM) (Perkin Elmer) Radioactive amino acid used for the evaluation of protein biosynthesis in live yeasts cells. This amino acid is transported into the cell and then used by the translational machinery during protein synthesis. TCA-precipitated proteins are then evaluated for the measurement of *in vivo* translation activity[288].

1.6. Others

- Sarkosyl (N-laurosyl-sarcosine). Detergent utilized for making the yeast cell membrane permeable. This detergent also disrupts the chromatin structure and blocks transcription re-initiation by RNA polymerases while allowing a first round of initiations from preassembled initiation complexes.

- Glycogen (molecular biology grade, Roche). Polysaccharide. Given its insoluble nature in ethanol and isopropanol, the addition of glycogen to RNA samples forms a white aggregate that pulls down the nucleic acids increasing the yield of the precipitation step, as well as helping the visualization of the RNA pellet, without contaminating the RNA sample with another nucleic acid (such as *E. coli* tRNA).

2. Microbiology techniques

2.1. Yeast strains used in this work

Microorganism	Strain	Genotype	Source
C. albicans	SC5314	Wild- type (Clinical isolate)	Jesús Pla (UCM)
S. cerevisiae	BY4741	*Mat a; his3Δ; leu2Δ ; met15Δ ; ura3Δ*	Euroscarf
	GYLR-4B	*Mat α trp1Δ::HISG URA::pGAL-YLR454W, his3Δ1, ura3Δ, lys2Δ, met15Δ* (Background BY4741)	Paco Malagón (Henry M. Jackson Foundation, USA)
	imd2Δ	*Imd2Δ::HphMX4* (Background BY4741)	Sebastián Chávez (US)
	BQS252	*Mat a; ura3Δ* (Background BY4741)	Our laboratory

2.2. Media and growth conditions

- For the transcriptomic analyses under hypoxic and oxidative stress in *C. albicans* (described in chapters 1 and 2, *C. albicans* SC5314 strain was grown in liquid YPD medium (1% yeast extract, 2% peptone, 2% glucose) at 37 °C. For hypoxia experiments, cells were grown in normoxia up to OD_{600} of 0.5. For normoxia control (time 0) a 50 mL aliquot was collected. For hypoxic treatment of the cell, a volume of cells containing 0.45 $O.D._{600}$ units was collected, supernatant was discarded, and cell pellet was transferred to a hypoxic chamber (*InvivO2 200 hypoxic chamber*, Baker Ruskim) in a new flask containing fresh YPD medium that had been previously equilibrated O/N (O/N) in hypoxia (0.2% O_2) the day before the experiment. From this point cells were kept growing in the hypoxic chamber at 37 °C with agitation for a total time of 3h. For oxidative stress treatment, cells were grown in the same conditions until $O.D._{600}$ 0.5. At this time point, a 50 mL aliquot was collected for the no stress control (time 0). In parallel, 100 mL of cells were transferred to a new flask containing tert-butyl hydroperoxide solution (t-BOOH 70% v/v, Sigma) to a final concentration 0.1 mM.

From this point cells were kept growing in the bath at 37 °C with agitation for a total time of 3h.

- For all experiments carried out for the optimization of the GRO technique described in Chapter 3 both *S. cerevisiae* BY4741 and *C. albicans* SC5314 cells were grown in liquid YPD medium, at 30 and 37 °C respectively unless stated otherwise.

- For all experiments concerning RNA pol II density measurements at different grow temperatures described in Chapter 4 *S. cerevisiae* GYLR-4B cells were grown in liquid YPD medium at 23, 30, 34 and 37 °C respectively. In all cases, cells were grown for at least seven generations at the correspondent growth temperature before performance of the assay.

- To test the effect of 6-azauracil (6AU) on RNA pol II density described in chapter 4, *S. cerevisiae imd2Δ* cells were grown at 30 °C in synthetic complete medium lacking uracil (SC-ura, yeast nitrogen base without amino acids 0.67%, 2% glucose), up to O.D. $_{600}$ 0.5. At this point, 6-AU (Sigma) in solution was added to the culture to a final concentration of 50 µg/mL.

- For all experiments performed to asses translation activity at different growth temperatures described in chapter 4 *S. cerevisiae* BQS252 cells were grown in liquid synthetic minimal medium containing uracil (SD+ura; yeast nitrogen base without amino acids 0.67%, 2% glucose, 20 µg/mL uracil) up to O.D. $_{600}$ 0.3 at 23, 30, 34 and 37 °C respectively. In all cases, cells were grown for at least seven generations at the correspondent growth temperature before performance of the assay.

3. Molecular biology techniques

3.1. Isolation of nucleic acids

3.1.1. DNA extraction from yeast cultures

Extraction of genomic DNA was needed for it to be used as a template for positive controls in PCR reactions. Total DNA was extracted from yeast cells using phenol:chloroform extraction. Cells were grown in YPD up to $O.D._{600}$ 0.5, and harvested by centrifugation at 4000 rpm for 3 min. Cell pellet was resuspended in 500 µL of *10 prep* solution (2% Triton X-100, 1% SDS, 0.1M NaCl, 1mM EDTA, 10mM Tris-HCl pH 8) and transferred to an 1.5 mL eppendorf tube containing 1 volume of phenol-chloroform (1:1) and 1 volume of glass beads (0.5 mm, Sartorius). Cells were broken in *Fast-Prep* device (MP-Biomedicals) in 3 rounds of 30 seconds each, 5.5 m/s. Cell debris was pelleted at 13000 rpm for 5 min in a table microcentrifuge. Supernatant was extracted with 1V of chloroform: isoamyl alcohol (24:1). Aqueous phase was precipitated by the addition of 0.1V of 3M NaAc and 2.5 volumes of cold EtOH 96 %, and stored at - 20 ^{o}C O/N. DNA was pelleted by centrifugation at 13000 rpm for 15 min, washed with EtOH 70 % (v/v), dried in Savant SPD111V *SpeedVac concentrator* (medium drying rate) for 2 min, and resuspended in 400 µL of TE buffer (10 mM Tris pH 8, 1 mM EDTA). So as to remove the contaminant RNA from the sample, 3 µL of 10 mg/mL DNase-free RNase A (Roche) were added to the sample, followed by incubation at 37 ^{o}C for 30 min. To remove the RNase A DNA was treated with 10 µL of 10 mg/mL Proteinase K solution (Roche), followed by incubation at 65 ^{o}C for 1h. DNA was precipitated by the addition of 0.1 volume of 3M NaAc and 2.5 volume of cold EtOH 96 %, and stored at - 20 ^{o}C O/N. DNA was pelleted by centrifugation at 13000 rpm for 15 min, washed with EtOH 70 % (v/v), dried in *SpeedVac concentrator* (medium drying rate) for 2 min, and resuspended in 1mL of TE buffer. DNA amount was quantified from 1 µL using *Nanodrop 2000 Spectrophotometer* (Thermo Scientific).

3.1.2. RNA extraction from yeast cultures

Total RNA was extracted from yeast cells using phenol:chloroform extraction. Cell pellet was resuspended in 500 µL of LETS (100 mM LiCl,10 mM EDTA,10 mM Tris-HCl pH 7.5,0.2% SDS) and transferred to an eppendorf tube containing 1 volume of acid phenol-chloroform (5:1) and 1 volume of glass beads (0.5 mm, Sartorius). Cells were broken in a *Fast-Prep* device in 3 rounds of 30 s each, 5.5 m/s. Cell debris was pelleted at 13000 rpm for 5 min in a table micro centrifuge. Supernatant was extracted with 1 volume of chloroform: isoamyl alcohol (24:1). Aqueous phase was precipitated by the addition of 0.1 volumes of 5M LiCl and 2.5 volumes of cold EtOH 96 %, and stored at - 20 ^{o}C O/N. RNA was pelleted by centrifugation at 15700 rcf for 15 min, washed with EtOH 70 % (v/v), dried in *SpeedVac concentrator* (medium drying rate) for 2 min, and resuspended in the desired volume of RNase free water. RNA amount was quantified using *Nanodrop 2000 Spectrophotometer*.

3.1.3. Specific isolation of Biotinylated RNA molecules

Specific isolation of biotinylated RNA was needed for BioGRO-seq experiments. After total RNA extraction (see previous protocol) the total amount of extracted RNA was bound to 100 µL of Streptavidin M-280 magnetic beads (Invitrogen), following the manufacturer's instructions. For elution of the biotinylated RNA molecules, the beads were washed once in high salt buffer (2 M NaCl, 50 mM Tris-HCl pH 7.4, 0.5% Triton X-100), once in Medium salt buffer (300 mM NaCl, 10 mM Tris-HCl pH 7.4, 0.1% Triton X-100), and once in Low salt buffer (5 mM Tris- HCl pH 7.4, 0.1% Triton X-100). After washing, beads were resuspended in 300 µL of RNase-free water and bound RNA was extracted from the beads by the addition of a mixture of 300 µL of *Trizol reagent* (Invitrogen) and 60 µL of chloroform: isoamyl alcohol (24:1), followed by vigorous shaking for 15s and incubation at RT for 3 min. To allow phase separation, the sample was centrifuged at 13000 rpm for 10 min at 4 ^{o}C. To remove Trizol

contamination RNA was extracted with 1 volume of chloroform: isoamyl alcohol (24:1), followed by addition of 0.1V of LiCl 5M and 750 µL of 96% Ethanol and O/N RNA precipitation at -20 ^{o}C. RNA was centrifuged at 13000 rpm for 15 min at 4 ^{o}C, washed once with 70% EtOH, dried in *SpeedVac concentrator* for 2 min at medium drying rate, and dissolved in 15 µL of RNase-free water.

3.1.4. Poly (A) RNA purification

For genome-wide analyses of mRNA abundance upon hypoxia and oxidative stress, total RNA was extracted from the cell using phenol-chloroform extraction as described in 3.1.2., followed by poly (A) RNA purification. Poly (A) purification was performed using *Dynabeads mRNA purification kit* (Invitrogen), which isolation protocol relies on base pairing between the poly (A) residues at the 3' end of most mRNA and the oligo $(dT)_{25}$ residues covalently coupled to the surface of the Dynabeads. Using the manufacturer's instructions, an average final amount of 1.5 µg of poly (A) RNA was usually recovered from the starting 75 µg of total RNA .

3.1.5. Isolation of polysome-bound mRNA

For the specific isolation of mRNA bound to polysomal fractions, different fractions were pooled. As depicted the A_{260nm} profile display in Figure 25, fractions containing ribosomal 40S subunits, 60S subunits, and mRNAs attached to just one ribosome (monosomes) were pooled in one defined as 'M'. Fractions containing mRNAs with 2 or 3 ribosomes attached were pooled in one defined as 'P1'; fractions containing mRNAs with 4, 5, 6 or 7 ribosomes attached were pooled in one defined as 'P2'; and fractions containing mRNA with 8 or more ribosomes attached were pooled in one defined as 'P3'. To make each pool, 100 µL per fraction was taken. For normalization of RNA extraction efficiency, 3 ng of an *in vitro* transcript of the *Bacillus subtilis* lysine gene (denoted as 'LYS' in this study) was added to each pool as a spike-in control. RNA extraction was performed using *Speed tools total RNA extraction kit* (Biotools), following the manufacturer's instructions and omitting the steps of cell lysis and lysate filtration. RNA was eluted in a final volume of 50 µL of water, and RNA

integrity was checked by running 4 µL of the eluate in a 1% agarose gel in 1X TAE buffer as described in 3.2.

3.2. Electrophoretic separation of nucleic acids

For fragments size determination and sorting, nucleic acids were run in horizontal agarose gels in TAE buffer (40 mM Tris-Acetate, 1mM EDTA). The concentration of agarose utilized was 1% and 2% for DNA and RNA gels respectively, and all of them contained 5 µg/mL EtBr. To visualize the nucleic acids gels were exposed to a UV light source in *Molecular Imager® Gel Doc XR$^+$ imaging system* (BioRad).

3.3. Polymerase chain reaction (PCR)

3.3.1. Quantitative PCR (qPCR)

To validate the transcriptomic and the proteomic data obtained in *C. albicans* cells upon oxidative stress described in chapter 1, the relative mRNA amount was estimated using quantitative PCR (qPCR). All qPCR experiments were carried out in a DNA engine thermal cycler (BioRad) and analyzed using the Opticon Monitor software (BioRad). For amplification, SYBR Premix Ex Taq™ II -Tli RNase H Plus (Takara) was used, in a mixture containing 5 µL of SYBR Premix, 0.2 µL of each primer, 2.5 µL of cDNA and 0.2 µL of water. Cycling conditions were as follows:

1) 95 ^{o}C for 10 min.
2) 95 ^{o}C for 10 s.
3) 55 ^{o}C for 20 s.
4) Go to 2 and repeat 40 times.
5) Ramp from 65 to 95 ^{o}C in increments of 0.5 s.

3.3.2. Primer sequences used for qPCR experiments

All primer sequences described belong to the *C. albicans* genome.

Name	Sequence	Target
ACT1-F	5'-TTGCTCCAGAAGAACATCCA-3'	*ACT1* ORF
ACT1-R	5'-CACCATCACCAGAATCCAAA-3'	
ARD1-F	5'-ACGGGTTGGGGTCATTCTAC-3'	*ARD1* ORF
ARD1-R	5'-CATTGTGGTTGTGGGTCGTT-3'	
BHM1-F	5'-CGATGATGCTGTTGCTGATT-3'	*BHM1* ORF
BHM1-R	5'-GCCTTACCTTCTGTTGGTTGA-3'	
CHT3	5'-TATTGGGGACAAAACTCTGG-3'	*CHT3* ORF
CHT3	5'-CAACAGCATCGGAATCACAA-3'	
CIP1-F	5'-ATGCTCCAGGGTTTTTAGGA-3'	*CIP1* ORF
CIP1-R	5'-CTCTTTCTTGGCGTCTTCTG-3'	
ENO1-F	5'-GTGACGAAGGTGGTGTTGCT-3'	*ENO1* ORF
ENO1-R	5'-TTAGATGGGTCGGATTCTGG-3'	
PRX1-F	5'-GGGCTATCTTGTTCTCACATCC-3'	*PRX1* ORF
PRX1-R	5'-ATTTCACCCCTCTCTTCGTG-3'	
SEC13-F	5'-GGAAGGAGCAACCAGAAACTC-3'	*SEC13* ORF
SEC13-R	5'-TTGAAATCCACCACTGAGACC-3'	
TSA1-F	5'-GCTGACACCAACCACTCCTT-3'	*TSA1* ORF
TSA1-R	5'-GGAAAGCCTCCAACAATCTC-3'	

3.4 Transcriptional shut-off

For the determination of the mRNA half-lives of the 6 genes used in chapter 4 of this work (namely *ACT1, RPL17, FHO88, RPL25, RPL5, RPB6*) in cells that have been growing at either 28 or 37 ^{o}C, a transcriptional shut-off with the transcriptional inhibitor Thiolutin was performed. To do so, BY4741 strain was grown in liquid YPD medium for, at least, 10 generations at the correspondent temperature until $O.D._{600}$ ~ 0.5. At that point, liquid Thiolutin solution was added to the culture to a final concentration of 5 µg/mL, and 20 mL culture aliquots were collected at time points 5, 12, 25, 45, 60, 90, and 120 min. RNA was extracted from the cells by Phenol-Chloroform extraction as described in 3.1.2, and subjected to northern blot as described in 3.5.3.

3.5 Northern blot

3.5.1. Northern blot for Biotinilated RNA detection

Northern blot was used to assess the proper incorporation of Biotin-11-UTP during BioGRO assays, as well as for estimation of the average size of biotinilated RNA. To do so, RNA is size-separated in an agarose gel which is then transferred to a nylon membrane, followed by incubation with an antibody coupled with Streptavidin. For this purpose, purified RNA was run in a 2% agarose gel in 1X TAE buffer (40 mM Tris-acetate, 1mM EDTA, pH 8), and transferred to a Nylon membrane (Hybond XL, GE Healthcare) by capillarity using 6X SSC (0.9 NaCl, 90 mM sodium citrate). RNA was cross-linked to the nylon membrane by irradiation with 50 mJ of 254 nm UV light in a GS Gene Linker chamber (BioRad). For blocking and incubation with the antibody, The membrane was incubated in agitation for 20 min in Blocking buffer (125 mM NaCl, 9 mM Na2HPO4, 7 mM NaH2PO4, 10% SDS), followed by incubation for 10 minutes with 5 mg/mL Immunopure Streptavidin – HRP (Pierce) in Blocking buffer (Antibody dilution 1:4000). For washing, the membrane was incubated 2-fold in blocking buffer for 10, followed by two washes by incubation with washing buffer (100 mM Tris-HCl, 100 mM NaCl, 10.5 mM MgCl2, pH 9.5) for 5 min each. For staining, the membrane was incubated with ECL (GE-healthcare) for 1 min and the signal was developed using *ImageQuant LAS-4000 mini* luminescent image analyzer.

3.5.2. Dot blot for Poly (A) RNA estimation

For mRNA quantity determination at different cell growth temperatures, total RNA was extracted as described in 3.1.2. from cells that had been growing at the correspondent temperature for, at least 7 generations. Extractions were performed in biological triplicates. For each replicate, total RNA was spotted in a Nylon membrane (Nytran SPC, GE Healthcare) using a *BioGrid* robot (BioRobotics). For spotting, RNA samples were taken by the robot arm from a 384 well plate containing 3 different dilutions of each RNA sample, namely 250, 125 and 75 ng/µL. Once in the well these concentrations were checked by quantifying the spotted sample using *NanoDrop*. After sample spotting, the membrane was cross-linked in

UV light using the C-L program of a *GS GeneLinker* (BioRad) and hybridized with specific oligo $d(T)_{40}$ (Integrated DNA Technologies) probes labelled terminally with γ^{32}P-ATP (Perkin Elmer) using Polynucleotide Kinase (Roche) following the manufacturer's instructions. Membranes were hybridized at 42 ^{o}C for 24h in hybridization solution (Phosphate buffer 0.5M pH 7.2, EDTA 1mM, SDS 7%), and washed once with washing solution 1 (1X SSC, 0.1% SDS) for 10 min, followed by 1 wash with washing solution 2 (0.5X SSC, 0.1% SDS) for 10 min. All washes were done at 42 °C. After washing, membranes were sealed in plastic film and exposed to an Imaging plate (BAS-MP, Fujifilm) for 5-7 days (depending on cps value recorded with a Geiger counter; 10 cps ~ 5 days exposure), and scanned with Fujifilm FLA3000TM Phosphorimager. Signal intensity of the spots was quantified with Image Gauge 4.0.

3.5.3. Northern blot for mRNA half-life determination

For mRNA half-life determination of the 6 individual genes used in chapter 4, 10 µg of total RNA from each time point was run in a 1% agarose denaturing gel in the presence of 2M formaldehyde (Panreac), and transferred to a Nylon membrane (Immobilon-NY+, Millipore). The membrane was hybridized to the specific gene probes, labelled with *Ready-to-go DNA labeling beads (-dCTP) kit* (GE Healthcare). Membranes were exposed to an Imaging plate (BAS-MP, Fujifilm) and scanned with Fujifilm FLA3000 Phosphorimager. Signal intensity of the spots was quantified with *Image Gauge 4.0* software. Intensity values were normalized against the loading control.

3.6. Removal of DNA from RNA samples

For cDNA conversion, total RNA was treated with RNase-free DNase I prior to reverse transcription. This step aims to deplete the DNA molecules that are present in the sample. For DNase treatment, 2.5 µg of DNA/RNA was digested with 1U of RNase-free DNase I (Roche), by incubating at 25 ^{o}C for 15 min. To inactivate the DNase, 0.62 µL of 5 nM EDTA was added, followed by incubation at 65 ^{o}C for 10 min. Seven µL of this DNA-free RNA was used for reverse transcription as described in 3.6.

3.7. Reverse transcription (RT)

For cDNA conversion, RNA was reverse-transcribed using oligo $d(T)_{25}$ primers and *Maxima reverse transcriptase* (Thermo Scientific). To do so, 7 µL of DNA-free RNA was mixed with 1.6 µL of 10 mM dNTPs,1 µL of 10 µM oligo d(T) VN (IDT DNA technologies), and 2.4 µL of water to a final volume of 12 µL. The mixture was incubated for 10 min at 65 ^{0}C. Then, 4 µL of 5X Reverse transcriptase reaction buffer and 2 µL of 0.1M DTT was added, followed by incubation at 42 ^{0}C for 2 min. Finally, 0.5 µL of 200 U/ µL *Maxima reverse transcriptase* (20 U) was added and the mix was incubated at 42 ^{0}C for 50 min, followed by incubation at 70 ^{o}C for 15 min to inactivate the enzyme. All incubation steps were carried out in a *Thermomixer* (Eppendorf) at 650 rpm agitation.

3.8. Polysome fractionation from yeast cells

For polysome fractioning from *C. albicans* cells upon oxidative stress (described in chapter 1), cells were grown and treated with t-BOOH as described in 2.2. For each time point, 100 mL of $O.D_{600}$ 0.5 culture was collected, and translation was stopped by the addition of 0.1 mg/mL of cycloheximide (Sigma) followed by incubation on ice for 5 min, with occasional mixing by inversion. Cells were collected by centrifugation at 4500 rpm at 4 ^{o}C for 3 min. Supernatant was discarded and cell pellet was resuspended in 2 mL of lysis buffer (0.02 M Tris-HCl pH 8, 0.14 M KCl, 5 mM $MgCl_2$, 0.5 mM DTT, 0.1 mg/ml cycloheximide, 0.5 mg/ml heparin (Sigma), 1 % Triton X-100) and centrifuged at 4000 rpm at 4 ^{o}C for 5 min. Cell pellet was resuspended again in 700 µL of lysis buffer, and transferred into a 2 mL tube containing 1 volume of glass beads (0.5 mm, Sartorius). Cells were disrupted by vortexing the tube 8-fold for 30 s each, with 30 s of incubation on ice between vortexing cycles. Cell extract was centrifuged at 5000 rpm for 5 min at 4 ^{o}C. Supernatant was transferred to a new eppendorf tube, and centrifuged at 8000 rpm for 5 min at 4 ^{o}C. Supernatant was transferred to a new eppendorf tube and RNA content was measured using *Nanodrop 2000 Spectrophotometer*, using lysis buffer for the blank. For freezing, glycerol was added to the sample to a final concentration of 5%, and sample was frozen in liquid N_2 and stored at -80 ^{o}C. For separation

of polysomes based on ribosome load, 10 U of A_{260} were loaded in a 5%- 50% sucrose gradient, and centrifuged at 4 ^{o}C for 2 h 40 min at 35000 rpm in a L-70 ultracentrifuge (Beckman coulter) using a SW41 rotor. Gradients were then fractionated using isotonic pumping of 60% sucrose from the bottom. Polysomal profiles were registered via detection of A_{260} nm using a *Density Gradient Fractionation System* (Teledyne Isco, Lincoln, NE).

3.8.1. Data normalization in polysome-bound mRNA qPCR experiments

Relative mRNA abundances listed in Table 6 obtained by means of qPCR of polysome-bound mRNA were normalized by the expression value of an internal RNA extraction control (*Bacillus subtilis* LYS gene *in vitro* transcript), volume correction factor and ribosome load. The volume correction factor was determined as the total volume per pool divided by the volume taken from each pool for RNA extraction (200 µL). The ribosome load correction factor was applied to account for the number of ribosomes attached per mRNA molecule in each different fraction, as determined visually from the polysome profile shown in Figure 25, as follows:

Fraction	Number of ribosomes/mRNA	Correction factor
M	1	1 (=1/1)
P1	6 (2+3)	2.5 (= 6/2)
P2	15 (4+5+6)	5 (=15/3)
P3	34 (7+8+9+10)	8.5 (=34/4)

To account for the total amount of each mRNA present in the samples before and after the treatment (t0 and t40 min), the final value of the fold change t40/t0 for each polysome-bound mRNA was multiplied by the fold change t40/t0 of total mRNA (obtained from the transcriptomic data).

Fold change t40/t0 polysome-bound mRNA (Translation rate per individual mRNA molecule, **TLRi**)	X	Fold change t40/t0 total mRNA	=	Total amount of mRNA in translation (**TLR**)

3.9. Methionine incorporation assay for translation rate measurement

For the evaluation of translation activity in *S. cerevisiae* BQS252, cells were grown in SC+ura medium to an $O.D_{600}$ between 0.3 and 0.4 at 23, 30, 34 or 37 ^{o}C. At that time 2 ml of culture was collected (small differences between cultures in the same experiment were corrected to get the same number of cells) and transferred to a 2 mL eppendorf tube and rapidly incubated for 10 min in a thermomixer set at the same temperature of the culture in order to avoid any temperature change during the process. Eleven µL of a 1/10 dilution of ^{35}S-Methionine stock solution (Perkin Elmer) were added to the culture and incubated for 10 min with agitation at 650 rpm. The incorporation reaction was stopped by submerging the tube in an ethanol-ice bath in order to lower the temperature below 4 ^{o}C as fast as possible. The total intake of radioactivity by the cells was determined by spotting three aliquots (technical replicates) of 330 µL of cell suspension into a 2.5 cm glass fiber filter that were washed three times with cold culture medium in order to wash out external radioactive methionine. After washing glass filters were dried in an aerated heater and submerged in 5 mL of scintillation solution for radioactive counting.

The incorporation of methionine to proteins was determined as the amount of TCA-precipitable radioactivity. To do so 3 aliquots (technical replicates) of 330 µL of cell suspension were transferred to a new 2 mL eppendorf tube containing 500 µL of 0.5 mm glass beads and disrupted mechanically over 3 steps of 30 s at 5.5 m/s in a Fast-Prep device at room temperature. The liquid phase was separated from the beads and transferred to a new 1.5 mL eppendorf by making a small hole at the bottom of the sample tube and centrifuging into the second tube during 30 s at 2000 rpm. Then 33 µL of 5 M NaOH was

added to the liquid and incubated for 10 min at 25 ^{o}C in order to hydrolyze tRNA-Met molecules that could account for the final ^{35}S- Met measurement. After this period 366 µL of ice-cold 25% TCA were added for protein precipitation and the tubes were kept on ice for 10 min. Then, all the liquid content was spotted into a 2.5 cm glass fiber filter that was washed once with cold 10% TCA and once with cold 96% ethanol. Then glass filters were dried in and aerated heater and submerged in 5 mL of scintillation solution for radioactive counting.

4. Genomic techniques

4.1 Filter run-on (MiniGRO)

MiniGRO assays were used to evaluate the labeling efficiency in the run-on experiments showed in chapter 4. For these experiments, *S. cerevisiae* and *C. albicans* cells were grown in liquid YPD medium up to O.D.$_{600}$ ~ 0.5 at the indicated growth temperatures. At that point, a 3 mL aliquot was collected by centrifugation at 4000 rpm for 2 min.

- For control sample (*Blank*): Cell pellet was resuspended in 5 mL of distilled water and collected again. Pellet was resuspended in 1 mL of distilled water and transferred to an eppendorf tube.
- For experimental samples: Cell pellet was resuspended in 5 mL of 0.5 % Sarkosyl and collected again. Pellet was resuspended in 1 mL of Sarkosyl and transferred to an eppendorf tube.

Cells in eppendorf tubes were centrifuged at 6000 rpm for 1 min. Supernatant containing Sarkosyl was removed and pellet was resuspended in 7.2 µL of distilled water. Run-on pulse was performed by adding 9.87 µL of transcription mix [c] (Final volume ~ 18 µL). To allow transcription elongation, mix was incubated in agitation (650 rpm) for 5 min at 30 ^{o}C (For *S. cerevisiae*) or 37 ^{o}C (For *C. albicans*). Pulse was stopped by the addition of 82 µL of cold distilled water to the mix and storage on ice. For the measurement of the total amount of radioactivity present in the mix ("*Total*"), 10 µL of the reaction was directly spotted on the paper and dried in an aerated heater at 65 ^{o}C. To measure the percentage of radioactivity that had been incorporated to the nascent RNA chain, another 10 µL of the mix was spotted

in the glass fiber paper and dried in heater at 65 ^{o}C for 25 min, followed by nucleic acid precipitation performed in technical duplicates. For nucleic acid precipitation, glass fiber paper was soaked in 4 mL of 10 % (v/v) of Trichloroacetic acid (TCA) and incubated at 4 ^{o}C for 20 min. TCA was removed by decanting and 4 mL of cold TCA (10 % v/v) was added again, followed by incubation at 4 ^{o}C for 10 min. TCA was removed and glass fiber paper was washed with 3 mL of cold 70 % (v/v) EtOH, followed by washing with 3 mL of cold 96 % (v/v) EtOH. Glass fiber paper was dried in heater at 65 ^{o}C for 25 min. Once dried, 5 mL of scintillation liquid was added to each vial for radioactive counting. For each individual sample, the percentage of incorporation was calculated as:

[(Precipitated / Total) x 100] – *Blank*

Where "*Blank*" was calculated as [(Precipitated / Total) x 100] in the control sample (the one treated with water instead of Sarkosyl).

[c] Transcription mix

Per sample:

- 7.5 µL of 2.5X Transcription buffer (50 mM Tris-HCl pH 7.7, 500 mM KCl, 80 mM $MgCl_2$)
- 1 µL of 10 mM rNTPs (ATP, CTP, GTP, 10 mM each)
- 0.375 µL DTT (0.1 M)
- 0.66 µL UTP (3 µM)
- 0.34 µL [α-^{33}P] UTP (Perkin Elmer, 3000 Ci /mmol, 10 µCi /µl, 3 µM)

4.2 Genomic Run-on (GRO)

GRO protocol was adapted from García-Martínez *et al.* (2011). Yeast cells were grown at the required growth temperature[a] in liquid YPD medium up to O.D. $_{600}$ ~ 0.5. At this point, a 50 mL aliquot was harvested at 4000 rpm for 2 min. For cell permeabilization, cell pellet was resuspended in 10 mL of Sarkosyl 0.5 % and collected again. Pellet was resuspended in 1 mL of Sarkosyl 0.5 %, transferred to an eppendorf tube, harvested at 6000 rpm for 1 min, and resuspended in 115 µL of distilled water. Run-on pulse was performed by adding 162 µL of transcription mix [b] (Final reaction volume including cells ~ 300 µL). To allow transcription elongation the mixture was incubated at 30 ^{o}C for 5 min. Reaction was stopped by addition of 1 mL of cold distilled water to the mix and tube was immediately put on ice. Cells were collected by centrifugation at 3300 rcf for 1 min. For total RNA extraction, cell pellet was resuspended in 500 µL of LETS (100 mM LiCl,10 mM EDTA,10 mM Tris-HCl pH 7.5,0.2% SDS) and transferred to an eppendorf tube containing 1V of acid phenol-chloroform (5:1) and 1 vol of glass beads (0.5 mm, Sartorius). Cells were broken in *Fast-Prep* device (MP-Biomedicals) in 3 rounds of 30 seconds each, 5.5 m/s. Cell debris was pelleted at 13000 rpm for 5 min. Supernatant was extracted with 1 vol of chloroform: isoamyl alcohol (24:1). Aqueous phase was precipitated by the addition of 0.1 vol of 5 M LiCl and 2.5 vol of cold EtOH 96 %, and stored at - 20 ^{0}C O/N. RNA was pelleted by centrifugation at 13000 rpm for 15 min, washed with EtOH 70 % (v/v), dried in *SpeedVac concentrator* (medium drying rate) for 2 min, and resuspended in 300 µL of RNase free water. Radioactively labeled RNA was used for hybridization on macroarrays.

[a] Growth temperature for *S. cerevisiae* cells was adjusted to 23, 30 or 37 ^{o}C depending on the experiment.

[b] Transcription mix

Per sample:

- 120 µL of 2.5X Transcription buffer (50 mM Tris-HCl pH 7.7, 500 mM KCl, 80 mM $MgCl_2$)
- 16 µL of 10 mM rNTPs (ATP, CTP, GTP, 10 mM each)
- 6 µL of 0.1 M DTT
- 20 µL of [α-^{33}P] UTP (Perkin Elmer, 3000 Ci/mmol, 10 µCi/µl, 3 µM)

4.3 Non- radioactive Genomic run-on (BioGRO)

BioGRO technique was adapted from Jordán-Pla *et al* (2015)[289]. Yeast cells were grown in liquid YPD up to $O.D._{600}$ ~ 0.5, at the correspondent temperature (30 ^{o}C and 37 ^{o}C for *S. cerevisiae* and *C. albicans* respectively). At that point, a 50 mL aliquot was collected by centrifugation at 4500 rpm for 3 min. Cell pellet was resuspended in 10 mL of 0.5 % Sarkosyl and collected again. For RNase treatment, cell pellet was resuspended in a mixture of 32 µL of 10 mg/ml RNase A (Roche) in 3.2 mL of 0.5 %. Sarkosyl. The mix was incubated for 10 min at 30 ^{o}C with slight agitation. To wash the RNase, cells were washed 3-fold with 45 mL of 0.5 % Sarkosyl. Cells were collected again, Sarkosyl containing remaining of RNase was discarded, and cell pellet was resuspended in 1 mL of 0.5 % Sarkosyl and transferred to an eppendorf tube, followed by cell collection one more time. Supernatant containing Sarkosyl was discarded and cell pellet was resuspended in 113.5 µL of distilled water and 5 µL of 40 U/ µL RNase OUT (Recombinant Ribonuclease inhibitor, Invitrogen). Run-on was performed by the addition of 162.25 µL of transcription mix[d] (Final reaction volume ~ 290 µL). From this step on, all steps were carried out protecting the tubes from light exposure, as Biotin molecules are light-sensitive. The mix was incubated at 30 ^{0}C for 5 min at 650 rpm. Reaction was stopped by addition of 1 mL of cold distilled water and immediately put on ice. For total RNA extraction, cells were collected by centrifugation at 4 ^{o}C for 2 min at 6000 rpm, supernatant was discarded and RNA was extracted from cells using *MasterPure^{TM} Yeast RNA purification Kit* (Epicentre). In our hands, the average amount of RNA obtained with this

protocol is 25- 30 µg. When size selection was needed, total RNA <200 nt was isolated using *NucleoSpin2 miRNA kit* (Macherey-Nagel), following the manufacturer's instructions described in section 6.4 of the manual (Purification of siRNA and large dsRNA from DICER reactions). Total RNA was quantified using *Nanodrop 2000 Spectrophotometer*. To check for biotin incorporation, total RNA was then subjected to northern blot analysis as described in 3.5.1.

[d] Transcription Mix

Per sample:

- 120 µL 2.5X Transcription buffer (50 mM Tris-HCl pH 7.7, 500 mM KCl, 80 mM $MgCl_2$)
- 6 µL 0.1 M DTT
- 16 µL rNTPs (ATP, CTP, GTP, 10 mM each)
- 20.25 µL 10 mM Biotin-11-UTP

4.4 BioGRO-seq

For sequencing, BioGRO RNA samples obtained as described in 4.3, were purified with Streptavidin M-280 magnetic beads (Invitrogen) as described in 3.1.3. For library preparation isolated Biotinylated RNA molecules were treated with T4 Polynucleotide kinase (Roche) , in a mixture containing 18 µL of Bio-RNA, 1 µL of PNK (10 U/ µL), 10 µL of 10X PNK buffer, 10µL of ATP (10 mM) and 61 µL of distilled water. The mix was incubated for 30 min at 37 ^{o}C, followed by incubation at 65 ^{o}C for 10 min to inactivate the enzyme. Samples were sequenced in the Gene Core facilities of the European Molecular Biology Laboratory (EMBL, Heidelberg, Germany) using the platform Illumina HiSeq2000 and small RNA sequencing library preparation, selecting an insert range for sequencing between 30 and 200 nt. Data analysis were done as described in 5.3 and 5.4.

4.5. Hybridization of labeled probes on nylon macroarrays.

Macroarrays used in this study were produced in the DNA chips service S.C.S.I.E. of the University of Valencia. Positively charged nylon membranes (Nytran SPC, Whatman) were spotted with 6144 dsDNA probes generated by PCR. Out of the total number of spots, 6009 correspond to ORFs of the *S. cerevisiae* genome. The rest of the spots contain spike-in controls, such as rDNA, and bacterial DNA from both *E. coli* and probes for three genes of *B. subtilis*. For reference see Alberola *et al*. 2004[290]. Macroarrays were pre-hybridized with 5 mL of hybridization solution (0.5 M Phosphate buffer pH 7.2, 1mM EDTA, 7% SDS) for 1 h at 65 ^{0}C. For hybridization, 300 µL of total RNA was dissolved in 3.2 mL of hybridization solution (Final volume 3.5 mL), added to the tube containing the nylon membrane and incubated for 24 h at 65 ^{0}C. Membrane was washed once with washing solution 1 (1X SSC, 0.1% SDS) for 20 min, followed by 1 wash with washing solution 2 (0.5X SSC, 0.1% SDS) for 10 min. All washes were done at 65 ^{0}C. After washing, membranes were sealed in plastic film and exposed to an *Imaging plate* (BAS-MP, Fujifilm) for 5-7 days (depending on cps value recorded with a Geiger counter; 10 cps ~ 5 days exposure), and scanned with Fujifilm FLA3000^{TM} Phosphorimager. Intensity of the spots was determined as described in García-Martínez *et al*. (2011). For re-utilization, membranes were stripped 2-fold by incubation in boiling stripping solution (5 mM Sodium phosphate buffer pH 7, 0.1 % SDS) with slight agitation for 10 min.

4.6. Hybridization of labeled probes on Affymetrix tiling arrays

Custom Affymetrix® strand-specific tiling arrays were specially designed for the purpose of this study. These arrays were designed to contain probes for the *C. albicans* genome (Assembly 21), being the probe design based on an set of Affymetrix tiling array that was already available for *S. cerevisiae*, developed in the group of Lars Steinmetz (EMBL, Heidelberg) The reference for the specific *C. albicans* arrays used in this study is: *GeneChip® Oxystressb520842F array* (Affymetrix, Santa Clara, California, USA), and they contain 2.3 x 10^6 25-mer probes that overlap by 13 bp, leaving a gap of 12 bp between probes.

All experimental steps dealing with these tiling arrays were done in the "Servicio de Análisis Multigénico" de la Facultad de Medicina, S.C.S.I.E. (Universitat de València).

5. Software and tools for data analysis

5.1. Macroarrays data analysis

Once the macroarrays image was obtained, intensity of the spots was quantified with *Array Vision™* image analysis software. sARM value (background corrected ARM density) was used. Spots which signal was lower than 1.5 times the background noise were discarded for analysis.

5.2. Affymetrix tiling array data analysis

The scanning of the microarray provides the user with a file in .CEL file format that contains information relative to the signal intensity obtained for each spot on the microarray. The .CEL files of all three replicates of each experiment were averaged and normalized together using *Affymetrix Tiling analysis software* v1.1.02 (TAS), using the default analysis parameters. The TAS software relies on quantile normalization of the microarray data. For the subsequent processing of the data, an R programming language-based script was used in R studio (http://www.r-project.org/). For the specific analysis of the data present in this work a special R script was created by Antonio Jordán at the University of Valencia and Vicent Pelechano (EMBL Heidelberg).

This script was adapted from the Bioconductor Tiling array analysis package (http://www.bioconductor.org/packages/2.11/bioc/html/tilingArray.html).

5.3. Next Generation Sequencing (NGS) data analysis in *S. cerevisiae*

Libraries were sequenced on the *Illumina HiSeq 2000 system* using single-end read mode. Fastq files where inspected with FastQC to assess quality (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Sequences of 15 nucleotides or less were discarded as considered too short. Sequences containing all N values were also filtered. Adapter sequences were trimmed using the fastx_clipper function implemented in the FASTX-Toolkit. The remaining high quality reads were mapped with Bowtie2[291] to the Ensembl *S. cerevisiae* genome assembly S288C (version R64-2-1), allowing for up to two mismatches in the alignment. BAM file alignments were then coordinate-sorted and indexed with SAMtools[292]. IGV tracks were generated for visual inspection of the aligned reads in IGV browser (https://www.broadinstitute.org/igv/). Read counts were generated using *HTSeq*[293]. Counts were normalized in every sample to their corresponding library size using the RPKM method[294]. Correlation scatter plots were generated with the R package LSD to compare normalized expression values between samples. Normalized average metagene plots around TSS and pA sites were generated in python and R with the package *ngs.plo*[295] using the Ensembl *SacCer3* genome annotation as reference. For that, physical coverage of reads around regions was calculated by extending each alignment to the average 50 bp fragment length, according to FastQC analysis. For metagene plots spanning the whole gene body vectors where normalized to have equal size with a spline fitting function, and then values were taken at equal intervals along the gene body and its flanking regions. Curve smoothing was used to remove outliers in the data by applying a sliding window approach. To further avoid distortion of the average profiles, the statistical robustness parameter was applied to all calculations. This function filters out 0.5% of the genes with the most extreme expression values. To plot the data taking as a reference the nucleosome +1 dyad, the Ensembl SacCer3 reference annotation was modified, substituting the TSS coordinates of every gene by the nucleosome +1 center (dyad) position, taken from the *Saccharomyces* genome nucleosome atlas (http://atlas.bx.psu.edu/cj/nucl_retrieval.html). Expression heat map graphs were

generated with the *ngs.plot* and *replot.r* script, and genes were ranked by the total sum of counts for the region of interest. All NGS data analyses were done by Antonio Jordán Pla at the Department of Molecular Biosciences of the Wenner-Gren Institute, University of Stockholm.

5.4. NGS data analysis in *C. albicans*

The data analysis strategy followed to analyze *C. albicans* RNA-seq data was essentially the same to the one used for *S. cerevisiae*. However, there were several steps that required different parameters or inputs used which are briefly described in this section. The mapping of high quality and adaptor-clipped reads was done against the SC5314 (A21) reference genome from *Ensembl*, taking as input the individual *Fastq* files from each biological replicate and time point. The same genome assembly was used as reference for the generation of average nascent transcription profiles with *ngs.plot*. In order to align and average the data around the nucleosome +1 positions, the canonical *Ensembl* annotation was modified to change the ORF-start coordinate by the nucleosome +1 dyad position, derived from a list of NFR positions that was kindly provided by Dr. Oliver Rando. The two biological replicates for each time point were merged into one at the level of the alignment BAM files using the MergeSamFiles function from *Picard* (Broad Institute), as implemented in Galaxy (https://galaxyproject.org). Once merged, the files where coordinate-sorted and indexed with *SAMtools (http://samtools.sourceforge.net)*. The merged alignment files were used to replicate the same downstream analysis performed with the individual replicate files. All RNA-seq analyses were done by Antonio Jordán Pla at the Department of Molecular Biosciences of the Wenner-Gren Institute, University of Stockholm.

5.5 Data bases and other tools for computational data analysis

- SGD - *Saccharomyces* genome database (www.yeastgenome.org)

- CGD - *Candida* genome database (www.candidagenome.org). To look for enrichment in functional categories, the CGD GO term finder tool was used.

- IGV - Integrative genomics viewer (http://www.broadinstitute.org/igv) – Tool used for visualization of the *C. albicans* genome, that allow the design of the intergenic primers used in this study.

- Bioconductor – Bioconductor is an open source, open development software project to provide tools for the analysis and comprehension of high-throughput genomic data. It is based primarily on the R programming language (https://www.r-project.org/). In this study, the Bioconductor tiling array analysis package was used to process the microarray data.

- R studio- (http://www.rstudio.com) – Software environment for statistical computing and graphics that uses R programming language. R studio was used to process the microarray data obtained in this work.

- **TilingScan** – (www.tilingscan.uv.es) – Online application developed in this work for the study of differentially expressed DNA regions in tiling microarray data.

- **Online Venn diagram generator**: http://bioinformatics.psb.ugent.be/webtools/Venn

Chapter 1

Comparative transcriptomic-proteomic study of the oxidative stress response in *C. albicans*

1.1. Motivation

The process of translation has been known for years to be the most energy costly process of cell physiology in bacteria and yeast[268,296,297]. Regardless of the individual energy need each organism displays, this huge energy cost that protein production implies is common to all organisms, and for this reason it is generally accepted that regulation of gene expression is mainly done at the transcriptional level, to avoid the cost of regulation at the later stage of translation[173,174,298]. However, oonly limited data are available describing protein burden in eukaryotic cells[299-301], and it is not clear whether results inferred from one cell type, or from specific conditions, can be generalized to other organisms and environments.

When studying how a certain organism responds to changes in the environment, changes at the transcriptional and translational level are usually addressed separately to account for the individual contributions of each process to the change. However, interrogating protein and transcript levels simultaneously can have the benefit of providing insight into the regulatory mechanisms that underlie the change[302]. Despite the potential these type of studies have to open new fields of research, only a few studies have been performed to address simultaneous changes in mRNA and protein levels during stress responses in yeast[303-305]. One of these studies was performed in *S. cerevisiae* using cells that had been subjected to mild oxidative stress induced by diamide[303]. Contrary to the evidence observed in bacteria, the results of this study indicated substantial post-transcriptional regulation of a large fraction of the studied genes, regardless of their up or down-regulation. The importance of translational regulation, however, has been recently challenged by statistical meta-analyses[173,174]. The current opinion nowadays is that most regulation is performed at the transcriptional level, while translation potentiates the response and has specific roles in some particular cases[173,265,266].

To date, no comparative transcriptomic-proteomic studies have been published in *C. albicans*. However, a proteomic analysis of the oxidative stress response to the treatment with diamide is available, in which this response is compared to the one triggered by a different oxidizing agent, hydrogen peroxide (H_2O_2). In this case, the authors observed that the pattern of up-regulated proteins was almost identical under both stressors, whereas down-regulation appeared to be

much less specific. The authors suggested that this difference could be explained by the wide range of potential targets of ROS toxicity.

Give the variability of the response mechanisms that different organisms can display, not only in general but also in response to specific environmental changes, it was decided to perform a transcriptomic-proteomic study to explore the specific response of *C. albicans* to the oxidative environment caused by the oxidizing agent used in this study, tert-butyl hydroperoxide (t- BOOH).

1.2. Simultaneous comparison of protein and mRNA levels under oxidative stress in *C. albicans*.

To explore the specific response of *C. albicans* to the treatment with t-BOOH, a comparative transcriptomic-proteomic study was performed in cells that had been subjected to the treatment with this agent over a time-course experiment. To do so, yeast cells were grown in YPD medium at 37 ^{o}C until mid-log phase. At this point, a no-stress control aliquot was collected, and t-BOOH was added to the culture to a final concentration of 0.1 mM. This concentration reduced the growth rate but did not completely abolish cell growth (Figure 22).

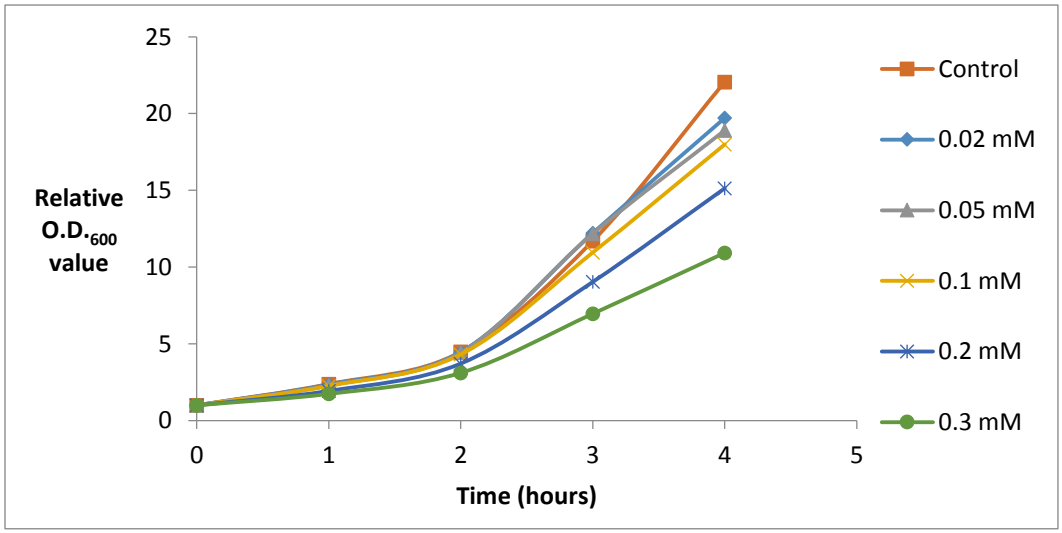


Figure 22. Growth curves of *C. albicans* SC5314 under oxidative stress caused by the addition of increasing concentrations of t-BOOH. Wild-type cells were grown at 37 ^{o}C in YPD medium for a total period of 4 hours. Growth curves were performed in three independent biological replicates. Average O.D. $_{600}$ measurements taken every hour are shown in respect to time 0.

Over a total time period of 3 hours under stress, two time points were chosen: one at time 40 min or 60 min (for the transcriptomics and the proteomics respectively) and a final aliquot at time 180 min. The reason for the divergence in the middle experimental point for the transcriptomics and the proteomics was that they respectively showed maximum peaks of expression of oxidative-responsive genes and proteins in response to this treatment at these time points (E. Herrero, Universitat de Lleida, personal communication). The election of this time points was also supported by a previous study that reported a 10-20 min delay between the mRNA and protein response to oxidative stress 303. Hence by doing this, relative mRNA levels were expected to be compared with their correspondent protein products.

For each time point of study, two aliquots were collected in parallel: one for total RNA extraction and one for total protein extraction that were used for the transcriptomic and proteomic analysis respectively (Figure 23).

This study was done within the framework of a collaborative study with the group of Enrique Herrero in the University of Lleida, where all proteomic analyses were performed using 2DE followed by protein identification by MS/MS. Only the transcriptomic analyses were done at the University of Valencia. For this reason, specific technical details on the proteomic analysis will not be discussed in this work. Even though the whole set of proteins in the cell culture were ran in 2D-gels for separation, only those which expression changed after the treatment were selected for identification. Due to financial constraints of the project, only some spots were identified, 29 of them being uniquely associated with a protein. Once these 29 proteins were uniquely detected, the expression levels of their correspondent mRNAs, as observed in the transcriptomic data, were compared at the time points of study. The expression levels of both proteins and transcripts along the time course are shown in Table 4.

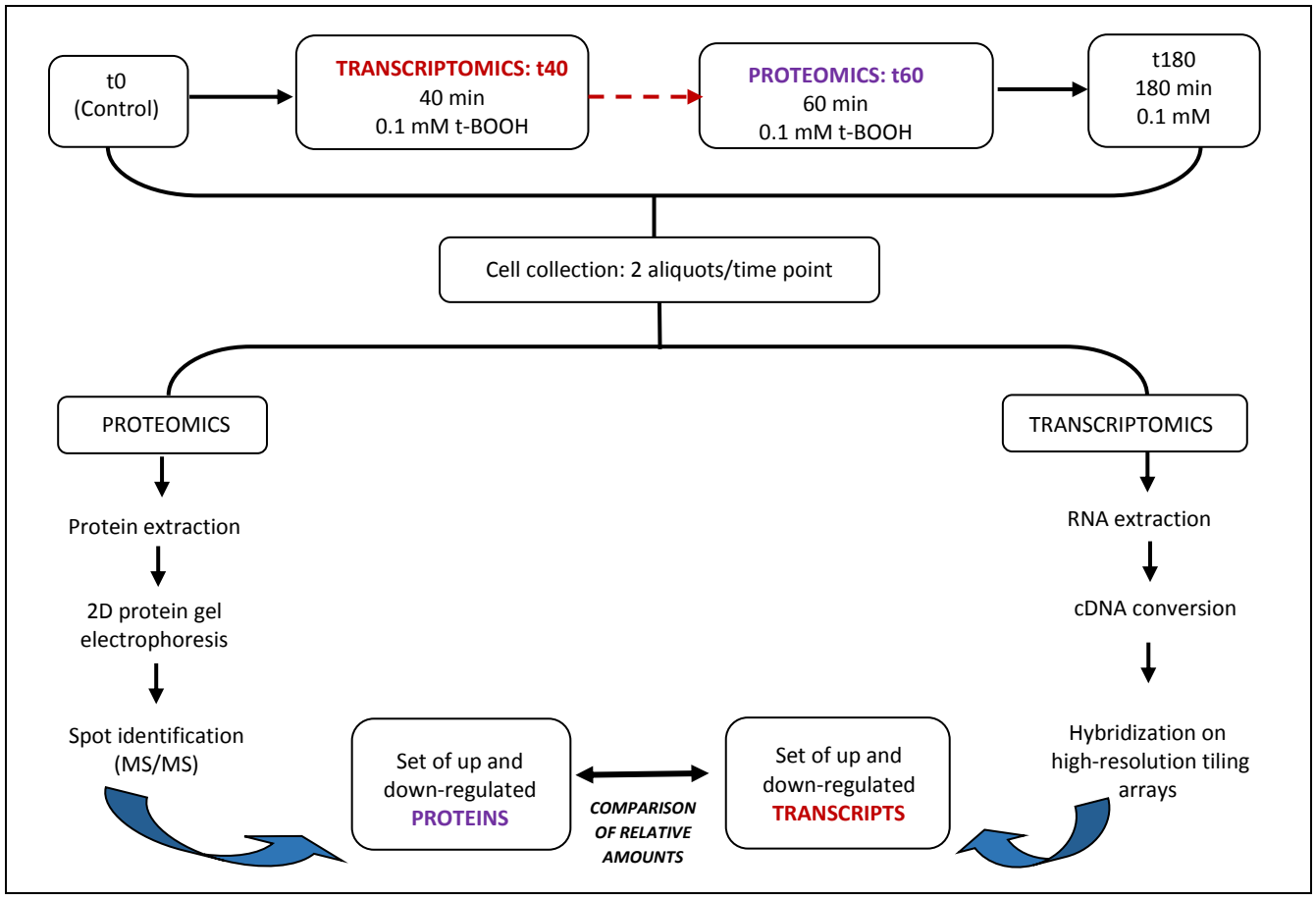


Figure 23. Experimental workflow for the transcriptomic-proteomic study under oxidative stress. Cells in exponential growth ($O.D._{600}$ 0.5) were treated with t-BOOH to a final concentration of 0.1 mM. After the treatment, 2 aliquots were collected at each time point, one for total RNA extraction and one for total protein extraction. **(Left)** Total cell protein content was extracted and separated by two-dimensional gel electrophoresis (2DE). Those protein spots separated in the gel which levels changed upon the treatment were then selected for potential identification through mass spectrometry (MS/MS). **(Right)** Total RNA was isolated, converted into cDNA and hybridized on high-resolution tiling arrays. A complete set of relative mRNA expression values was obtained for all *C. albicans* genes, and compared to their correspondent protein levels in those cases in which the proteins were identified. This transcriptomic analysis was performed in three independent biological replicates.

The proteomic results include many proteins that were already expected to be up-regulated, such as the peroxidases Tsa1[233] and Prx1[232], and the heat shock proteins Ssc1 and Ssz1[306]. In agreement with previous findings is also the down-regulation of the enolase Eno1, which was already found to decrease its expression levels in response to high concentrations of H_2O_2[306]. Interestingly however, this same previous study reported the up-regulation of the protein Ssa1 upon the addition of H_2O_2[306], which in this case appears down-regulated under the effect of t-BOOH. Another major disagreement with previously published data is the observed down-regulation of the superoxide dismutase Sod1, which in a previous study had proven to be necessary for the protection of *C. albicans* against H_2O_2[226]. In both cases, however, the stressor used was different from the one used in this study, which may explain the differences. In the same line of argument is the comparison of this work's data set with the one published by Kush *et al.* in 2007, where they performed a proteomic study to study the response of *C. albicans* to the treatment with H_2O_2 and diamide[307]. While Cip1, Tsa1 and Ebp1 were also up-regulated in that study, the expression levels of Pdb1, Ssb1, and Ssa1 went in opposite directions in that study compared to the one presented herein. While Pdb1 and Ssa1 were down-regulated in the presence of both H_2O_2 and diamide, it appeared up-regulated in the presence of t-BOOH. And the opposite happened with Ssa1, up-regulated in the presence of H_2O_2 and diamide (in agreement with the results previously mentioned[306]), and down-regulated in the presence of t-BOOH. All these differences observed between this and other studies are in line with the idea of the oxidative stress response in *C. albicans* being stressor-specific rather than general.

When looking at the transcript levels of the proteins listed in Table 4, it can be seen that, except for 3 genes highlighted in red (namely *TSA1*, *CIP1* and *SSC1*), all the rest of proteins did not change their transcript levels along the time course, regardless of their up or down-regulation. Based on this, these results suggest that except for the three above-mentioned genes, 26 (90%) of the genes studied are regulated at the post-transcriptional level in the response to t-BOOH in *C. albicans*. This result is similar to that observed in previous studies performed in *S. cerevisiae* under oxidative stress caused by diamide and sorbic acid [303,304], but differs from the conclusions drawn in one study done in fission yeast with H_2O_2 [305]. All this suggest that the overall response to oxidative stress is genus-specific and, in the particular case of *C. albicans*, stressor-specific too.

These findings further remark the importance of the simultaneous interrogation of protein and mRNA levels to get clues on the underlying regulatory mechanisms that control adaptation to stress in different organisms, rather than focusing exclusively on transcript levels as predictors.

Lastly, it is important to remark the protein abundance values observed at time 180 min. A population of cells growing in log phase is often said to be at a steady state, being this state defined by a zero net change of a parameter in a system, such as mRNA or protein concentrations. When a population of cells has been subjected to a stimulus (for example, stress), the concentrations of proteins across the population change over a specific period of time, and thus are not at the steady state, until the stimulus ceases and the population reaches the steady state again[308]. This steady state may reflect similar levels of proteins to those observed before the stimulus, or might differ from it. The time point 180 min in this study was chosen to monitor the recovery of the cell back to the steady state level. The results obtained show, that still 3 hours after the shock, the protein levels do not recover back to normal.

					PROTEOMICS		TRANSCRIPTOMICS	
	PL	UniProt Code	Gene name	Function or homology	t60/t0	t180/t0	t40/t0	t180/t0
1	UP	Q5AKX1	GCV3	Glycine decarboxylase, subunit H	22,68	23,90	0,94	0,95
2	UP	Q9Y7F0	TSA1	TSA/alkyl hydroperoxide peroxidase C (AhPC) family protein	14,99	18,94	11,51	1,01
3	UP	Q5AG68	YNK1	Nucleoside diphosphate kinase (NDP kinase)	10,24	7,02	1,00	0,99
4	UP	C4YCX0	CDC48	Putative microsomal ATPase	7,59	8,66	0,80	0,90
5	UP	Q5ANP2	EGD2	Nascent polypeptide associated complex protein alpha subunit	7,06	6,21	0,84	0,94
6	UP	XP_722820	MDH1-1	Malate dehydrogenase	5,01	4,42	0,90	1,00
7	UP	XP_714320	CIP1	Possible oxidoreductase	4,35	2,87	13,90	0,99
8	UP	P78590	EFB1	Elongation factor 1-beta	4,35	4,35	0,76	0,94
9	UP	P43084	EBP1	Probable NADPH dehydrogenase	3,86	2,96	1,47	0,99
10	UP	P43066	ARD1	D-arabinitol 2-dehydrogenase	3,38	3,61	1,00	0,99
11	UP	XP_711076	ARO10	Phenylpiruvate decarboxilase	2,46	2,28	0,92	1,07
12	UP	P87222	SSB1	Ribosome-associated molecular chaperone SSB1	2,38	2,93	0,83	0,92
13	UP	XP_717002	PRX1	Thioredoxin peroxidase	2,35	3,35	1,03	1,01
14	UP	P43065	LYS1	Saccharopine dehydrogenase	2,04	1,88	1,21	0,96
15	UP	C4YH62	PDB1	Putative pyruvate dehydrogenase	1,80	2,06	0,73	0,97
16	UP	XP_717293	SSZ1	Putative HSP70 chaperone	1,77	2,1	0,69	0,96
17	UP	XP_716953	PDI1	Putative protein disulfide-isomerase	1,71	2,03	1,02	0,95
18	UP	P83784	SSC1	Heat shock protein	1,48	2,23	3,53	0,90
19	UP	XP_715736	TFS1	Putative carboxypeptidase y inhibitor	*	0,94**	1,26	1,04
20	DOWN	O59924	SOD1	Cytosolic copper- and zinc-containing superoxide dismutase	0,69	0,28	1,14	1,00
21	DOWN	P41797	SSA1	HSP70 family chaperone	0,54	0,48	0,94	0,94
22	DOWN	XP_712370	LPD1	Putative dihydrolipoamide dehydrogenase	0,52	0,4	0,94	0,95
23	DOWN	C4YI21	SEC72	ER protein-translocation complex component	0,49	0,67	0,94	1,03
24	DOWN	O13287	GND1	6-phosphogluconate dehydrogenase	0,39	1,31	1,11	0,99
25	DOWN	Q59WW5	ILV5	Ketol-acid reductoisomerase	0,37	0,40	0,85	0,90
26	DOWN	Q5A0M1	IDH2	Putative mitochondrial NAD-isocitrate dehydrogenase subunit	0,37	0,4	0,87	0,95
27	DOWN	Q5AEF2	SEC13	Putative protein transport factor	0,26	0,37	0,93	0,99
28	DOWN	P30575	ENO1	Enolase	0,26	0,15	1,03	0,95
29	DOWN	O42766	BMH1	14-3-3 protein	0,10	0,00	1,00	0,96

Table 4. Summary table of the comparison between the relative levels of mRNA and proteins under the treatment with t-BOOH in *C. albicans*. Proteins identified in the proteomic study were assigned to their correspondent gene identifier and classified based on their up (red) or down (blue) regulation upon the treatment. Average fold change values are shown on the right in descent order. Gene names and description were taken from the CGD database, using the name and ORF identifier from the assembly 22 of the *C. albicans* genome. Transcriptomic values in red are indicated to highlight the only transcripts on the list which expression changed after the treatment. * = Protein not present at time 0 of the experiment. Being the control value (t0) equal to zero, the ratio t60/t0 could not be calculated. **= Protein not present at time 0 of the experiment. Being the control value (t0) equal to zero, the ratio t60/t0 could not be calculated. The value shown (0, 94) corresponds to the ratio t180/t60). All proteomic data were provided by the group of Enrique Herrero at the University of Lleida, Spain.

1.3. Experimental validation of the transcriptomic and proteomic data using single-gene techniques

Despite the current robustness of genome-wide experimental approaches, it is always advisable to validate their results with single-gene techniques to account for the potential errors that can be introduced during data pre-processing and analysis. In order to further verify the results obtained using genome-wide approaches, single-gene techniques were used for the experimental validation of the data. For the validation of transcriptomic data, quantitative PCR (qPCR) analyses were performed to measure the mRNA abundance of specific genes individually. To validate the proteomic data, western-blot analyses using specific antibodies against the identified proteins would have been the ideal approach. However, we were unable to find available antibodies against any of the proteins presented in this study. For this reason, and considering our results indicating that these genes are regulated at the post-transcriptional level, it was decided to approach the validation of the proteomic data in an indirect manner, using polysome profiling. To do so, and aiming to reproduce the experiment shown in Figure 23, the experiment depicted in Figure 24 was designed. Briefly, cells were grown in the same experimental conditions (YPD, 37 ^{o}C) up to O.D.$_{600}$ 0.5. At this point, t-BOOH was added to the culture to a final concentration of 0.1 mM. Two different time points were analyzed: time 0 (no stress control) and time 40 min (40 min after the addition of t-BOOH). At each time point, two aliquots were collected: one for total RNA extraction and one for polysome extraction, which were used for the validation of the transcriptomic and the proteomic data respectively. The results of each independent validation are shown in the two following sections.

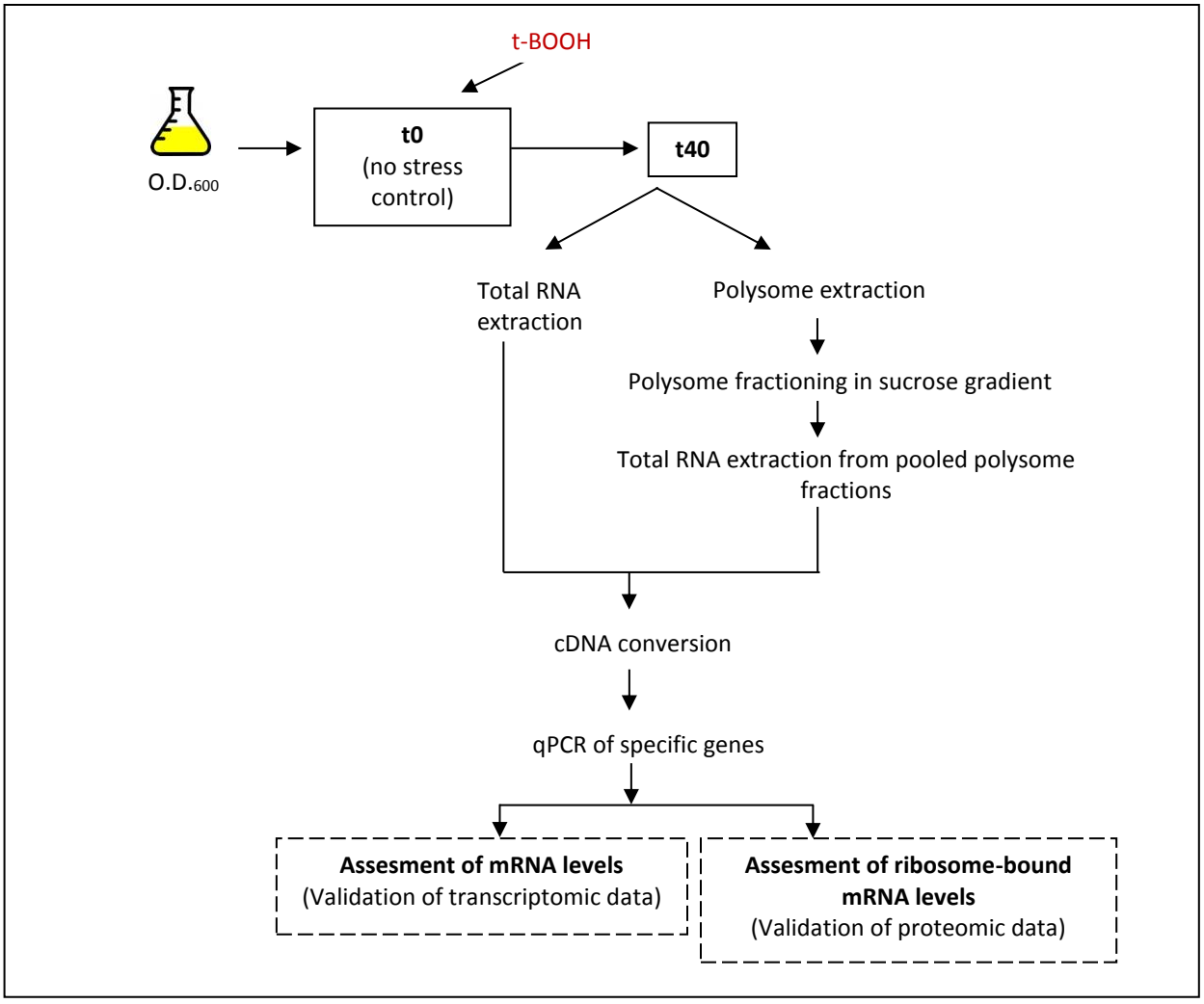


Figure 24. Experimental workflow for the validation of the transcriptomic and proteomic data. Cells in exponential growth phase ($O.D._{600}$ 0.5) were treated with t-BOOH to a final concentration of 0.1 mM for 40 min after collection of a no-stress control. After the treatment, 2 aliquots were collected at each time point, one for total RNA extraction (left) and one for polysome extraction (right). Upon polysome extraction, total polysome cell content was fractionated and pooled for polysome-bound mRNA isolation. In both cases (left and right) total mRNA was isolated, converted into cDNA and subjected to qPCR analysis. This experiment was performed in 3 independent biological triplicates.

1.3.1. Experimental validation of the transcriptomic data

For the validation of the mRNA amount determined at a genomic-scale using tiling arrays, the relative transcript levels of 4 randomly selected genes was determined individually using qPCR. The measurement was done for 4 genes contained in the transcriptomic data set, as follows:

- Two genes which mRNA levels were significantly up and down-regulated after the treatment (*CIP1* and *CHT3* respectively). This aimed to serve as a positive control of change.
- One gene which mRNA levels remained unchanged but protein levels were up-regulated after the treatment (*ARD1*).
- One gene which mRNA levels remained unchanged but protein levels were down-regulated after the treatment (*SEC13*).

Relative mRNA levels for these genes at time 40 were compared to the levels observed at time 0 after normalization by a no-change control (the house-keeping gene *ACT1*). As shown in Table 5, the fold-changes observed using qPCR validated those observed using high-resolution tiling arrays.

	CIP1	***CHT3***	***ARD1***	***SEC13***
qPCR	15.01 ± 4.57	0.44 ± 0.1	1.45 ± 0.43	1.37 ± 0.47
Tiling array	13.90 ± 3.63	0.39 ± 0.89	1.03 ± 0.57	0.93 ± 0.33

Table 5. Experimental validation of mRNA abundance using qPCR. Fold-changes of mRNA levels after 40 min under the treatment with 0.1mM t-BOOH are shown. The correspondent data obtained by means of tiling array hybridization is shown underneath for reference. Relative expression levels in qPCR experiments were normalized by those of *ACT1*. Values reflect the average of three individual biological replicates.

1.3.2 Experimental validation of the proteomic data

For the validation of the protein levels determined at a global scale using 2DE-MS/MS, the relative levels of ribosome-bound mRNAs before and after the oxidative stress treatment were determined for 5 genes selected at random. To do so, the total polysome population contained within the cells was isolated at each time point, and loaded into a sucrose gradient for fractionation. For total RNA extraction, different fractions were pooled and named M, P1, P2 and P3, standing for a pool of the fractions containing monosomes (M), and 3 pools of the fractions containing polysomes (P1, P2, P3). Profiles of absorbance at 260 nm (A_{260}) of the different fractions are shown in Figure 25.

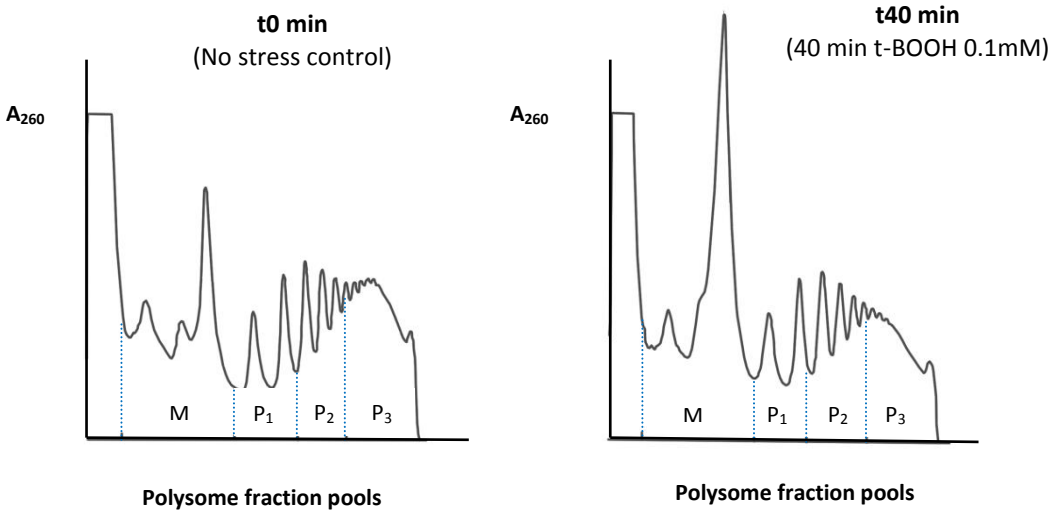


Figure 25. Polysome profiles before and after the treatment with 0.1mM t-BOOH for 40 min. Profiles of A_{260} reflecting ribosome-bound mRNA abundancy in each pool of polysome fractions are shown for the no stress control (left) and the time point 40 min after the addition of 0.1mM t-BOOH (right).

For each gene subjected to analysis, the relative amounts of free or translationally inactive RNA (mRNA bound to monosomes) were compared to the relative amounts of translationally active RNA (mRNA bound to polysomes). The genes selected for this analysis were as follows:

- An example of transcriptional regulation: gene which both protein and mRNA levels were up-regulated after the treatment (*TSA1*).
- Four different examples of post-transcriptional regulation: Two genes which mRNA levels remained unchanged over the time course, but which protein levels were up-regulated (*ARD1* and *PRX1*); and two genes which mRNA levels remained unchanged over the time course, but which protein levels were down-regulated (*ENO1* and *BHM1*).

The fold-changes of ribosome-bound mRNA for each selected gene are shown in Table 6.

	TSA1	***ARD1***	***PRX1***	***ENO1***	***BMH1***
qPCR (t40/t0)	9.11± 0.98	2.26 ± 0.59	3.2 ± 0.43	0.32 ± 0.27	0.2 ± 0.11
2DE-MS/MS (t60/t0)	14.99	3.38	2.35	0.26	0.1

Table 6. Fold-changes of ribosome-bound mRNA levels upon the addition of 0.1 mM t-BOOH for 40 min. For 5 different genes present in the comparative transcriptomic-proteomic study, the relative abundance of ribosome-bound mRNA was compared between the time 0 (no stress control) and time 40 min after the addition of 0.1mM t-BOOH. Relative mRNA values were normalized by an internal RNA extraction control, volume correction factor and ribosome load (see materials and methods for detailed description of data normalization).

It can be seen in the results that the qPCR data of ribosome-bound mRNA is in agreement with the protein level results, validating the results obtained using 2DE-MS/MS. t should be noted that polysome fractioning followed by qPCR is nothing but an indirect way to estimate the level of translation of the studied genes. By doing this, changes in levels of ribosome-bound mRNA are assumed to be reflected in changes in the final protein product, which will not be correct in the case of those genes which regulation is occurring via mRNA or protein stabilization. However, for *ARD1, PRX1, ENO1*, and *BMH1*, polysome profiling seemed to be a good approach since in these genes changes in ribosome-bound mRNA were later reflected in changes in protein levels, verifying that, at least in these cases, regulation is performed at the translational level.

1.4. Discussion and future perspectives

The ultimate aim of the characterization of a biological system is not the analysis of mRNA transcript levels alone but the supplement of this information with determination of both protein expression levels and their respective activities. Even though measured mRNA levels can in some cases be extrapolated to indicate the levels of the corresponding protein, many current studies have proven the lack of correlation between mRNA and protein abundances, remarking the importance of the interrogation of both[303,305,309-311] . Despite discarding the possibility of using mRNA amounts as proxies of protein activity, this opens a new field of possibilities of study, as this lack of strict correlation can provide clues on the underlying regulatory mechanisms that govern the modulation of protein abundance.

In this study, protein and mRNA levels were interrogated simultaneously in the context of an oxidative stress environment caused by the addition of t-BOOH in *C. albicans*. This, to our knowledge, was the first comparative transcriptomic-proteomic study ever performed in this organism to date. By doing it, two main pieces of information were aimed to be obtained. First, studying the transcriptomic and proteomic changes of this organism under this type of stress caused by this particular oxidizing agent; and second, addressing whether the regulation of gene expression under these circumstances happens at the transcriptional or the translational level. That is, whether changes in mRNA levels are reflected in changes in the correspondent protein abundances or not. Despite the obvious advantages of parallel profiling of transcripts and proteins on a global 'omic' scale, there are different practical challenges involved in their application. Transcriptomics is now a robust, cost-effective technology, capable of simultaneously quantifying tens of thousands of mRNA species. Proteomic analyses, conversely, are currently much more limited in breadth and depth of coverage due to variations in protein abundance, hydrophobicity, stability, size and charge. The cost of the use of this technique to its final stage, protein spot identification, is also significantly higher than the one for a transcriptomic analysis. For this reason, even though this study was initially planned to be done at a global scale, economic constraints only allowed tens of proteins to be identified in this study, being this number far lower

than the one presented in a similar previous proteomic study[307]. The herein presented proteomic analysis only detected a minority of the potentially regulated proteins, which were amongst the most highly expressed and thus, the easiest to be detected in 2D-gels. Hence, the statistical relevance and the potential bias of this study preclude a generalization of these results to a transcriptome/proteome comparison. Due to the low number of elements analyzed, the conclusions drawn in this study cannot be general but more specific to the set of genes and proteins listed.

When comparing the results with those seen in previous oxidative stress response studies where different oxidizing agents were used, the up or down-regulation proteins sometimes coincided with the one shown in this study, sometimes not. Although there is still limited data to draw conclusions, these differences between studies suggest that the oxidative stress response in *C. albicans* is stressor-specific rather than general. This further supports the idea pointed out by the authors in Kush *et al.* 2007[307] that different oxidizing agents may target different elements within the cell, which in turn could explain the differences observed between studies. It is unknown however, if these differences are due exclusively to the type of stressor used, or also to the concentration of it, as the damage caused will vary with different concentrations of different reagents.

As for the type of regulation these genes are subjected to, 90% of the genes analyzed in this study seemed to be regulated at a post-transcriptional level, since their protein levels were up or down-regulated under stress whereas their transcripts levels did not vary. This being stated, the fact that protein metabolism is much slower than the mRNA counterpart implied drawback for this comparison. Proteins are synthesized later than mRNAs after stress. This was aimed to be addressed by taking a 20 min time window between mRNA and protein sampling. However, proteins are also much more stable than mRNAs, what makes mRNA response peaks much broader[305]. This may explain why many mRNAs corresponding to up or down regulated proteins did not show relative changes at 180 min after stress.

Whereas most studies still point at transcription being the main point of regulation under stress conditions in yeast[173], results like the ones shown in this study may constitute the first exceptions to the rule. It is important to keep in mind however, that not just the type of stressor applied, but

also its concentration and the time it has been present for will account for different degrees of severity of the treatment. And in fact, recent studies have pointed out that the relative contribution of mRNA- versus protein level regulation seems to be dependent on the temporal scale, complexity level of the biological system and on the type of perturbation applied[298]. For this reason, the difference observed between different yeast species or different treatments may just be responding to that. The post-transcriptional regulation of the genes listed in this study could be specific to this type of stress conditions, or general to the regulatory mechanisms of this organism, but the data obtained here is for now insufficient to cast any light into it. Not only this study should be broaden to a higher number of proteins, but also different oxidizing agents will have to be tested in the same way to be able to draw conclusions in that sense. Post-transcriptional mechanisms controlling gene expression in this particular scenario may include mRNA stabilization, translational control and control of protein half-life, which should be explored from now on.

Chapter 2

Development of a bioinformatic application for the study of antisense transcription under hypoxic and oxidative stress in *Candida albicans*

2.1. Motivation

Current genomic technologies allow laboratories to produce large-scale high-resolution transcriptomic data sets, either by the use of next-generation sequencing or tiling microarray platforms. To explore these data sets and get maximum value from the data, researchers need to view their results alongside all the known features of a given reference genome. To meet this need, several genome browsers have been developed over the past decade[312,313]. These platforms allow the user to visualize their own results along public domain gene annotations, sequences, and other previously published genomic data sets, facilitating researchers to determine how their results agree or disagree with previous findings and current understanding of genomic structure. When visualizing a given expression data set, these platforms align the putative transcripts to the reference genome, enabling the user to see whether these transcripts have been transcribed from already annotated genomic features (i.e. ORFs, tRNAs, snRNAs, etc.), or from other unpredicted sites of transcription (i.e. transcripts arising from intergenic regions).

To study the transcriptional changes that occur under a given experimental condition, very often researchers look for regions of the genome that are differentially expressed between conditions (i.e. stress versus no stress, mutant strain versus wild-type strain). For the detection of these regions, several algorithms have been developed over the years both for tiling microarray and for RNAseq data sets (reviewed in[314,315]) along with some bioinformatics platforms that have enabled their use[316]. However, currently available applications for comparative microarray analysis exclusively focus on changes in gene expression within known transcribed regions of predicted protein-coding genes, the list of which has to be previously created by the user. By just focusing on changes that occur on annotated ORFs, they miss changes in other potentially interesting but unpredictable features, such as ncRNAs. Given the recently discovered prevalence and functional importance of ncRNAs, it is of great utility to be able to search for differential expression in

an unbiased manner, not to miss the presence of transcripts in those regions of the genome that do not correspond to ORFs.

Tiling array data profiles generally have a main intrinsic technical constraint: they tend to show high background noise levels, often masking the presence of those transcripts that are transcribed in low numbers. For this reason, intensity profiles are usually very noisy, displaying high fluctuations of signal on every oligonucleotide probe, which can lead to the misidentification of individual peaks of intensity. All the above-mentioned browsers allow the easy adjustment of the background noise level, but do not provide the user with a way to avoid signal fluctuations. As a consequence, the simple visual inspection of intensity profiles in these platforms often leads to misinterpretation of differentially expressed regions.

A way to avoid the appearance of artifactual fluctuations is to smooth the data set prior to its visualization. The smoothing of a data set consists of applying an approximating function that removes noise from the data, allowing significant patterns to stand out. This function is referred to as an algorithm, which modifies the data points of a signal so that outliers (presumably because of noise) are reduced, and points that are higher or lower than the adjacent points are decreased or increased accordingly resulting in a smoother signal. Several packages for microarray data analysis that incorporate smoothing algorithms have been developed and made publicly available at *Bioconductor*, which is an open source software for bioinformatics that provides tools for the analysis and comprehension of high-throughput genomic data. However, despite being created to address biological issues, the purpose of these algorithms along with the way they work are often complex and difficult to understand for biologists. These packages are written in the statistical R programming language (www.R-project.org), which requires the user to spend large amounts of time to learn how the algorithm works and how to execute it on a standard computer.

The vast majority of data described in the present work was obtained using microarray platforms, which displayed all technical constraints and difficulties to their analysis described above. In order to analyze these data and to obtain as much biological information as possible, the development of a new computing tool was proposed.

2.2. Development of a bioinformatics application for the detection of differentially expressed regions in tiling microarray data

Aiming to address all the previously mentioned limitations, a new bioinformatics web application was developed in collaboration with José Juanes at the Department of Informatics of the University of Valencia. This application was named *TilingScan*[317], and its design was based on the following criteria:

1. It should incorporate a smoothing algorithm for the pre-processing of the tiling array signal intensity profile.
2. It should provide the user with an automatic tool to rapidly and accurately locate differentially expressed regions in unbiased manner, searching for changes in expression and then allowing the comparison of the results with current genome annotations.
3. It should be simple, intuitive, and have a user-friendly interface, not requiring the user to have any previous background on computer science to be able to use it.

The application was named "*TilingScan*" as it scans the results of tiling microarray experiments in search for differential expression. The created application enables the visualization of data from both tiling array and NGS as genome graphs, which display probe intensity alongside the genomic sequence, aligning it with a reference genome provided by the user. Once the data is uploaded to the server, the application allows custom visualization of either specific chromosomes or specific genes of interest via the "*Visualize chromosome"* and "*Visualize gene"* tools (Figure 26).

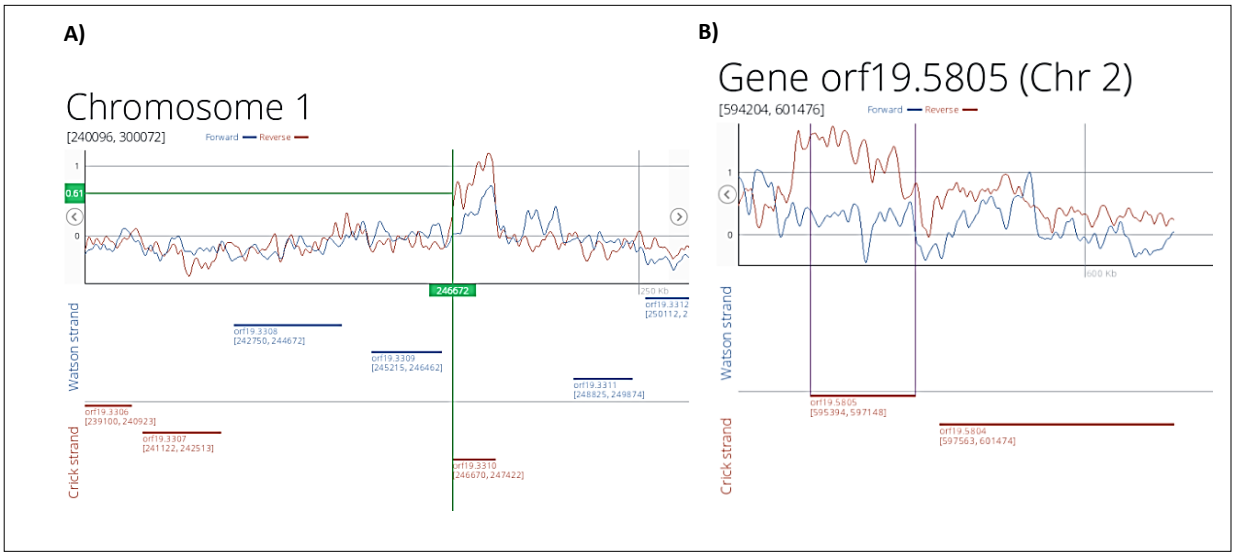


Figure 26. TilingScan graphical outputs for data visualization. By selecting the chromosome (**A**) or the gene (**B**) of interest, strand-specific signal intensity profiles will be shown, displaying the intensity profile on Forward (Watson) strand in blue and the Reverse(Crick) strand in red. Start and end chromosomal coordinates of the visualized region are shown in brackets (top, left). Intensity profiles are aligned to the selected reference genome (bottom).Sliding the mouse over the signal profile, X and Y coordinates for the selected data point will be shown (green rectangles on A panel), corresponding to the position in the genome (in bp) and the signal intensity value respectively.

By uploading data sets that reflect the ratio between two experimental conditions, peaks and valleys of intensity display differentially expressed regions. Along with visualization, the application provides the user with a search tool that accurately locates and identifies these regions via the "*Window search*" option. This tool is based on a modified version of the geometric moving average algorithm, which constitutes one of the several methods developed for the detection of abrupt changes in a signal[318]. Briefly, the search algorithm identifies those regions which intensity value surpasses (or drops below) a certain fold-change threshold that is manually selected, which then allows the identification of both up and down-regulated regions across the genome. For clarification, a more detailed description of the search algorithm is included in the Appendix I section of this work.

Upon performing the search, the application displays both a list of all the graphical outputs of the detected regions (Figure 27) and a table in which the following information is registered:

- *Project name* (data set from which the regions were identified)
- *Chromosome* (Chromosome in which the regions are encoded)
- *Length* (length of the detected region)
- *Start* (Chromosomal coordinates of the start point of the region)
- *End*(Chromosomal coordinates of the end point of the region)
- *Watson Y- mean*(Mean intensity value across the region on the Forward strand)
- *Crick Y-mean* (Mean intensity value across the region on the Reverse strand)
- *ORFs* (Systematic ID of all ORFs found to be encoded within or overlapping the detected region, if any)

These features will not only be automatically registered after the user runs the analysis but can also be manually annotated if some other region of interest is found. This can be achieved by manually delimiting the flanks of the region and selecting "*Annotate!*, a tool that records all the above mentioned information when the region is selected whilst listing it in a table (Figure 28).

If smoothing of the signal is desired, Gaussian filtering can be selected prior to visualization of the data. The Gaussian filter approximates the data to a Gaussian function, minimizing the rise and fall time, and reducing noise. The filter can be permuted as many times as desired, so that the higher the applied filter value is the smoother the signal looks.

This web application was originally created to meet the needs of the present study, and therefore was mainly focused on the analysis of tiling array data sets. To broaden the scope of the application, *TilingScan* was also made compatible for the analysis of data obtained through RNA-seq technologies via the "*Cover2Tiling" option*, which allows the conversion of NGS data (from a sorted BAM file) into the *TilingScan* input format.

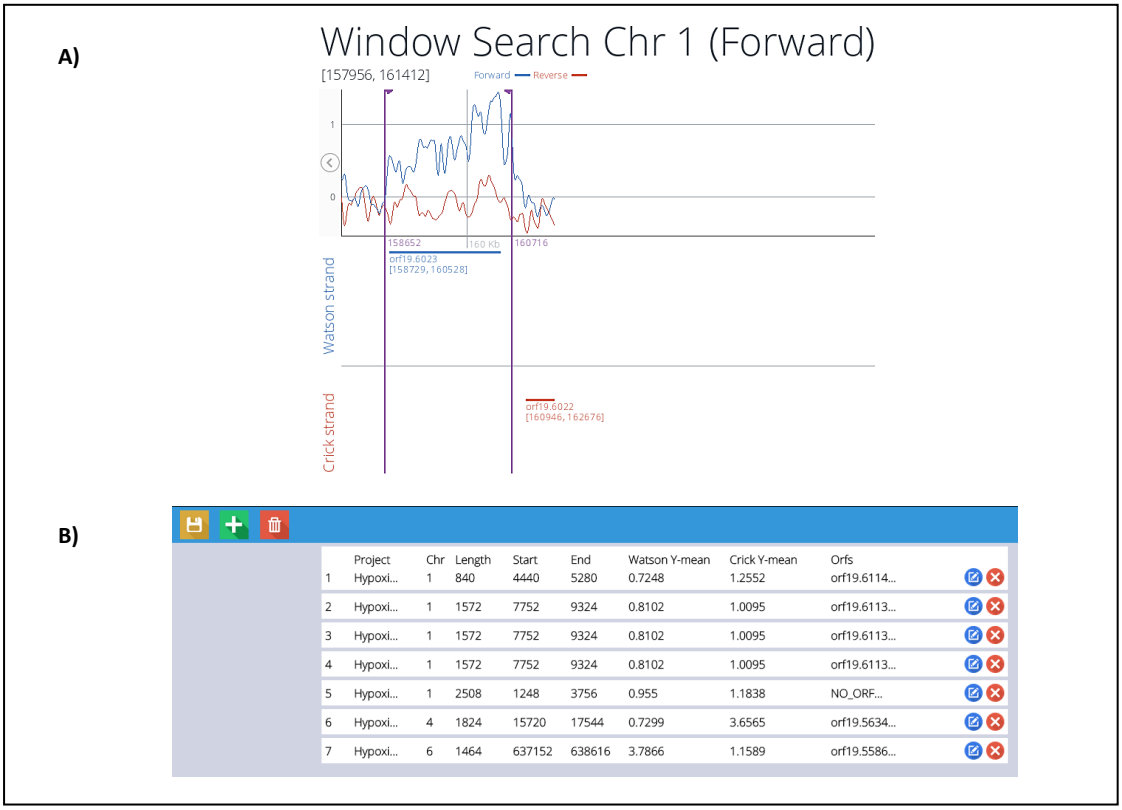


Figure 28. Outputs of the search for differentially expressed regions. A) Graphical output of a region detected on *C. albicans* chromosome 1 after 40 min treatment under oxidative stress. Intensity profiles on Forward (Watson) strand and Reverse (Crick) strand are shown in blue and red respectively. Start and end chromosomal coordinates of the whole visualized section and the detected region are shown in brackets (top, left) and by the edges of the delimiting purple vertical lines respectively. Intensity profiles are aligned to the selected reference genome at the bottom. **B)** Image section of the *Annotate!* tool displaying recorded features for some of the regions detected in *C. albicans* after 30 min treatment in hypoxia.

2.3. Use of the developed application for the detection of differential expression of antisense transcripts and other ncRNAs in response to the studied stresses in *C. albicans*

To date, several studies have addressed the transcriptional response to hypoxic and oxidative stress in *C. albicans*[192-194]. However, most of them focused on the changes observed on protein-coding genes, leaving other potentially interesting transcripts unexplored. Nowadays there is an increasing number of studies suggesting different types of ncRNAs to be involved in the regulation of specific stress responses in other yeast species[319-322]. Some of these studies argued that, in addition to reprogramming gene expression pertinent to the changing conditions, the presence of these transcripts could also lead to greater expression variability for antisense (AS)-containing genes, which in turn could help species to adapt to changing environments. Along with these findings, a recent genome-wide study[151] reported the presence of several AS transcripts as well as other ncRNAs in *C. albicans* under conditions that were relevant to its pathogenicity, such as biofilm formation and serum-induced hyphae. Using RNA-seq technology, other authors also reported the discovery of novel transcriptionally active regions during the exposure of *C. albicans* cells to different environmental stresses, among which oxidative stress was included[323]. All together these findings led us to hypothesize that some ncRNAs could be involved in the general stress response in *C. albicans*, and in particular in response to hypoxic and oxidative stress. To explore the putative existence of these transcripts in *C. albicans* upon these two stresses, we used the *TilingScan* software to explore the genome in search for all the differentially expressed regions in a strand-specific manner using the "*Windowsearch"* tool. Prior to the analysis, two data sets were created for each type of stress, in which the RNA amount measured in two different time points under stress were compared to the no stress control (Figure 29).

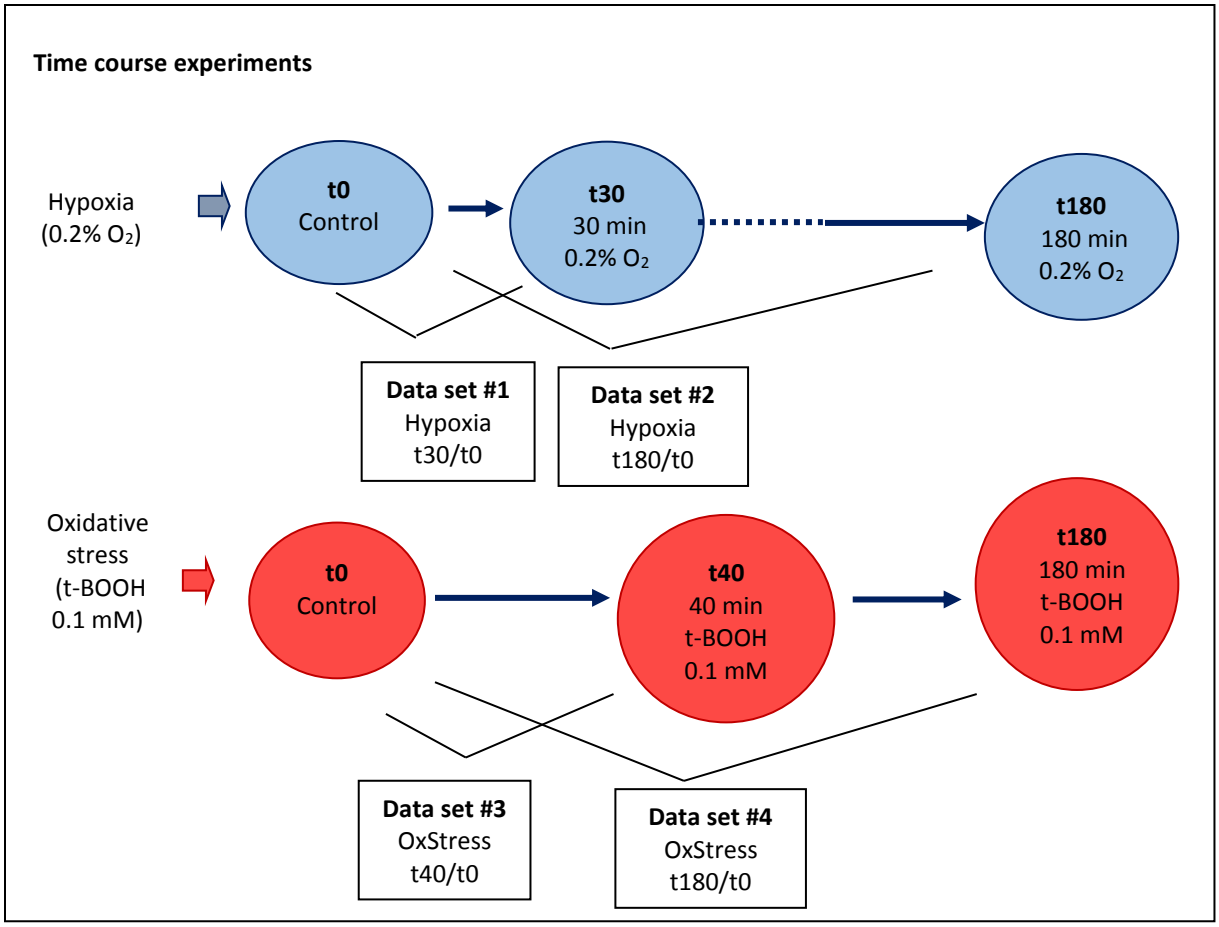


Figure 29. Samples utilized for the creation of data sets subjected to the search for antisense and intergenic lncRNAs. For each type of stress (oxidative and hypoxic), the two different time points under stress were compared to the no stress control (t0), yielding a final number of 4 different datasets that were individually subjected to the search analysis.

The resulting output was four different data sets, which were subjected to the search for the detection of differentially expressed regions. From the observed regions, we manually delimited and annotated those regions expressing transcripts which met the following criteria:

1. Were either antisense to annotated ORFs or transcribed from an intergenic region. In the case of intergenic transcripts, these regions were required to be at least 250 nt away from an annotated gene.
2. Were at least 200 bases in length (thus considered lncRNAs)
3. Had been either induced or repressed upon the treatment.
4. Had a fold-change higher than 2 (over at least a length of 200 bp).

As these transcripts were manually selected, and so as to avoid arbitrary delimitation of the start and end coordinates, start and end points were selected as those coordinates at which the signal intensity first surpasses (and then first drops below) the selected threshold cut-off value.

By using the search algorithm on the expression data sets obtained in this study (hypoxic and oxidative stress treatment in *C. albicans*) we discovered the existence of potentially interesting genetic elements, such as antisense (AS) transcripts and intergenic ncRNAs. Some examples are shown in Figure 30.

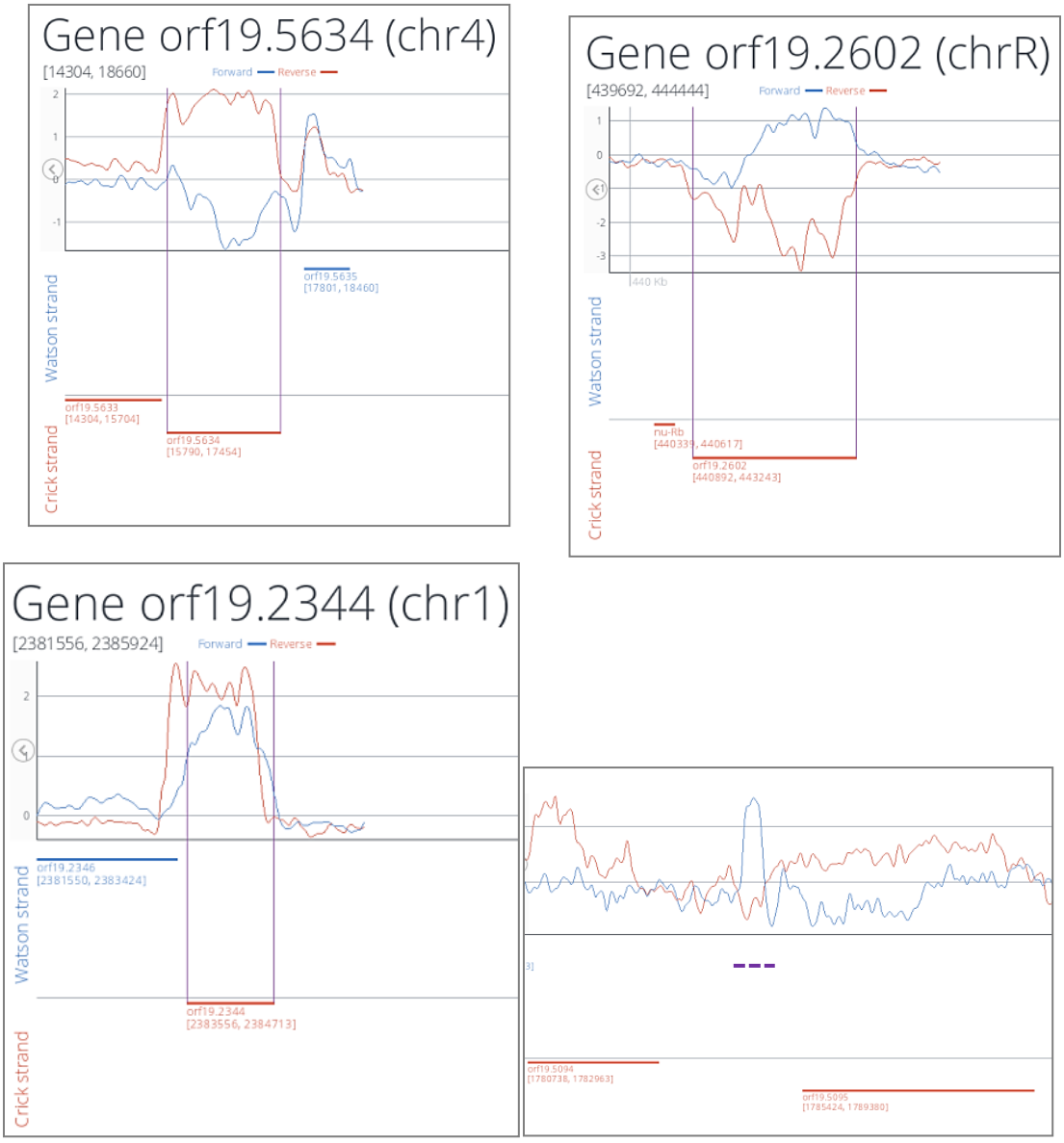


Figure 30. Examples of AS and intergenic ncRNAs found with the search algorithm of the developed application. Regions of differential gene expression detected after 180 min of treatment in hypoxia are shown. (Top, left) Example of AS transcript which expression was down-regulated upon the treatment, whereas its sense transcript was induced. (Top, right) Example of AS transcript which expression was up-regulated upon the treatment, whereas its sense transcript was down-regulated. (Bottom, left) Example of simultaneous up-regulation of both sense and antisense transcripts. (Bottom, right) Example of intergenic lncRNA detected in chromosome 1 (underlined in purple dotted line). Intensity values are shown in $\log_2$ scale.

The whole list of detected ncRNAs in each data set that followed these criteria is in Appendix II, and summarized in Table 7.

Data set	Total number of transcripts	AS transcripts	Intergenic lncRNAs
Hypoxia t30/t0	16	8	8
Hypoxia t180/t0	138	81	57
Oxidative stress t40/t0	141	84	57
Oxidative stress t180/t0	18	8	10

Table 7. Summary of the detected ncRNAs that met the selection criteria. For each data set analyzed, the total number of detected transcripts is shown. Those that were antisense to previously annotated ORFs are denoted as antisense (AS) transcripts, and those transcripts detected in intergenic regions are denoted as intergenic lncRNAs.

The number of novel expressed regions identified is low compared to that observed in other studies[151,323]. However, this was expected due to the high stringency selection criteria applied in this analysis. In addition, it should be noted that all RNA molecules analyzed in this study were subjected to a round of poly(A) RNA selection, hence the detection of non-polyadenylated transcripts was not expected. For the same reason, transcripts which short poly(A) tails may as well be underrepresented in these data sets. Considering all these factors together, it is likely that more differentially expressed ncRNAs exist and have been missed in this analysis. Yet, the decision was made to follow these criteria so as to be as confident as possible about the existence of the ncRNAs presented herein.

The highest number of AS transcripts or ncRNAs was observed in the data sets corresponding to the comparison t180/t0 in hypoxia and t40/t0 in oxidative stress, which coincided with the number of protein-coding genes which expression changed upon the treatment (data not shown).Thus, both sense and antisense transcription seem to change in the same manner. Previous studies remarked that sense and antisense expression can associate in different ways, according to the mechanism specified by the environmental condition: some may require the concomitant presence of sense and antisense transcripts, while others may work by reciprocal exclusion[321]. Regarding the possible effect of the expression of the AS transcript on the expression of the sense, and based on visual

inspection, no clear pattern was found in this study. In some cases, sense transcripts were up-regulated whereas their AS transcripts were down-regulated and *vice versa*; in some other cases the up or down regulation of the AS transcript was not accompanied by a change in the levels of the sense transcript (data not shown). Overall no clear correlation was observed that could lead us to draw conclusions on how AS transcription may be influencing the expression of the sense transcripts in our conditions of study.

In search for common elements, Gene ontology (GO) analyses were performed in all the above-mentioned data sets. The analysis of the AS transcripts detected in hypoxia showed no significant categories for the genes containing them. However, the 'component' search for the transcripts detected under oxidative stress showed statistical enrichment of the genes annotated under the terms cell wall (corrected p-value 2.65 x 10^{-5}), fungal-type cell wall (corrected p-value 2.65 x 10^{-5}), external encapsulating structure (corrected p-value 3.09x 10^{-5}), and cell surface (corrected p-value 9.5 x 10^{-5}). (Table 8)

Terms from the Component Ontology					
Gene Ontology term	Cluster frequency	Background frequency	Corrected P-value	False discovery rate	Genes annotated to the term
cell wall | AmiGO	12 out of 85 genes, 14.1%	142 out of 6473 background genes, 2.2%	2.65e-05	0.00%	ALS2, ALS4, C2_10320C_A, C7_00190W_A, CHT3, CRH12, ENG1, GPM2, PDC11, PGK1, SCW11, YWP1
fungal-type cell wall | AmiGO	12 out of 85 genes, 14.1%	142 out of 6473 background genes, 2.2%	2.65e-05	0.00%	ALS2, ALS4, C2_10320C_A, C7_00190W_A, CHT3, CRH12, ENG1, GPM2, PDC11, PGK1, SCW11, YWP1
external encapsulating structure | AmiGO	12 out of 85 genes, 14.1%	144 out of 6473 background genes, 2.2%	3.09e-05	0.00%	ALS2, ALS4, C2_10320C_A, C7_00190W_A, CHT3, CRH12, ENG1, GPM2, PDC11, PGK1, SCW11, YWP1
cell surface | AmiGO	12 out of 85 genes, 14.1%	199 out of 6473 background genes, 3.1%	0.00095	0.00%	ALS2, ALS4, CHT3, CRH12, ENG1, FGR41, PDC11, PGA26, PGA38, PGK1, SCW11, YWP1

Table 8. Gene ontology term finder results for the component search performed from the antisense-containing gene set in oxidative stress conditions. All genes found in the oxidative stress analysis containing an overlapping antisense transcript were subjected to the GO search, including in the list all together both genes found at time points 40 min and 180 min. The background set of genes was that provided by default in the *Candida* Genome Database.

Based on this observation, a *GO Slim mapper* search was performed with the transcripts detected in hypoxia, this time looking specifically at categories related to the cell surface. The search results showed that 16 out of 89 (18%) transcripts had also functions related to the cell surface. (Table 9). In general, these results indicate that under both types of stresses there is a small group of genes which function takes place in the cell surface that change the expression levels of their AS transcripts under stress. Although there is no evident explanation as to why this would happen, it makes sense that the number of these genes in the oxidative stress data set doubles the number of genes found in hypoxia, as oxidative stress is directly sensed in the cell surface whereas hypoxia is sensed internally.

Results for the mapping of 85 *Candida albicans* genes to the Candida GO-Slim Component

GO-Slim term	Cluster frequency	Genes annotated to the term
plasma membrane GO Term View	8 out of 85 genes, 9.4%	C1_12430W_A, ENA2, ITR1, OPT1, RBT5, RHD3, SSU1, TPO3
cell wall GO Term View	5 out of 85 genes, 5.9%	CYC3, DDR48, IFF5, RBT5, RHD3
extracellular region GO Term View	3 out of 85 genes, 3.5%	GDH3, RBT5, RHD3

Table 9. Gene ontology slim mapper results for the component search performed from the antisense-containing gene set in hypoxic conditions. All genes found in the hypoxic stress analysis containing an overlapping antisense transcript were subjected to the GO search, including in the list all together both genes found at time points 30 min and 180 min. The GO slim terms selected for the search were cell cortex, cell wall, plasma membrane and extracellular region. Obtained from the *Candida* Genome Database (CGD).

To search for common elements between this study and other previously published data sets we compared the entire set of AS transcripts detected under hypoxia and oxidative stress with those found in Sellam *et al.* 2010[151]. These authors reported the appearance of antisense transcripts under other stress conditions that are relevant to *C. albicans* pathogenicity, such as biofilm formation and serum-induced hyphae. Results showed that 12 out of the total number of putative AS transcripts detected in this work had been previously identified, 3 of them being present in all three studies. (Figure 31). Very little information was obtained from the identification of the ORFs in which AS transcription was

detected, since the majority of them arose from uncharacterized ORFs (Table 9). However, the majority of those with characterized function once again corresponded to proteins that are located in the cell surface, suggesting that AS transcription could somehow regulate the abundance of these proteins in response to environmental stresses. The reason these 3 data sets displayed such little overlap could be explained by the fact that changes in gene expression during stress responses in *C. albicans* seem to behave in a condition-specific manner, as previously reported by other authors[323,324]. Unlike in other yeasts such as *S. cerevisiae* or *S. pombe*, these comparative transcriptomic studies under different stress conditions in *C. albicans* showed that the commonly known as "core stress response genes" are only a very small proportion compared to those observed in the other two yeast species[325]. For this reason, even though some genes were expected to be similarly expressed in response to the different environmental stresses, this number would still be expected to be low.

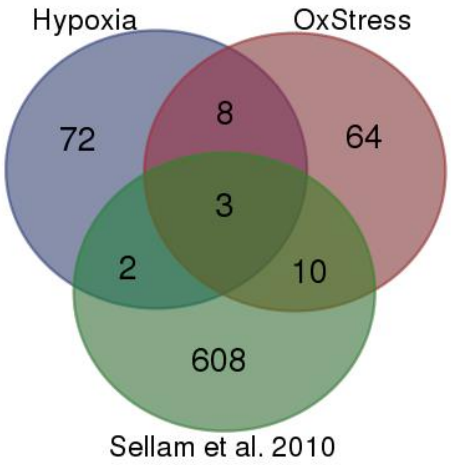


Figure 31. Venn diagram showing the common and unique AS transcripts of this study compared to those found in Sellam *et al* 2010[151].

Data Sets compared	Common elements	ORFs	Systematic ID	Function
Hypoxia OxStress Sellam *et al*. 2010	**3**	orf19.3475	C6_02330W_A	Uncharacterized
		orf19.164	CR_02570C_A	Uncharacterized
		orf19.7350	*RCT1*	Uncharacterized
Hypoxia OxStress	8	orf19.3329	*LCB3*	Uncharacterized
		orf19.5325	*KIN3*	Uncharacterized
		orf19.6349	*RVS162*	Uncharacterized*
		orf19.4450.2	C1_07150W_A	Uncharacterized
		orf19.600	*TRK1*	Potassium transporter
		orf19.1636	*STE50*	Uncharacterized**
		orf19.4789	*ALD97*	Uncharacterized***
		orf19.3517	CR_05460W_A	Uncharacterized
Hypoxia Sellam *et al*. 2010	2	orf19.5305	*RHD3*	GPI-anchored cell wall protein
		orf19.5636	*RBT5*	GPI-linked cell wall protein
OxStress Sellam *et al*. 2010	10	orf19.986	*GLY1*	L-threonine aldolase
		orf19.994	C1_10510W_A	Uncharacterized
		orf19.3966	*CRH12*	CRH family cell wall protein
		orf19.3651	*PGK1*	Phosphoglycerate kinase
		orf19.4555	*ALS4*	GPI-anchored adhesin
		orf19.2877	*PDC11*	Pyruvate decarboxylase
		orf19.5114.1	C1_08350C_A	Uncharacterized
		orf19.1097	*ALS2*	ALS family protein
		orf19.1067	*GPM2*	Putative phosphoglycerate mutase
		orf19.3618	*YWP1*	Secreted yeast wall protein

* Protein containing a BAR domain, which is found in proteins involved in membrane curvature

** Protein with sterile alpha motif (SAM) and Ras-associated domain (RAD); similar to S. cerevisiae Rad50p, which is involved in signal transduction via interaction with and regulation of MAPKKK

***Has domain(s) with predicted metal ion binding activity

Table 9. List of genes containing overlapping antisense transcripts common to two or more of the environmental stresses studied. The genes found to contain an overlapping antisense transcript under hypoxia and oxidative stress in this study were compared to the list of antisense-containing genes described in Sellam *et al.* 2010[151] aiming to search for common elements in all three studies. The comparison analysis was done using an online Venn diagram tool at (http://bioinformatics.psb.ugent.be/webtools/Venn).

2.4. Discussion and future perspectives

One of the biggest surprises of the post genomic era has been the enormous number and diversity of transcriptional products arising from previously presumed 'wastelands' of the non-protein-coding genome. These include a plethora of small regulatory RNAs and tens of thousands of polyadenylated and non-polyadenylated lncRNAs that are antisense, intronic, intergenic and overlapping with respect to protein-coding loci[326].

The functions of these transcripts are largely unknown, although there is increasing *in vitro* and *in vivo* evidence that lncRNAs have key roles across diverse biological processes, with an emerging theme of interfacing with epigenetic regulatory pathways[327,328]. Thus, the sheer number and the increasing pace of discovery of new lncRNAs are accompanied by the growing challenge of their definition and annotation.

The broad term "lncRNA" refers to a transcript>200 nt in length that does not appear to contain a protein-coding sequence[328]. The size threshold is an arbitrary biophysical cut-off that excludes most known, although still poorly understood, classes of small known RNAs, such as tRNAs, snRNAs, snoRNAs and their derivatives, microRNAs, short interfering RNAs, Piwi-interacting RNAs and small RNAs that regulate splicing[326]. It also includes most of the cryptic unstable ncRNAs that are usually very short because of the co-transcriptional surveillance mechanisms[329]. Although these transcripts may be collectively referred to as 'noncoding' RNAs, this denotation is often temporary until they are better characterized. Indeed, a small percentage of transcripts originally annotated as lncRNAs have later been found to encode new proteins[330,331]. Some lncRNAs that are relatively highly expressed appear to be structural and functional scaffolds for specialized sub-nuclear domains, whereas others with more restricted expression have been found to be associated with chromatin-modifying complexes and to regulate alternative splicing[332,333].

Tiling microarray technologies allow the interrogation of the entire genomic sequence that is available for a given organism. In this sense, they enable the exploration of both gene-containing and intergenic regions, yielding transcript information of either predictable or non-predictable ORFs. Given the possibilities of study these technologies offer and the

recent discovery of the pervasive presence of long non-coding transcripts, it is important to develop bioinformatics tools that foster their exploration.

To meet this need, *TilingScan* was developed as a web application that allows the detection and annotation of differential gene expression in an unbiased manner[317]. By applying a newly implemented search algorithm, the application scans the genome in search for differentially expressed DNA regions, whether they are located in previously annotated genomic features or not. Using this tool, we scanned the *C. albicans* genome in search for putative lncRNAs which had been either up or down regulated upon the two stresses applied in this study: hypoxic and oxidative, under the hypothesis that they could play a role in the adaptation of the organism to these environmental insults.

This search yielded a list of lncRNAs which number was interestingly correlated with the overall number of protein-coding genes which expression changed after the treatment. In addition, some of the detected transcripts had already been found in a previous study where other authors reported the observation of antisense transcription in *C. albicans* cells that had been subjected to other types of stresses[151]. This could suggest that the transcripts found in this study, or at least those that were found in both studies, may play a general role in the response to stress in this organism. Supporting this idea, some studies in *S. cerevisiae* already proposed that AS transcription could be important in a stress-related condition[147,148,334], although the relative contribution of these to the adaptation to stress still remains controversial[335]. Unfortunately, the lack of this type of studies in *C. albicans* makes the comparison between studies in this organism still not possible.

Even though this analysis provides more evidence on the existence of antisense transcription in *C. albicans*, further experiments will be needed to substantiate their potential function, if any. Moreover, the transcripts found in this study via genome-wide approaches will now have to be experimentally validated using single-gene techniques. Once validated, it will be necessary to address whether the appearance of these transcripts is dependent on the presence of the agent (or condition) causing stress, and not just due to the particular growth conditions.

Finally, given the current importance and gradual increase of the use of RNA-seq technologies the website hosting the *TilingScan* software was made compatible with them by creating a BAM file convertor that allows the conversion of RNA-seq data into the tiling scan input format. This broadened the scope of the application and was of great importance as now the same type of search can be made in RNA-seq data sets. Provided that many ncRNAs are just now being discovered, the sequencing approach will probably be more suitable for these type of studies as the obtained data sets can be re-analyzed several times as these elements are added to the current genomic annotation of *C. albicans*.

Chapter 3

Optimization of the GRO technique for coupling with RNA-seq technologies in *C. albicans*

3.1 Motivation

As detailed in the introduction, the radioactive approach of the GRO technique has been successfully used over the years in the model yeast species *S. cerevisiae*[86,162,163,165,166]. After over a decade of use, it has always been the usual approach of choice in our laboratory, as big amounts of data have been generated in several yeast strains and experimental conditions that can be used for comparison as new results are obtained. With the arrival of next-generation technologies, the technique was made compatible with tiling array platforms in the new *BioGRO* version, which has already yielded useful information on basic aspects of RNApol II dynamics at higher resolution[159]. Yet, this technique suffers from lack of sensitivity. To this date, the current BioGRO approach only allows the acquisition of reliable TR data for high TR genes, which in *S. cerevisiae* only account for the 12% of the genome approximately[162]. The reason for such low number, despite still unknown to us, is likely to be due to the relatively narrow dynamic range of tiling array platforms, which display loss of signal at the low end failing to detect transcripts that are in low abundance. Due to this limitation, in order to broaden the scope of the technique as well as to obtain TR data for the whole genome of *S. cerevisiae*, the implementation of the BioGRO approach was aimed to be made suitable for sequencing technologies (for this purpose, RNA-seq). This approach has been implemented in other organisms, such as human cells[84,336], *Drosophila*[337] and *Arabidopsis*[338], but not in yeast cells. In parallel to this, and provided that none of the variants of the GRO technique had ever been used to study other yeast species, the most up-to-date version of the GRO achieved to this date was made compatible for the use with *C. albicans*.

3.2. Comparison of run-on efficiency in two different yeast species: *C. albicans vs S. cerevisiae*

As previously mentioned, GRO and all of its variants have been successfully used in *S. cerevisiae* over the years in our laboratory. One of these variants, the Filter-run on (*MiniGRO*) is performed prior to any GRO or BioGRO assay to assess the labeling efficiency that can be achieved with the specific type of cells that are going to be used in the

experiment. This efficiency is measured as the amount of ^{33}P- UTP that is incorporated into the nascent RNA chain upon the addition of the nucleotide pulse. In this sense, the labeling efficiency is defined in terms of "Percentage of incorporation", where the amount of labeled RNA is referred to as "Incorporated" (that is, integrated in the nascent molecule), and the total amount of ^{33}P-UTP present in the reaction mix, incorporated or not, is referred to as "Total". Based on these two parameters, the percentage of incorporation is then calculated as: (Incorporated/Total) x 100 (Figure 32).

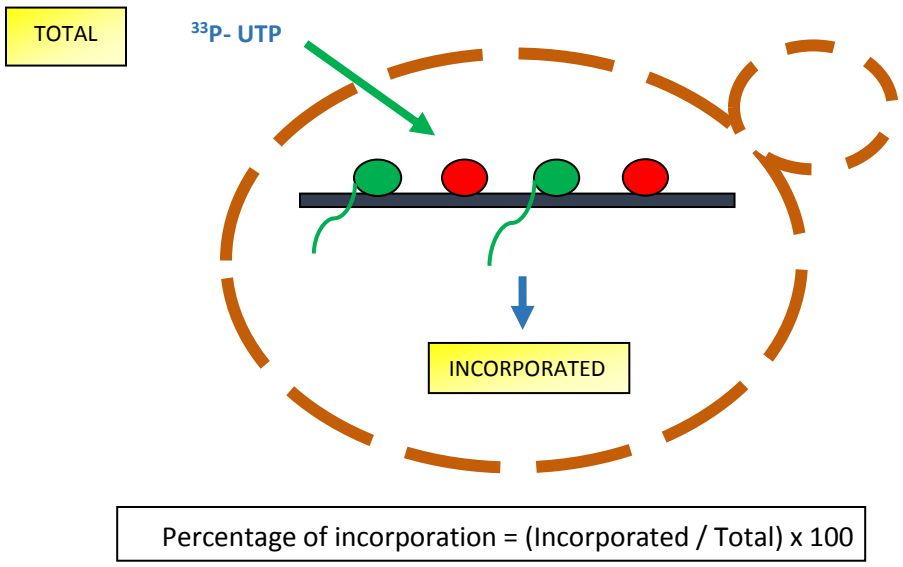


Figure 32. Parameters measured to address run-on efficiency. The amount of labeled RNA is referred to as "Incorporated" whereas the overall intake of ^{33}P-UTP in the cell, incorporated or not, is referred to as "Total". The percentage of incorporation is then calculated based on these two parameters using the equation described above.

The reason the labeling efficiency is always firstly measured using the radioactive MiniGRO approach is that, by doing so, quantitative (rather than qualitative) data can be quickly and inexpensively obtained by measuring the radioactivity of the sample in a scintillation counter. Once confirmed that the radioactive approach works in the conditions or organisms of study and that the labeling efficiency obtained is enough to then acquire the level of sensitivity desired, the BioGRO approach is subsequently used to obtain gene-specific data.

Over the years it has been observed that the percentages of incorporation obtained in a MiniGRO assay usually differ from one yeast strain to another, as well as from one experimental condition to another. For instance, mutant strains lacking some of the RNA pol II subunits that as a consequence are defective in transcription tend to yield lower incorporation values overall (Daniel Medina, personal communication).

In our hands, it has been well established that using the MiniGRO technique in *S. cerevisiae*, the percentage of incorporation that is usually achieved ranges from 3% to 10%. It was later on observed that even when obtaining the highest percentage of incorporation possible (10%) the number of genes that yielded reliable TR data by means of BioGRO studies (using biotin-UTP instead of ^{33}P-UTP) were only 809 genes of the *S. cerevisiae* genome. This number accounts only for 12% of this organism's genome, so despite very informative, it is still far from the 'genome-wide' feature that was aimed to be achieved with the implementation of this technique. This meant that, for the BioGRO technique to be used for *C. albicans*, at least the same labeling efficiency as that obtained for *S. cerevisiae* was needed. To measure the percentage of ^{33}P-UTP incorporation in *C. albicans* and so as to compare it to the one obtained in *S. cerevisiae* a MiniGRO experiment using *C. albicans* cells was undertaken. For this experiment, the same number of cells of *S. cerevisiae* was used as the experimental control, these cells being grown in the same experimental conditions as *C. albicans*.

Because all the previous data obtained in our laboratory for *S. cerevisiae* had been acquired in cells grown at the standard growth temperature for this organism (28-30 ^{o}C), the temperature at which *C. albicans* cells were grown for comparison was the same. However, as the final aim of this work was to study the transcriptional features of *C. albicans* at its physiological temperature (37 ^{o}C), and no previous reference existed on whether a change in cell growth temperature could yield different incorporation values, the MiniGRO tests performed to compare labeling efficiency in both yeast species were done at both cell growth temperatures, 30 ^{o}C and 37 ^{o}C (Figure 33).

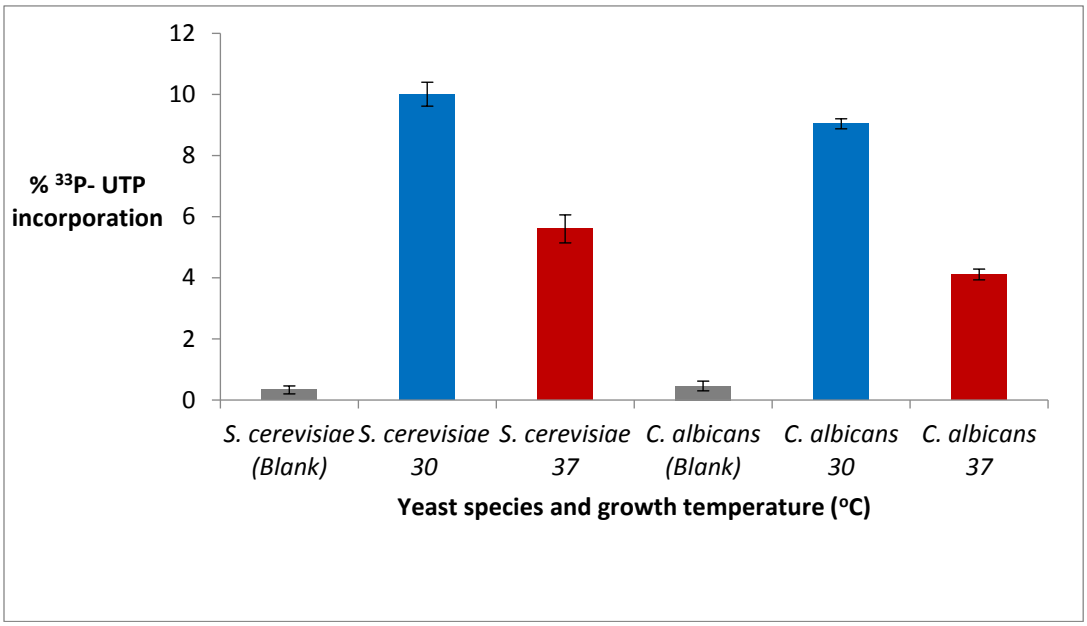


Figure 33. Percentage of ^{33}P- UTP incorporation in *C. albicans* compared to *S. cerevisiae* at two different growth temperatures. Bars show average values of incorporation in both yeast species grown at 30 and 37 ^{0}C. Blanks refer to those samples in which cells were not treated with detergent, which are used to assess the background noise of the measurement. Experiments were performed in three biological replicates, each replicate being done in technical triplicates.

The results in this experiment yielded two main findings. Firstly, even though all incorporation values were found to be within the expected range of labeling efficiency, the average value in *C. albicans* was slightly lower (18% less on average) compared to that in *S. cerevisiae* at both growth temperatures. This implies that the sensitivity of the BioGRO technique in *C. albicans* is likely to be even lower, or in the best case equal to the one achieved in *S. cerevisiae*.

Secondly, it was observed that the labeling efficiency decreased in both yeast species when cells were grown at 37 ^{o}C compared to 30 ^{o}C. In both species the percentage of incorporation in cells grown at 37 ^{o}C was only 60% of that observed in cells grown at 30 ^{o}C. This second finding raised the question of whether, provided that GRO assays reflect density of polymerases at a global scale, RNA pol density could vary with growth temperature. This question is addressed in chapter 4.

3.3. Run-on assays from frozen cells

Nowadays, many laboratories collaborate in order to distribute the large amount of work that a research project carries. Often, these laboratories are located in different countries. For this reason, it is of great utility for some of them to be able to freeze and then deliver biological samples. The present study was carried out within the framework of an international collaborative project that required *Candida* cell samples to be processed and collected in a foreign laboratory and delivered to ours, for the GRO experiments to be performed at the University of Valencia. For shipping, cells had to be frozen after harvesting, which implied that the GRO experiments planned for the project should be able to be done from frozen cell pellets. Until the current project was proposed, all GRO experiments performed in our laboratory had been done from freshly collected cell samples, and no previous evidence existed on whether this assay would work using frozen cells as starting material. To test this, several MiniGRO assays were undertaken in the same manner as described in section 3.2 of this chapter, this time adding an additional aliquot for both yeast strains, which had been flash-frozen in liquid nitrogen immediately after harvesting. For the purpose of this experiment, freshly collected cells were used as the experimental control. The percentages of incorporation obtained in each experimental condition for each yeast species are shown in Figure 34.

The results observed in this experiment yielded three main findings. Firstly, it can be observed that not only GRO assays can be performed using frozen cells as starting material, but also that the labeling efficiency increases by 60% on average in both yeast species when the cells are previously frozen compared to when they are freshly collected. This is of great advantage to us as it enables GRO samples to be delivered from one laboratory to another allowing the improvement of the labeling efficiency in a simple step. In addition, it can shorten and facilitate cell sampling. This is because when several samples have to be analyzed in the same GRO assay, cells can be collected one day, frozen and stored, and then processed all at once on a different day if desired. A good example of this is when different time points of a time course experiment have to subjected to the run-on pulse at different

times which differ by several hours. Prior to this finding, the researcher would have had to wait for the cells to grow, be harvested and run-on-pulsed during the entire duration of the time course, whereas this way all samples can be collected one day and then subjected to the pulse and subsequent RNA extraction all at once on a different day, reducing the time required for the experiment and facilitating the pulse itself. Provided this phenomenon was observed, all GRO experiments described in the following sections of this study were done using frozen cell samples as starting material, both in *S. cerevisiae* and *C. albicans*.

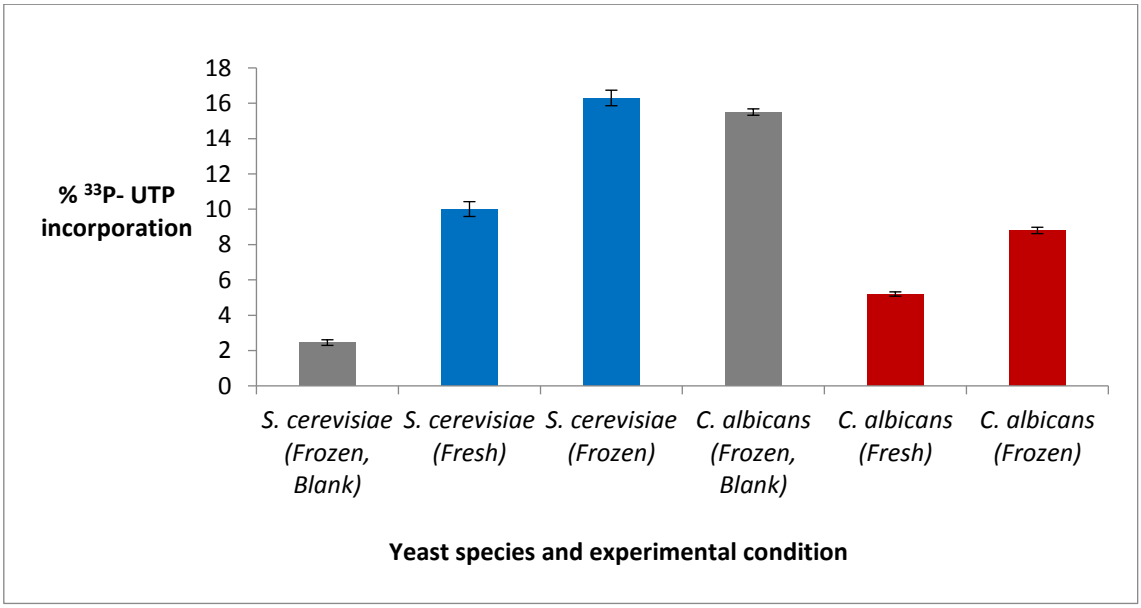


Figure 34. Percentage of ^{33}P- UTP incorporation in freshly harvested and frozen cells samples of *C. albicans* and *S. cerevisiae*. Bars show average values of incorporation in both yeast species when grown at their correspondent physiological growth temperatures (30 and 37 ^{o}C for *S. cerevisiae* and *C. albicans* respectively). Blanks refer to those samples in which the cells had not been treated with detergent, used to estimate the background noise of the measurement. These blanks were done using frozen cells in both cases. The experiment was performed in three biological replicates, each replicate being done in technical triplicates.

Another interesting feature observed in this experiment was the values of incorporation obtained in the blanks. In a MiniGRO assay, the blanks are experimental controls that are added alongside the experimental samples in order to evaluate the background noise of the measurement of the ^{33}P –UTP incorporation. Blank samples contain cells that have not been treated with detergent (Sarkosyl), therefore, presumably, are not made permeable to the nucleotide pulse. Since making the cell permeable is the step of the protocol that allows

labeled nucleotides into the cell, one would expect to observe no incorporation in the blanks. That was indeed what occurred in the experiment shown in Figure 33 when cells were freshly collected. Interestingly, however, in the case of the experiment shown in Figure 34 where all cells used had been previously frozen, the values of incorporation observed were far above the ones obtained in Figure 33, indicating that in this case incorporation was occurring. This somehow implies that the cell surface structures were affected by the freezing procedures these cells were subjected to, suggesting freezing in liquid nitrogen to be an aggressive treatment that disrupts both cell wall and membrane allowing nucleotides into the cell. Given these results, it can be concluded that GRO experiments can be performed, especially in *C. albicans*, without the use of detergent, so long as the cell is made permeable by other means.

Finally, it can also be observed that in the case of *C. albicans* when cells are previously frozen the percentage of incorporation obtained in the blank is even higher than the one obtained in cells that had not been frozen but just treated with detergent. This questions whether the effect of Sarkosyl in this species is different from the one usually provoked in *S. cerevisiae*. As discussed in the introduction, Sarkosyl is a detergent that has several effects on the cell: First, it makes the membrane permeable, stopping all physiological processes at the moment it is added; and second, it disrupts chromatin structure by washing away nucleosomes, facilitating the elongation of polymerases[156]. An additional effect of this detergent that has been reported by several authors is that it prevents new transcription initiation events from occurring[339,340]. When evaluating the percentage of incorporation during a run-on reaction, these effects would work as opposing forces: the two firstly mentioned would result in an increase of the overall incorporation, whereas the last one would decrease it (Figure 35).

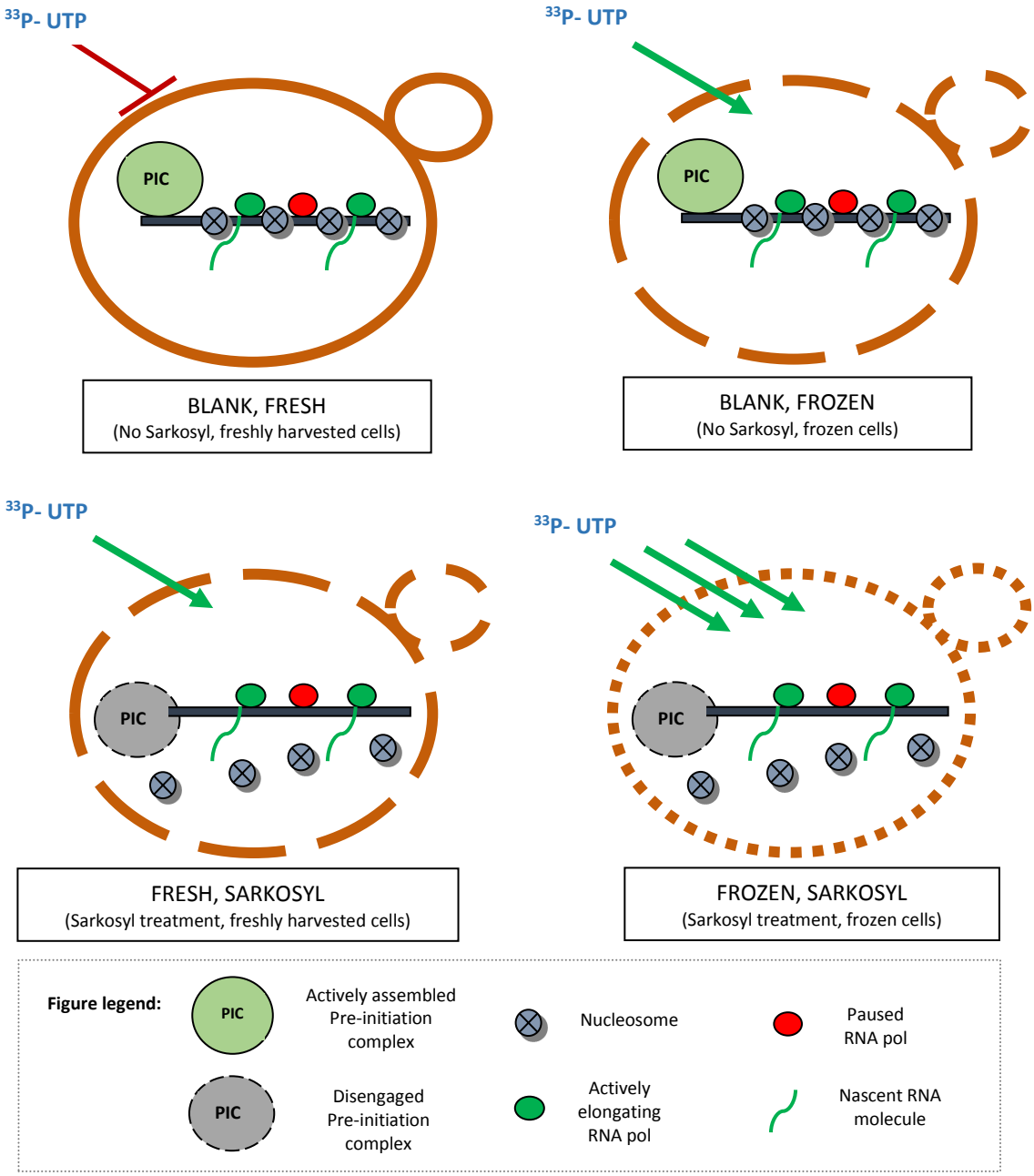


Figure 35. Schematic modeling of ^{33}P- UTP incorporation under different experimental conditions. (Top, left) When Sarkosyl is not added, cells are not made permeable thus no incorporation occurs. **(Top, right)** Freezing disrupts the cell surface making the cells permeable to a certain degree, which allows nucleotides in. **(Bottom, left)** When Sarkosyl is added, cells are made permeable, chromatin structure is disrupted facilitating elongation of polymerases, and PIC complexes are washed away from promoters preventing re-initiation events from occurring. **(Bottom, right)** When cells are frozen and Sarkosyl is added, cells are made permeable (presumably more than when only Sarkosyl is added), chromatin structure is disrupted facilitating elongation of polymerases, and PIC complexes are washed away from promoters preventing re-initiation events from occurring. The overall contribution of all these effects together may explain the differences in incorporation observed. (Note that the figures represent the average of cell population. It is unknown whether the sum of effects caused by sarkosyl and freezing improves cell permeability in all cells or, alternatively, increases the percentage of permeable cells).

In the case of the blanks used here, when Sarkosyl is not added the negative effect should presumably not take place: not only elongating polymerases resume transcription, but also new initiation events may be occurring, increasing the overall incorporation value. This could explain the value of incorporation obtained in the case of the blanks of *C. albicans*, however, it should also be expected to occur in *S. cerevisiae*, and this was not the case. This difference between yeast species may suggest that the overall sum of the two opposing effects of Sarkosyl differs from one species to the other. This being the case, the reason for it remains unkown.

When performing a run-on assay one of the main goals is to achieve the highest labeling efficiency possible. This is because it has been proven over the years that in general terms, the higher the labeling achieved the better the sensitivity of the technique is, resulting in a more informative experiment. Since the highest incorporation value in this experiment was obtained when cells were frozen and not treated with detergent, these results would suggest that this should be the preferred treatment *Candida* cells should be subjected to. This stated, in the absence of Sarkosyl during a run-on experiment new transcription initiation events could take place, hence the final assessment of RNApol activity would not reflect actively elongating RNA pols exclusively, but also potentially those that start a new cycle at the moment of the pulse. In this sense, the use of Sarkosyl in a GRO experiment is essential to avoid this potential artifact, and thus Sarkosyl treatment should continue to be the option of choice despite the improvement its absence would imply.

3.4. Comparative GRO analysis: freshly harvested vs. frozen cells.

Once known that cells can be frozen, stored and used later on for run-on assays, it was needed to ensure that this freezing procedure, soaking the cell pellet in liquid N_2 upon harvesting, would not affect the transcriptional profile (TP) that would be observed if the cells are given the usual treatment (freshly harvested the processed straight away). This is an issue to address since the aim of this work was to study transcriptional features in the

natural physiological state of the cells, which is achieved by limiting cell manipulations to the minimum number of steps necessary for the purpose of each experiment, among others. Since the use of liquid N_2 is an instantaneous freezing procedure, it seemed reasonable to think that no change in the TP should be observed when cells are frozen this way, presumably because the cells are not given any time to respond to the 'cold shock' in such short time lapse. That being true, the TP of cells frozen using this procedure should almost completely reflect the one observed in freshly collected cells. To explore this, a GRO experiment was designed, in which two different states of the cells were compared: frozen versus freshly collected (Figure 36).

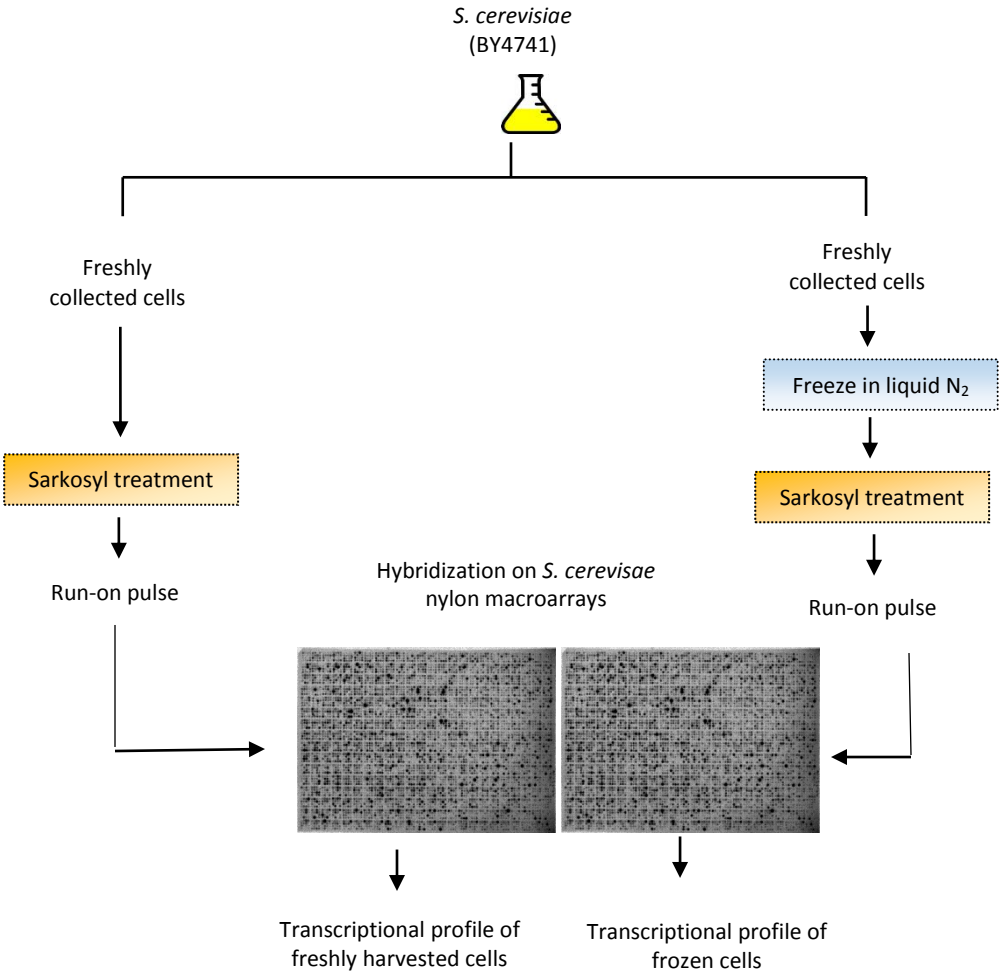


Figure 36. Experimental design of the GRO experiment performed for the comparison of freshly collected and frozen cell samples. Fresh cells were used straight after harvesting (**left**), whereas frozen cells (**right**) were collected, frozen in liquid N_2 , stored at -20 ^{o}C for at least 2 hours and processed later on.

Because no macroarrays are available for *C. albicans* (that is, DNA chips that are compatible with radioactive labeling), and it had already been proven that the overall sensitivity that could be possibly achieved with the BioGRO approach would not yield enough information for the purpose of this study (see section 3.2 of this chapter), this experiment was performed using *S. cerevisiae*, for which macroarrays were readily available in our laboratory. The hybridization signal of both samples was processed and quantified yielding relative TR values for a final number of 5430 genes of the *S. cerevisiae* genome that were plotted for comparison between experimental conditions (Figure 37).

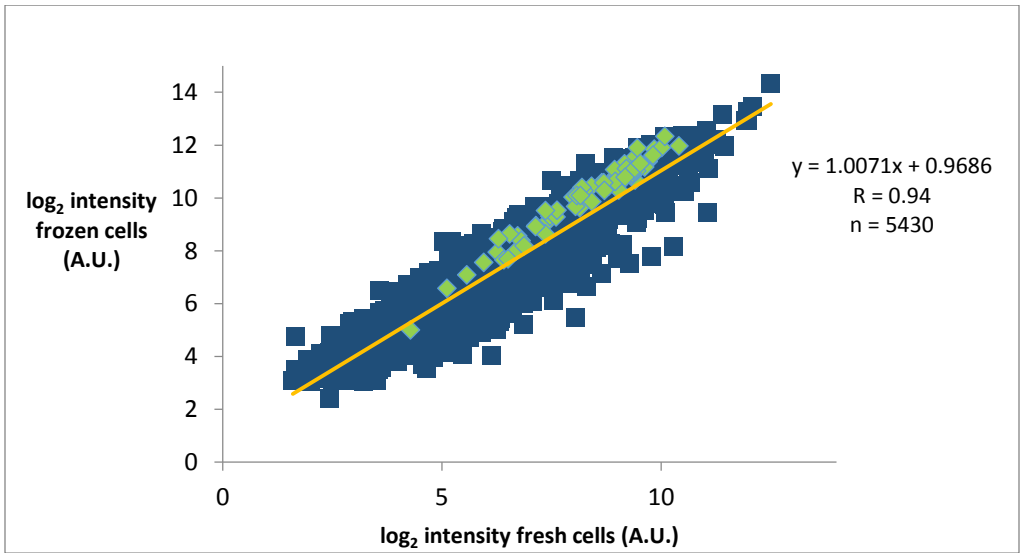


Figure 37. Correlation plot of the comparison between the transcriptional profile of freshly collected and frozen cell samples. The average GRO signal obtained from 2 independent biological replicates of the experiment is shown for freshly harvested cells (X axis, experimental control) and cells that were frozen before the experiment (Y axis). After data filtering a total number of 5430 genes were plotted for comparison. The average signal of ribosomal protein genes (shown in green) was higher in frozen cells, as seen in their tendency towards the Y axis.

The correlation coefficient observed (R = 0.94) indicates that freezing the cells implies no major disturbance of the TP with respect to the traditional treatment of processing the cells straight after harvesting. Even though a value of R = 0.94 would not seem to reflect total similarity between conditions, it should be noted that such correlation value is within the

expected technical error range usually obtained through the use of macroarrays. For reference, due to technical variability, the highest correlation coefficient that can be achieved between one sample and itself hybridized onto a different macroarray is R =0.98. In fact, when the technical replicates of the herein presented experiment were compared, the highest correlation coefficient observed was the same (R = 0.94, data not shown). Therefore, it can be concluded that freezing the cells after collection does not affect the purpose of a GRO experiment, which is studying the TP of the cells under a given condition. And for this reason, frozen cells can be used for GRO assays the same way freshly collected cells have been used over the years.

However, when comparing the two transcriptional profiles obtained in search for differences between them, two minor differences were detected. Firstly, the signal intensity of the whole population of ribosomal protein genes (RP genes) was slightly higher in frozen cells (Figure 37, green dots). This probably reflects that these genes are repressed instantaneously when the stress response is triggered as a consequence of the manipulation of the cells. The flash-freezing procedure limits the stress caused by the manipulation. And secondly, when representing the gene signal intensity in respect to its relative TR value (using the data set of TR values published in Pelechano *et al*. 2010[162]) it could be seen that the signal of very low TR genes (relative units 0 to 15 approximately) was also higher in frozen cell samples (Figure 38). This could just be a simple artifact, or may indicate that freezing the cells has a positive effect on elongating polymerases that is for some reason enhanced in this particular group of genes. These genes, however, are usually the least informative in most experiments because their signal is usually barely over the background signal. Therefore, the potential importance of the bias these genes could introduce for comparisons is very low.

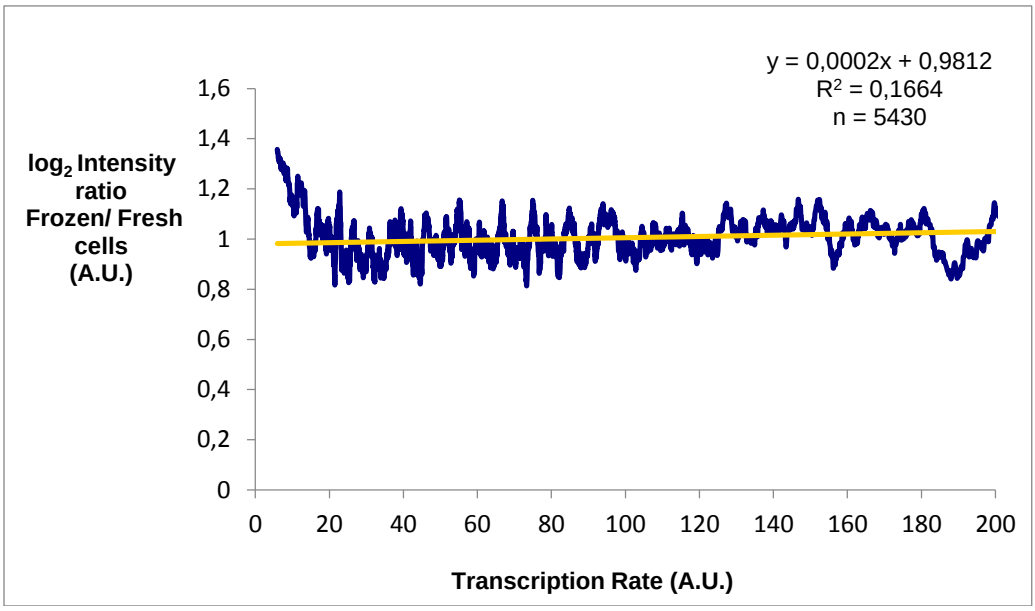


Figure 39. Plot of the ratio between the average signal intensity of the frozen cell sample compared to freshly collected cell sample in respect to each gene TR value. Signal intensity ratio of frozen cells compared to fresh cells is shown in respect to the TR value described in Pelechano et al. 2010[162], using average sliding windows of 50 elements. While the overall profile for most genes is relatively flat indicating no change between conditions, the ratio value of those gene with TR comprised in between 0 and 15 A.U approximately was higher in frozen cells compared to fresh cells.

Given these two last observations and despite the general TP is mostly not altered, it can be concluded that fresh and frozen cell samples should not be mixed in the same experiment, as, by doing so, some slight differences between samples may be seen which the observer may not be able to unambiguously assign to a neither biological nor technical reason. When comparing different cell samples of the same experiment, all samples should be processed the same way (either freshly collected or frozen) but never mixed.

3.5. BioGROseq: Adaptation of the BioGRO technique for coupling with RNA-seq technologies

As mentioned in previous sections, the lack of sensitivity of the tiling array-based BioGRO approach limits the number of genes which expression changes can be studied to only a 12% of the *S. cerevisiae* genome, particularly only those displaying high TR[167]. Given this

limitation, it was made clear that the technique had to be made compatible with RNA-seq technologies for the rest of the lower TR genes to also be explored.

Since the ultimate aim of a GRO experiment is the study of the 'nascentome' (that is, the pool of nascent RNA molecules present in the cell at the time of the extraction) a sequencing approach will only be adequate if this particular population of RNA molecules is isolated from the rest of the RNA present in the cell, so it can be independently sequenced later on. In order to isolate the population of nascent RNAs, a protocol for isolation of biotinilated RNA was used, as described in Kwak *et al*. 2013[341]. Briefly, the entire pool of RNA molecules that is extracted after the run-on pulse is subjected to a round of Biotin- UTP purification through the use of magnetic beads that are coupled to Streptavidin molecules. By doing so, mature RNA molecules are depleted while nascent molecules are retrieved after elution of the biotinilated RNA from the beads. This purified nascent RNA is then subjected to library preparation and sequencing (Figure 40).

Despite the low sensitivity of the tiling array-based approach of BioGRO, the main advantage of it is that direct hybridization of the sample avoids the introduction of potential mature RNA noise by sample amplification protocols. This, in turn, is the main disadvantage of the sequencing approach. For this reason, it became necessary to check whether the sequencing approach would yield the same results as the tiling array-based one. Once again, the final aim of the set-up of the BioGROseq in the work presented herein was to use it in *C. albicans*. However, since the only previous tiling array BioGRO data was exclusively available in *S. cerevisiae*, the experiment to test this was made using this yeast species. The experiment was done using the *S. cerevisiae* strain BQS252, which is the strain the tiling array data was available in, following the experimental pipeline presented in Figure 40. Once the sequencing data was obtained, these data were plotted against the tiling array data available for the same strain. The correlation coefficient obtained when comparing both techniques was R = 0.77, indicating a good correlation between the both techniques, even better than expected[153] (Figure 41A).

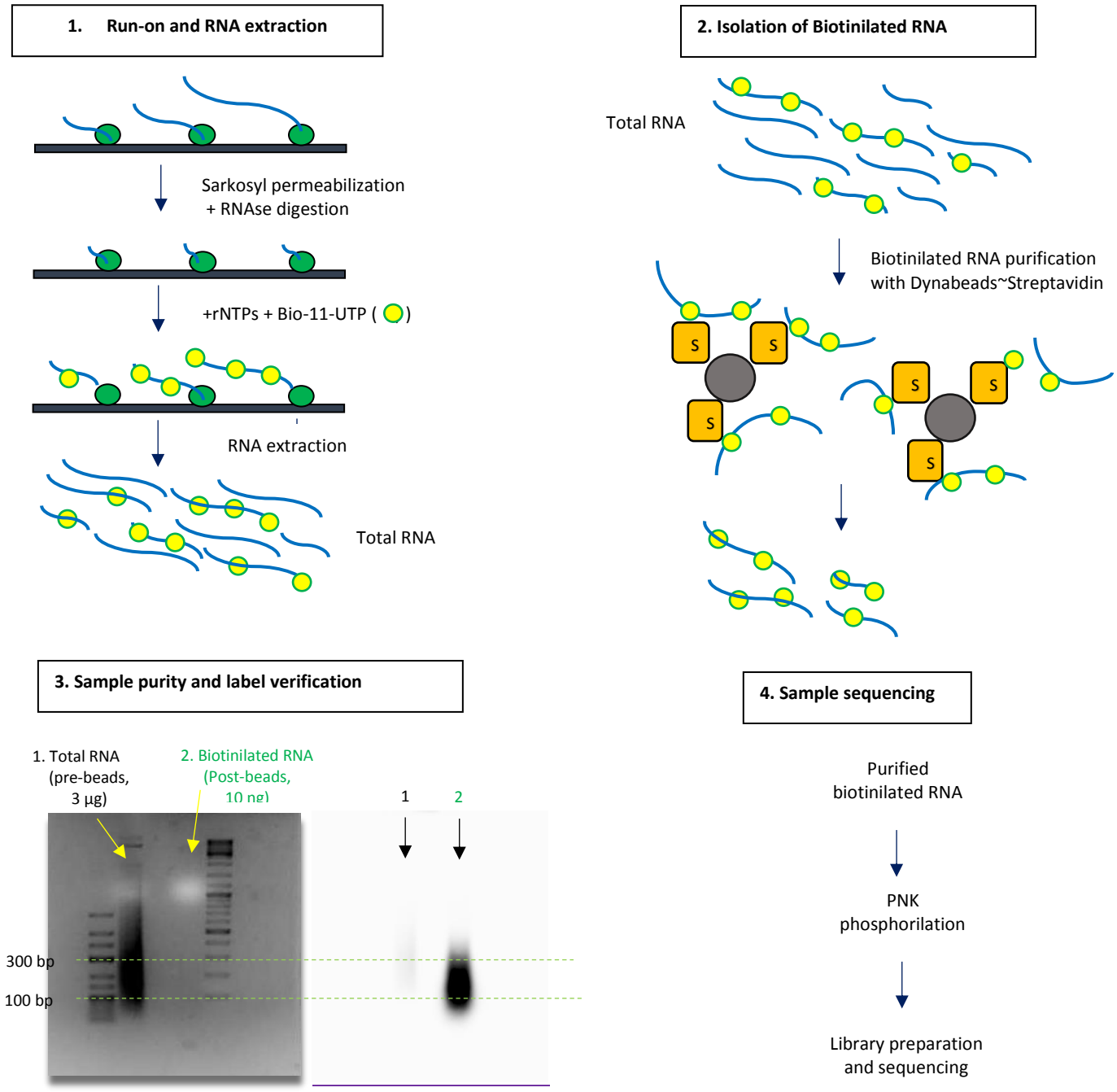
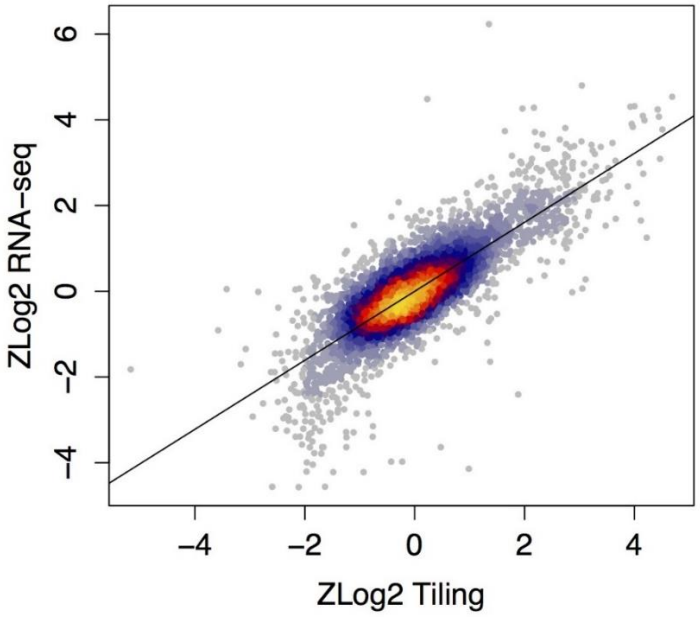


Figure 40. Experimental pipeline of a BioGROseq experiment. (Top, left). Prior to the run-on pulse, cells are treated with RNAse in order to deplete the pre-existent RNA molecules in the cell. After the pulse, all RNA content is isolated, including both mature RNA and the newly created (nascent) one that is biotinilated. **(Top, right)** In order to enrich the sample in the nascent RNA, biotinilated molecules are subjected to a round of selection through incubation with dynabeads coupled with steptavidin. **(Bottom, left)** Upon purification, the presence of the biotin label in the nascent RNA is verified along with its size. Generally, the size of the nascent RNA obtained ranges betweem 100 and 300 bp on average. **(Bottom, right)** Once the quality of the sample is verified, it is then phosphorilated and subjected to library preparation and sequencing.

A)



B)

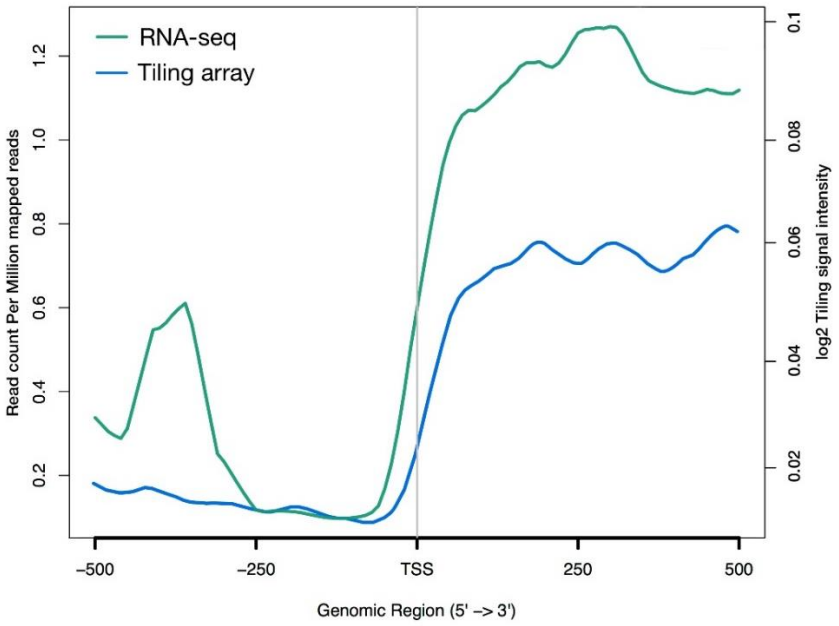


Figure 41. Comparison of BioGRO and BioGROseq data in *S cerevisiae*. A) Correlation plot of TR data obtained by means of tiling array hybridization (BioGRO, X axys) and RNA-seq (BioGROseq, Y axis). Only high TR genes (corresponding to 809 genes of the *S. cerevisiae* genome) were represented for the comparison as they were the only ones for which reliable tiling array data was available. For representation, the Z-score of the $\log_2$ intensity values were plotted. Note that the lowest expressed genes show a lower limit in tiling array signal, which is not seen in BioGROseq signal. **B)** Metagene analysis of TR data obtained by means of tiling array hybridization (BioGRO, X blue line) and RNA-seq (BioGROseq, green line) aligned by the TSS. Figures provided by Antonio Jordán at the Department of Molecular Biosciences of the Wenner-Gren Institute, University of Stockholm.

A complementary way in which the quality of BioGROseq data can be checked is by plotting a metagene profile along the first 500 bp of the high TR genes, from the TSS into the gene body. A metagene analysis is an average of quantitative data over one or more genomic regions (often genes or transcripts) aligned at some internal feature. The utility of this type of representation for the purpose of this study relies on the fact that, when it was first set up, one of the most interesting features that the BioGRO technique allowed to explore was the accumulation of active RNApol alternating with the position of nucleosomes[159]. Since the BioGRO signal originates exclusively from transcriptionally active RNA pol, this signal is an indicator of the regions at which active RNA pol is present, and inversely, the lack of it indicates areas in which RNA pols are either absent of inactive. It is known that when RNA pol II transcribes through a nucleosomal template, it pauses at certain sites, showing peaks of stalling just before the nucleosome dyad of nucleosomes +2, +3 and +4, and just after the dyad of nucleosome +1[132,342,343]. These together create a typical antinucleosomal profile that shows a regularly spaced peak and valley shape when representing the BioGRO average signal for all genes from the 5' end towards the gene body (Figure 42). This type of representation, therefore, constitutes an alternative way to prove that only nascent (and not mature contaminant RNA) is being represented in the graph. For this reason, this type of representation was also done for the comparison of tiling and RNAseq BioGRO data (Figure 41B). The profiles observed further support the quality of the TR data obtained by means of RNA-seq. Once proved that the sequencing technique reflects nascent RNA features, the next step was trying this approach with *C. albicans*. To do so, a BioGROseq experiment was performed in a wild-type strain of this species following the experimental pipeline described in Figure 40. Once the sequencing data was obtained though, the limitation was that no BioGRO tiling array data was available for these species. Therefore, a correlation plot like the one shown in Figure 41A could not be represented to check how similar the two approaches are. For this reason, a metagene representation was done, plotting the *C. albicans* BioGROseq data alongside the data obtained in *S. cerevisiae* shown in Figure 41B.

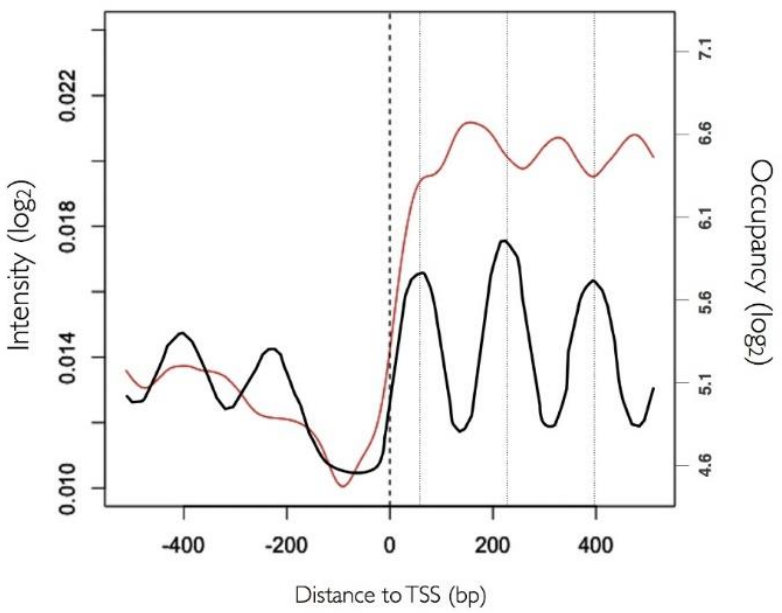


Figure 42. **Typical antinucleosomal profile of BioGRO data.** BioGRO metagene profile in *S.cerevisiae* BQS252 (red line, on the $\log_2$ scale of arbitrary intensity units) compared to the average nucleosome positioning (black line, on the $\log_2$ scale of occupancy levels). Taken from Jordán Pla et al. 2015[159].

Currently there is no annotation of TSSs positions available in *C. albicans*. For this reason and so as to represent both yeast species in the same graph the data were aligned by the position of nucleosome +1 instead. Profiles of expression in both species alongside the first 500 bp from the nucleosome +1 dyad position are shown in Figure 43. It can be seen how the typical antinucleosomal pattern observed in *S. cerevisiae* is also present in *C. albicans*, proving that the technique can now also be used for this yeast species. This now makes this technique useful for the study of the nascentome of this species, which can from now on be compared to all the features discovered in *S. cerevisiae* so far.

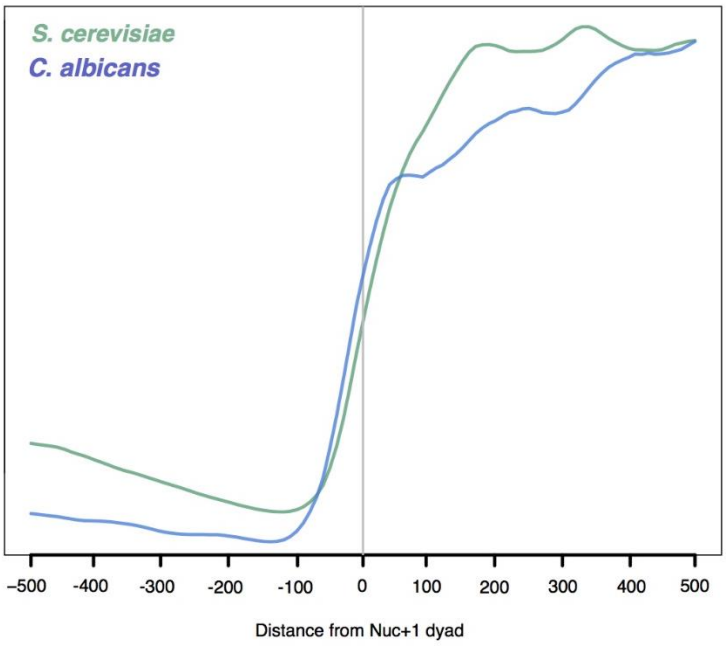


Figure 43. **BioGROseq data metagene plot in *S. cerevisiae* and *C. albicans***. Average nascent RNA expression (in RPKM) is shown alongside the first 500 bp of the gene from the position of the nucleosome +1 dyad in *S. cerevisiae* (green line) and *C. albicans* (blue line). Only high TR genes were plotted, corresponding to 2154 genes in *S. cerevisiae* and 1743 in *C. albicans* (as obtained from the expression analysis of the data set included in Cottier *et al.* 2015[344].)

3.6. Discussion and future perspectives

The main aim of the work described in this chapter was to implement the BioGRO protocol for its coupling with sequencing techniques as well as for it to be used with a yeast species it had never used before, *C. albicans*. The reasons and advantages of setting up this technique in this new species are many. Firstly, it allows the exploration of the nascentome of *C. albicans*, something that has not been done to date. Secondly, using as a reference everything that has been studied so far in *S. cerevisiae*, any already discovered feature about this yeast can now be compared between species to establish similarities and differences in regulatory mechanisms. And thirdly, being this a sequencing approach, the possibilities of discovering new regulatory elements becomes more remarkable in both species regardless of how extensively their genome has been mapped. To do this, several steps had to be taken, including testing whether run-on experiments can be performed in

C. albicans, making sure GROs could be done in the context of this study, and implementing the BioGRO protocol to make it compatible with sequencing platforms. All aspects were implemented at the technical level with a successful outcome. However, all these findings still remain at the technical level of implementing a technique, and from now on the technique will have to be used to address biological issues to prove its reliability answering biological questions.

Additionally, several technical issues still need to be addressed in all different sections of this chapter. Firstly, it is still unknown why the incorporation efficiency is lower in *C. albicans* compared to *S. cerevisiae* (Figures 33 and 34). Secondly, it is still to be addressed whether the decrease in RNA pol density at different growth temperatures (Figure 33) responds to a biological or a technical reason. This is addressed in Chapter 4. And thirdly, it is still to be explored why the effect of Sarkosyl differs from one species to the other (Figure 34). But more importantly, it is now essential to make sure the final BioGROseq version is really optimal for the study of the features that are aimed to be studied and that the technique was implemented for. It should be kept in mind that the main reason for the implementation of this technique was to be able to obtain TR data for medium and low TR genes, something that was missing with the tiling array approach. The fact that the average intensity profile depicts a clear wavy pattern that proves the nature of the technique (Figure 43) does not necessarily mean that the medium and low TR genes are accounting for it. In fact, preliminary analyses of the *C. albicans* BioGRO seq data seem to show that the wavy pattern is mostly created due to the contribution of a few high TR genes that condition the profile (data not shown). For this reason, further technical improvement will now need to be made to keep pursuing the aim of the technique herein presented. Yet, the potential of the BioGROseq technique is now invaluable compared to the limitations presented by previous versions of this protocol. The implementation of this new version also now opens a whole world of possibilities regarding the study of *C. albicans*. Once improved, this technique will be used for a broader study of the response to oxidative stress and hypoxia in *C. albicans*, as well as to explore gene regulatory mechanisms in this species, such as the cross-talk between transcription and mRNA degradation.

Chapter 4

Study of changes in transcription and translation rate as a function of the growth temperature in yeast

4. 1 Motivation

GRO assays serve as a reflection of the average density of active RNA polymerases sitting on the genes at a given time. When addressing GRO labeling efficiency in *C. albicans* for the purposes of this study, it was found that in both *C. albicans* and *S. cerevisiae* the average density of polymerases decreased when cell cultures were grown at higher temperature (Figure 33). Regardless of the yeast species used, the run on labeling efficiency in cells grown at 37 ^{o}C was only 60% of that observed in cells grown at 30 ^{o}C, suggesting that the estimation of the TR is affected by the temperature the cells were grown at. This observation prompted the assessment of whether other parameters that are measured when studying gene expression would also be affected by a change in cell growth temperature. These include gene expression estimators such as the SR or the DR, the cellular concentration of RNA and proteins, or others that would influence nTR estimations through the use of GRO based techniques (i.e. RNA pol II speed and density).

Even though the final aim of this study was characterizing changes in gene expression in *C. albicans*, the relationship between GRO parameters and cell growth temperature was decided to be addressed in *S. cerevisiae* as previous data were already available for this species in our laboratory, and all the needed techniques for the study were already implemented for this organism when this work started. This avoided the GRO sensitivity limitations already confirmed in *C. albicans* (Chapter 3) and allowed the comparison of the results with previous data sets[2,345]. For this reason, most of the experiments described in the present chapter were performed using *S. cerevisiae*.

4.2. RNA pol II density is negatively correlated with cell growth temperature

As previously discussed, the nTR of a given gene is usually determined via the estimation of the density of RNA pol II molecules sitting on it at a given time. This density, along, depends on three different factors. First, on the number of RNA pol initiation events, second, on the number of abortive elongations that cause the polymerase to drop-off from the template, and third, on the elongation speed (Figure 44). Being the two first parameters constant, the

RNA pol density and elongation speed are mutually inverse parameters: the higher the elongation speed the lower the density of RNA pol molecules and *vice versa*[346]. This correlation applies to both average measurements from whole gene lengths and single gene estimations all along the gene length on a given gene of study. Indeed, this relationship between RNA pol density and speed has been recently utilized for the interpretation of RNA pol II presence profiles, showing characteristic accumulations of slowed (or paused) RNA pol II at the 5' and 3' ends of all genes, both in yeast[159] and in higher eukaryotes (reviewed in Jonkers & Lis, 2015[346]). Given this relationship between density and speed, if changes in any factor -either internal or external to the cell- affect the speed of the RNA pol II, one would expect it to see a concomitant change in RNA pol II density, so long as transcription initiation events and the drop-off index remain constant.

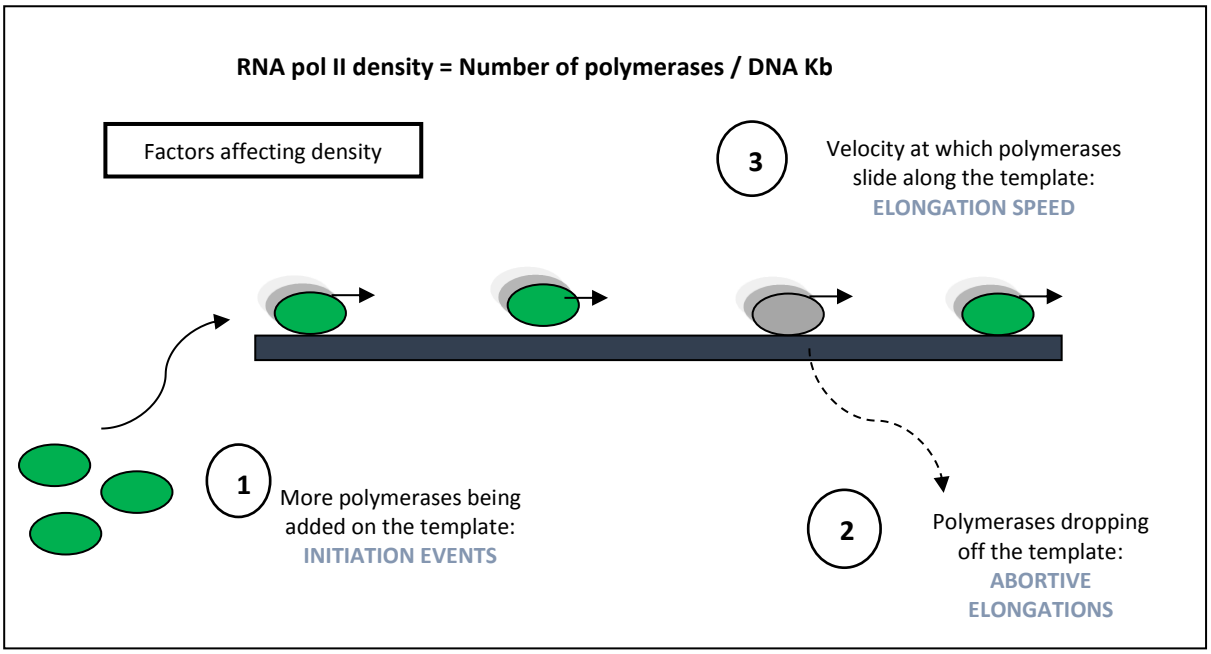


Figure 44. Factors affecting RNA pol II density on a given DNA template.

Parallel to this finding, using single gene techniques it was found in our laboratory that along the 8 kb-long YLR454W gene the RNA pol II speed increases with growth temperature[347] in *S. cerevisiae*, in a way that almost perfectly fits the classical view of how an increase in the reaction temperature provokes an increase in enzyme activity[348] (Arrhenius law). In

particular, elongation speed increased linearly from 0.67 to 1.31 kb/min as the growth temperature increased between 23 and 37°C[86]. In addition, in this work conditions, not only RNA pol II speed increased, but also the drop off index remained constant[347]. Given this finding and the above mentioned relationship between density and speed it was decided to investigate the dependence of the density of RNA pol with the cell growth temperature. Once known that the drop off index does not vary at different growth temperatures, this raised two possibilities: either RNA pol II density changes with growth temperature in a way that compensates the change in elongation speed; or it remains constant due to a parallel increase in the number of initiation events. In the former case the nTR will remain constant. In the latter nTR will increase proportionally in accordance to the Arrhenius factor.

To explore these two alternatives, a GRO experiment was performed to determine average RNA pol II densities in cells that had been grown at increasing temperatures: 23, 30, 34 and 37°C. This experiment was done in the *S. cerevisiae* strain GYLR-4B strain so as to utilize the same strain that was used in the assays performed to address RNA pol II speed in the *YLR454W* gene[86]. Given that *S. cerevisiae* grows at different rates (growth rate, GR) depending on the temperature of the culture, prior to the performance of the experiment the GRs of the GYLR-4B strain at the temperatures of study were checked. Figure 45 shows that between 28 and 37°C (considered the optimal range of growth temperatures for this species) the GR barely varied, showing a significant decrease in growth at temperatures below 28 °C. The GRs observed were similar to those obtained in a previous study for the *S. cerevisiae* BY4741 strain[349].

For each growth temperature, the total density of elongating RNA pol II molecules were measured on all genes of the *S. cerevisiae* genome for which nucleotide probes were contained in the macroarrays used in this study (769 ORFs). Since the run-on signal accounts for all the elongating RNA pol molecules, and these macroarrays contain whole-length probes for all these genes, the median of all gene signals serves as an estimator of the average total RNA pol II density in canonical genes. Figure 46 illustrates how the median run-on signal drastically dropped from 23°C to 37°C along the whole set of canonical RNA

pol II genes. This result shows that RNA pol II density decreases linearly with growth temperature.

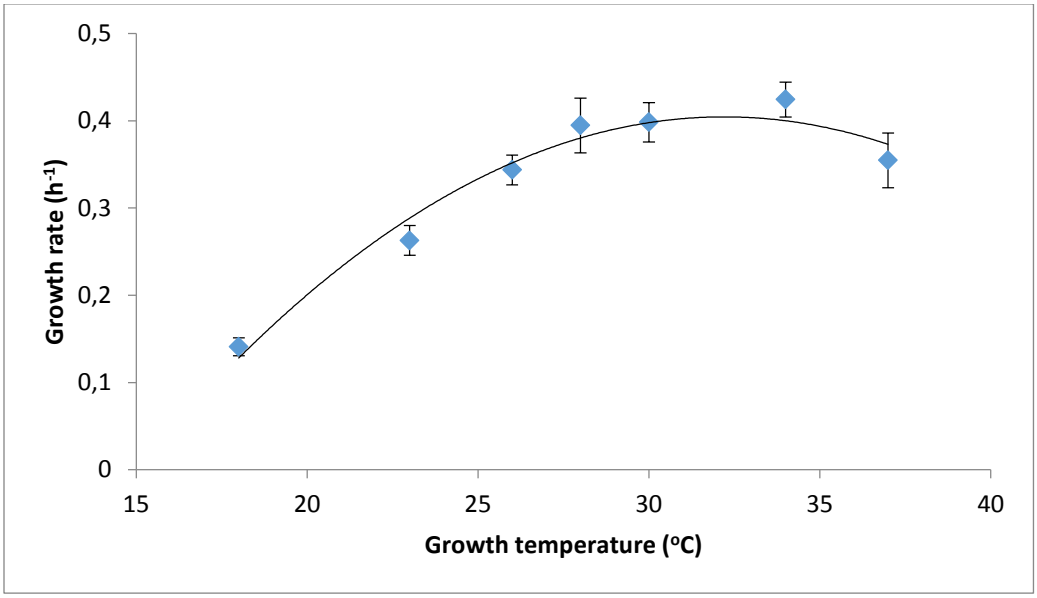


Figure 45. Growth rates of *S. cerevisiae* GYLR-4B at different incubation temperatures. Cells were grown in exponential phase in YPD medium at 18, 23, 26, 28, 30, 34 and 37 ºC for a minimum of 7 generations. The GR was calculated as the inverse of the slope in a representation of the Ln of the $O.D._{600}$ value *versus* the time in hours. For each temperature the experiment was performed in biological triplicates.

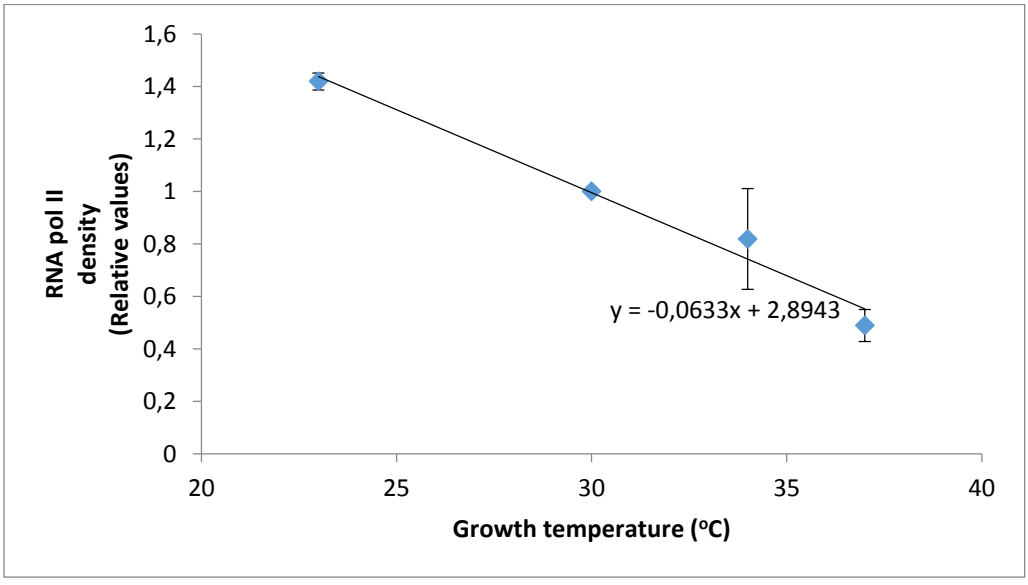


Figure 46. RNA pol II density drops as the growth temperature increases. *S. cerevisiae* GYLR-4B cells were grown in exponential phase for at least 7 generations in YPD medium at 23, 30, 34 and 37 ºC respectively. The total RNA pol II density was calculated as the sum of the run-on incorporation on the 769 *S. cerevisiae* gene

probes as described in materials and methods. Incorporation values are shown as relative to that at 30 ºC, which was arbitrarily taken as 1. The experiment was performed in biological triplicates.

To further confirm this observation, a MiniGRO experiment was performed using *C. albicans* cells in the same growth conditions, using *S. cerevisiae* as the experimental control. Figure 47 shows how the results in *C. albicans* followed the same pattern observed in *S. cerevisiae*, further confirming the influence of growth temperature in RNA pol II density. Therefore, it can be hypothesized that a change in RNA pol II speed will be observed in a similar fashion if the correspondent experiment (RNA pol II speed determination) was reproduced in this species.

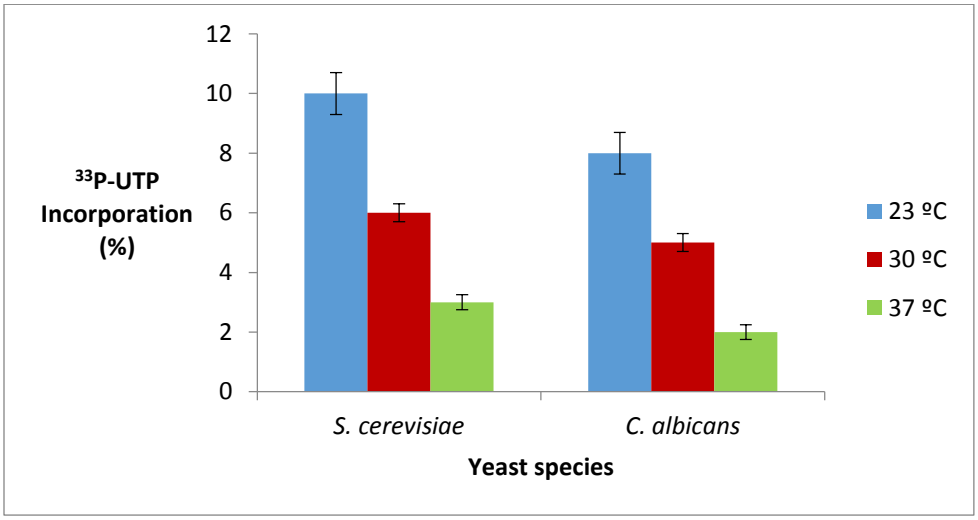


Figure 47. Average RNA pol density in respect to the growth temperature in the two yeast species used in this study. Both *S. cerevisiae* GYLR-4B and *C. albicans* cells were grown in exponential phase for at least 7 generations in YPD medium at 23, 30, and 37 ºC respectively. The percentages of ^{33}P- UTP incorporation as determined by a MiniGRO assay are shown. The experiment was performed in 3 biological triplicates.

4.3. Changes in RNA pol II elongation speed are compensated by changes in density

In the previous section it was verified that an external factor causing a change in RNA pol II speed was reflected in a change in RNA pol II density, indicating the existence of an inverse relationship between these two parameters. To further verify this relationship, it was decided to investigate whether another cause of RNA pol II speed reduction, this time one internal to the cell, would reproduce the same effect.

It has been reported that the depletion of NTP pools within the cell causes a reduction in RNA pol elongation speed[345]. This depletion can be provoked by treating the cells in culture with the nucleotide-depleting drug 6-Azauracil (6AU). In our hands, the treatment with 50 µg/mL of this drug provoked a 60% reduction in RNA pol II speed in a *S. cerevisiae* wild-type strain after 30 min of treatment[86], results that are similar to those previously published by other authors[345].

Given the previously observed relationship between density and elongation speed, it was hypothesized that the NTP pool depletion causing a reduction in speed would provoke an increase in density. To address this, global RNA pol II densities were determined in a GRO experiment using *S. cerevisiae* cells before and after 30 min of treatment with 50 µg/mL 6AU, time at which the depletion of the cellular NTP pools had provoked a decrease in elongation speed[345]. The assay conditions were the same as those previously used in our laboratory to address the change in RNA pol speed after the treatment with the drug, in order to mimic the NTP depletion[86]. The experiment was done in an *imd2Δ* mutant strain to avoid the partial recovery of NTP pools that occurs naturally in a wild-type strain due to the induction of the *IMD2* gene upon nucleotide depletion[350,351]. In this particular strain the drug concentration of use still allowed the cells to grow (at a slightly lower rate than in its absence), and provoked constant [NTP] depletion (data not shown). Figure 48 shows that, compared to the control, the global average RNA pol II density increased by 40% after 30 min of the treatment with the drug.

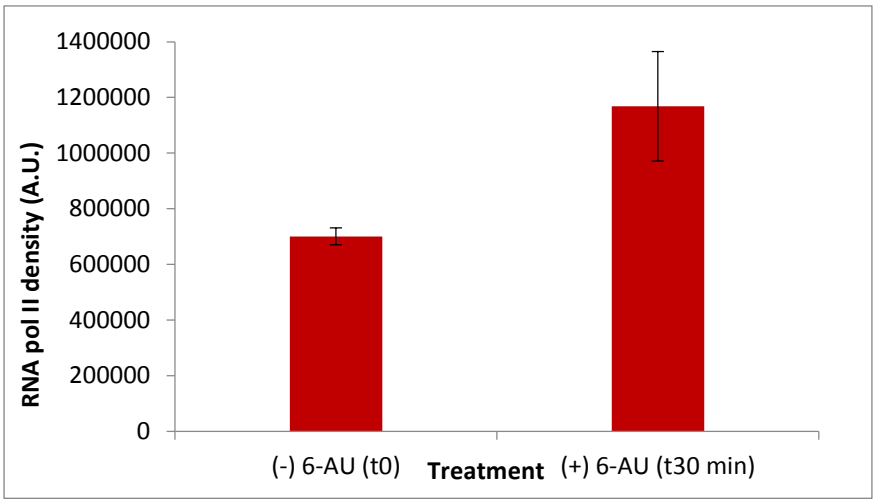


Figure 48. The depletion of NTP pools in the cell causing a reduction in RNA pol II speed is reflected in a change in RNA pol II density. Bars show total RNA pol II density as determined in a GRO experiment done in *imd2Δ* cells before (left) and after (right) the addition of 50 µg/µL of the nucleotide-depleting drug 6-AU.

This result confirms both the idea that the elongation speed of RNA pol II is inversely related to its density and that changes in elongation speed are the cause of the reduced run-on labeling at higher temperatures previously seen in both *S. cerevisiae* and *C. albicans* (Figures 33 and 47).

4.4. In response to changes in growth temperature nTR is controlled at the RNA pol II initiation level

As previously discussed, the transcriptional activity on a gene is directly proportional to the density of RNA pol II molecules present on it. This density along depends on the number of transcription initiation events at the gene promoter, changes of RNA pol II speed during transcription, and the number of abortive elongations observed (drop off index).

When taken together, the changes in RNA pol II speed and density in respect to the growth temperature described in previous sections seem to compensate each other, maintaining certain homeostasis in transcription rate levels (Figure 49). Therefore, it can be concluded that the increase in elongation speed is compensated by a decrease in the RNA pol II density, likely in order to maintain nTR levels. This homeostasis, however, only applied within the optimal range of growth temperatures (26- 30 ^{o}C, Figure 45), as a decrease of approximately

the 25% in nTR was observed at 37 ºC. Since the change in elongation rate proved to be only passively responding to the changes in temperature following the Arrhenius law[86], and the drop- off index does not vary as the growth temperature changes[86], it can be concluded that this decrease in nTR observed at 37 ºC is due to a reduction in the number of initiation events (Figure E1). This means that the main control point of the change in nTR in response to changes in growth temperature is RNA pol II initiation, which is usually the main regulation step in the transcriptional process (discussed in Pérez-Ortín *et al.* 2013[352]).

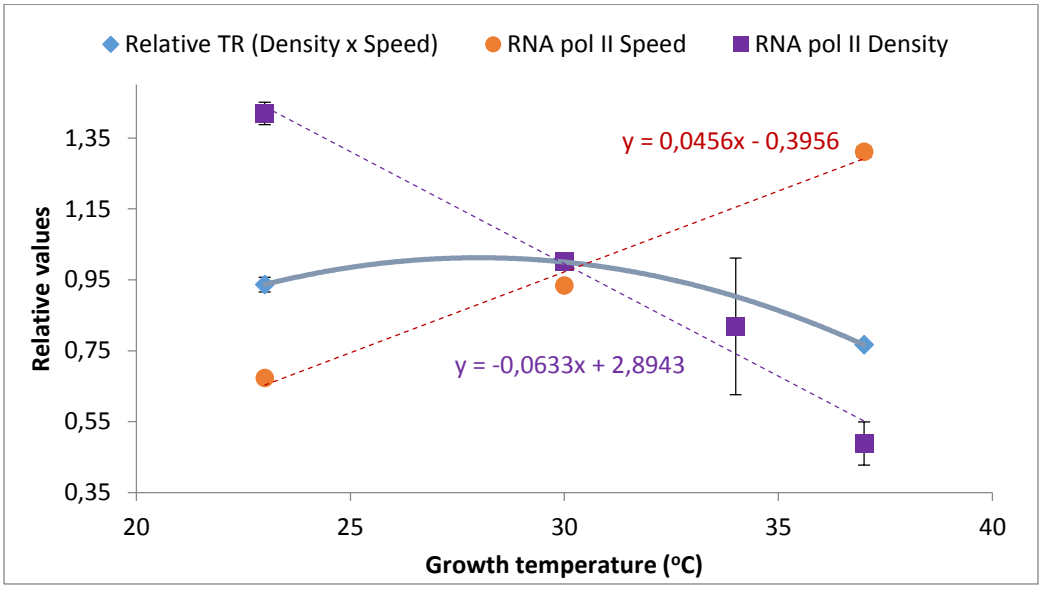


Figure 49. Nascent transcription rate levels are adjusted as growth temperature changes to maintain homeostatic nTR levels. RNA pol II speed increases with growth temperature (red dotted line), as RNA pol II density decreases to compensate for the increase in speed (purple dotted line). When taken the contribution of these two parameters together for nTR estimation (density x speed) nTR levels remain unchanged within the optimal range of growth temperatures (23-30 ºC), decreasing by a 25% at 37 ºC (blue tendency line). RNA pol II density values were taken from those previously shown in **Figure 45.** RNA pol II speed values were taken from Miguel *et al.* 2013[86]

All together these results suggest that, when measuring TR in terms of RNA pol II density using GRO in different samples exposed to different conditions, it should be made sure that RNA pol II pol speed is not differentially affected in the different samples tested. A clear example of this would be when performing GRO in cells subjected to a heat shock treatment, where cells would have to be transferred from normal growth temperature (i.e.

28 °C) to heat shock conditions (i.e. 37 °C). In this particular case, for the estimation of nTR values of RNA pol II densities obtained at 37 °C the obtained data should be corrected by the change of speed that is caused by the increase in growth temperature[166].

4.5 Total RNA and mRNA amounts are maintained at homeostatic levels within the optimal range of growth temperatures

During exponential growth in yeast, mRNA levels are known to be approximately at a steady state[165]. Steady state is a dynamic situation in which mRNA synthesis (SR) and degradation rates (DR) are equal, maintaining a stable level of RNA molecules within the cell:

Steady state: $SR = DR = k_d \cdot [RA]$

where k_d is the first order degradation constant for RNA molecules and RA stands for RNA amount (concentration). This equation, in fact, applies to every kind of RNA species (i.e. rRNA, tRNA, or mRNA) and every RNA pol (I, II or III) that synthesizes them. Because the SR for RNA pol II can be inferred from its nTR value (discussed in Pérez-Ortín et al, 2013[353]) the relatively constant nTR levels observed in Figure 49 suggest that for mRNA molecules DR is also kept constant within the optimal range of growth temperatures. Within the cell, this situation could be achieved via compensatory changes in either RA or k_d.

To test this hypothesis, the overall RA present in cells growing at different growth temperatures was determined. This was achieved by performing total RNA extractions from identical cell mass aliquots grown to exponential phase at 18, 23, 26, 30, 34 and 37 °C. The total RNA obtained from this type of extractions is mainly composed of rRNA and tRNA molecules (which account for the 80% and 15% of the total RNA population within a given cell, respectively[354]). For this reason, these extractions serve also as good estimators of ribosome concentration in the cytoplasm. In parallel, the amount of (poly-A) RNA was measured, as it serves as an estimator of mRNA amount within the cell. The relative amount of poly-A RNA compared to the total RNA amount was determined using a dot blot technique, in which total RNA samples were spotted on a nylon membrane and hybridized to labeled oligo d(T) probes. The relative mRNA values obtained by dot blot were referred to the total RNA amount per cell to get a final value of total mRNA per cell. Assuming that

the cell volume does not vary with growth temperature this total mRNA/cell value is equivalent to mRNA concentration ([mRNA]). Figure 50 shows that the values of total RNA/cell and mRNA/cell remain constant over the usual range of growth temperatures used for laboratory cultures of *S. cerevisiae*, but both RNA values decrease at lower and higher temperatures.

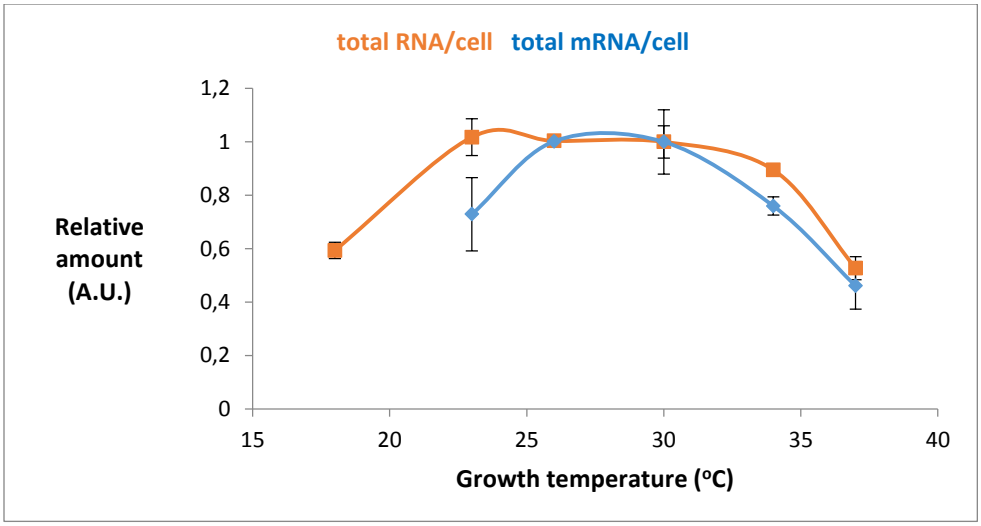


Figure 50. Total RNA and mRNA amounts per cell at different growth temperatures. *S. cerevisiae* GYLR-4B cells were grown in YPD medium to exponential growth phase at the indicated temperatures. The red line shows relative values of total RNA amount per cell as determined by total RNA extraction corrected by the number of cells (determined as $O.D._{600}$). The blue line shows total mRNA amounts per cell as determined by dot blot analysis using labeled oligo d(T) probes against the total RNA samples obtained for total RNA quantification. The values of both data series are shown as relative to the values obtained at 30 ºC that were taken as 1 for reference. Both experiments were performed in technical and biological triplicates.

While the decreased levels of total RNA and mRNA at lower temperatures could be interpreted as caused by the reduced growth rate (Figure 45), what still remained to be explained was the observed decrease in nTR and mRNA amounts at 37°C, where the GR remained similar to that of the reference temperature. This phenomenon could be due to the increased degradation rate for mRNAs at this growth temperature, that had been already reported in a previous study[355]. To confirm this hypothesis, the relative stability of several mRNAs of independent genes was assessed in the conditions of study, using a transcriptional shut-off approach with the drug Thiolutin. Transcriptional shut off

techniques allow the determination of mRNA half-life, which indirectly reflect the rate at which they are being degraded (k_d). For the purpose of this study, the relative stability of 6 independent genes at two different growth temperatures (28 and 37 °C) was assessed. The genes included in the experiment and their relative stabilities at the two growth temperatures are shown in Table 10.

Growth temperature	***RPL17***	***ACT1***	***PHO88***	***RPL25***	***RPL5***	***RPB6***	**Average ratio**
28 °C	44	128	34	51	77	40	
37 °C	25	58	21	24	29	20	
Ratio (37 °C /28 °C)	0.56	0.45	0.61	0.48	0.38	0.50	**0.50**

Table 10. Average half-lives of 6 independent genes at two different growth temperatures. The half-life (in min) of the genes indicated at the top was determined by transcriptional shut-off in cells grown at two different temperatures, 28 and 37°C. The displayed values were calculated as described in materials and methods. The ratio of change between growth temperatures for each independent gene is shown at the bottom. The average ratio of change from 28 to 37°C is shown on the right, where the ratios obtained for all six genes were included for the calculation.

As it can be observed in the results, the stability of each independent mRNA species when cells are grown at 37°C decreases compared to the one at 28°C. On average, and considering the genes included in this study it can be concluded that in the conditions used the stability of mRNAs is reduced by 50% at 37°C compared to 28°C (Table 10), or in other words, the degradation rate constant (kd) increases by a 50%.

In summary, taken all together, these data show that at 37 °C the average SR decreases by 25% (as inferred by the nTR values in Figure 49) while the degradation rate increases by 50% (Table 10). Thus, there is no exact compensation of the synthesis and degradation rates at this growth temperature, which fits with the decrease of mRNA levels. (Figure 49). Regardless, the variation in RA is still maintained within a range of ±50%, which is a deviation range that is considered to be within homeostatic levels, as previously observed in our laboratory[356] and as previously reported by other authors[357]. Hence, overall, it can be concluded that total RNA and mRNA amounts are maintained at homeostatic levels within the optimal range of growth temperatures.

The decrease in total RNA concentration at high temperature, on the other hand, suggests that the yeast cell requires less ribosomes and tRNAs at those temperatures, likely because the Arrhenius effect on translation increases the elongation speed of each ribosome. To study the protein biosynthesis capacity in yeast cells an experiment was conducted, which is described in the following section.

4.6. Translation rate increases with growth temperature

Despite its importance in gene expression, RNA is just a transient molecule in the process. What really enables the cell to keep growing and sustain living is the maintenance of protein production at an adequate rate, since the concentration of proteins is the main parameter that needs to be adjusted when changes in cell physiology are required[358]. In the previous sections it was observed that the RA is reduced by 50% in cells grown at 37 ^{o}C compared to those grown at 30 ^{o}C (Figure 50). Despite this decrease, cells seemed to grow just as fast at this growth temperature, maintaining a very similar growth rate (Figure 45). These results suggest that the yeast cell is adjusting its protein levels to be able to sustain the same growth rate at 37 ^{o}C compared to 30 ^{o}C. This being the case, this adjustment could be achieved in two different ways, either by increasing protein stability or by increasing protein synthesis at 37 ^{o}C. However, and similarly to what happened with mRNA molecules, it had been previously reported that protein stability decreases at 37 ^{o}C compared to 30 ^{o}C growth temperature[359]. For this reason, it was decided to test whether an increase in protein production could explain why cell growth rate was maintained. To do so, a translation activity assay was conducted to measure translation rates (TLR) in cells growing at different temperatures, using 4 of the temperatures of study (23, 30, 34 and 37 ^{o}C). In this assay, TLR was determined by giving the cells in culture a pulse of amino acids which contains one that is radioactively labeled, ^{35}S-Methionine (^{35}S-Met). This pulse allowed this amino acid to be incorporated into new proteins to a rate that was proportional to the rate at which they were being made (that is, the TLR). The TLR was then determined as the percentage of TCA-precipitable radioactivity in relation to the total incorporated radioactivity in the cells[288]. In

order to perform this assay, the cells needed to be grown in a medium lacking Methionine, so as to avoid the labeled amino acid to be diluted with the cold one already present in the medium and maximize incorporation efficiency. It was for this reason that this assay could not be performed in the yeast strain used in the previous experiments described in this chapter (GYLR-4B, which is *met15Δ*), so the most similar strain present in our laboratory was used, BQS252 which has identical genetic background but is Met^+. Figure 51 shows how the TLR increases with growth temperature, opposite to the decrease observed for TR.

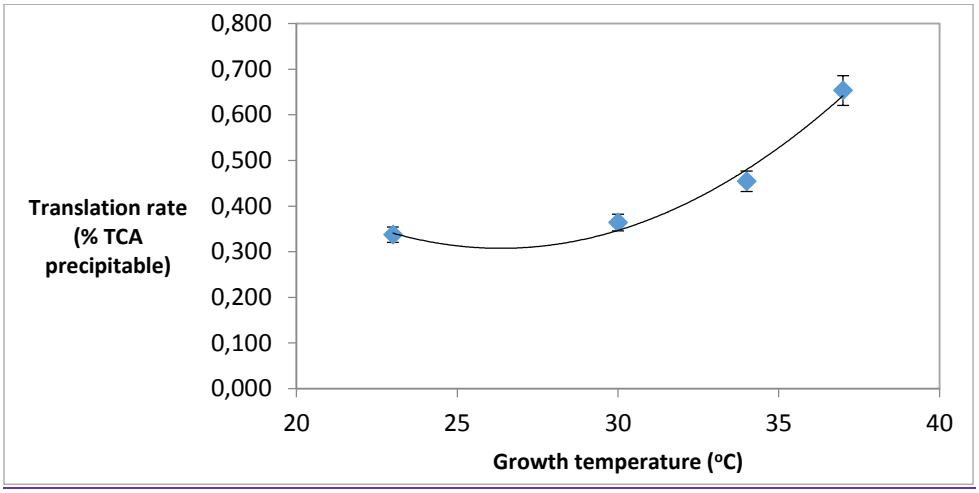


Figure 51. Variation of TLR with growth temperature as determined by ^{35}S-Met incorporation. The TLR was determined as the percentage of TCA precipitable radioactivity, which was calculated as relative to the overall ^{35}S-Met intake. This experiment was performed in *S. cerevisiae* BQS252 in biological triplicates, each of which was done in technical triplicates.

This observed increase in TLR could be explained as a passive consequence of the enhanced enzymatic activity at higher temperatures (as predicted in the Arrhenius law, this time referring to ribosomes). Alternatively, an increase in translation could be a consequence of an increase in the amount of the translation machinery (ribosomes) or mRNA content present in the cells. As previously described, the total amount of RNA per cell serves as reflection of the ribosome content, as the vast majority of RNA molecules account for rRNA. As shown in Figure 50, the amount of total RNA per cell not only did not increase with growth temperature but actually dropped at 37 ºC, and the same was observed in mRNA

molecules. In agreement with these results it has been reported that in the bacterium *E. coli* the ribosome content does not vary within the range of physiological growth temperatures for this species (37-42 ^{o}C[360]). Hence, the observed increase in TLR is more likely to be a passive consequence of the increase in enzymatic activity with temperature, even compensating the decrease in active ribosomes. This idea is further supported by other studies that showed that ribosome speed is strictly dependent on temperature in both prokaryotes[360] and eukaryotes[358]. In addition, and despite the decrease in mRNA levels at 37 ^{o}C, it has also been proven that in a yeast cell growing at 30^{o}C ribosomes work at saturating levels of mRNA[361], hence this decrease is not even likely to affect translation. An additional and compatible explanation for the increase in TLR while the growth rate stays stable could be that the increase in TLR compensates the decrease in protein stability that is observed at 37 ^{o}C compared to 30 ^{o}C[359], as protein concentration in yeast cells is almost constant in between 28 and 37 ^{o}C[86]. Interestingly, the parallel change in rRNA and mRNA seen from 26 to 37 ^{o}C (Figure 50) suggests that the saturation level of ribosomes with mRNAs is kept constant over a wide range of growth temperatures.

4.7. Discussion and future perspectives

The present chapter was dedicated to the study of the interdependence between growth temperature and gene expression parameters in the model yeast species *S. cerevisiae*. In particular, it was studied how TR and TLR are affected by changes in external temperature, which casted into light the similarities and differences between the processes regulating mRNA and protein production, respectively.

Based on the observed results, both processes seemed to be affected by classic physiochemical laws, such as the Arrhenius law for the dependence of enzymatic reactions with temperature. Because this phenomenon unavoidably occurs, the cell needs to make use of regulatory mechanisms to ensure homeostasis for these two types of molecules. The case of temperature as an external variable is difficult to study, as it provokes changes in all enzymatic reactions occurring within the cell, not only those committed to biomolecule synthesis but also to those involved in their degradation. Furthermore, changes in

temperature can affect the structural stability of macromolecules, being these less stable at higher temperatures. For this reason, the different parts of the process (synthesis, degradation, stability) have to be often addressed separately.

As found in this study, as well as in others studying other different external variables[356,357], homeostasis of mRNA molecules is maintained within a range of ±50%. Protein concentration, however, seems to be more finely adjusted, displaying a narrower range of variation[86]. This makes sense in the context of cell physiology, as protein production is energetically more costly to the cell[354], and is the determinant step in the context of gene expression. The global conclusion that could be drawn in this study is that, within the usual range of laboratory growth temperatures for *S. cerevisiae* (26-34 ^{o}C), both TR and TLR remain mostly invariable (Figures 49 and 51). This requires, most likely, control mechanisms at the initiation level in both processes. Interestingly, however, this did not apply beyond this range of growth temperatures. Cells growing at 37 ^{o}C displayed a different strategy, in which nTR decreased and TLR increased when compared to lower growth temperatures. The decrease in nTR at 37 ^{o}C is paralleled by a decrease in RNA pol II concentration in yeast cells[86] which suggests that the main pathway for total nTR control in *S. cerevisiae* is the regulation of RNA pol II levels. The reason for the difference in transcription and translation rates control remains unknown, it is likely to be related to the different roles that mRNA and proteins perform in the cell. Messenger RNA is an intermediary in the information flux, and it is a quite unstable molecule: mRNA half-lives in yeast usually range between 2 and 100 min[362,363]. Protein molecules, on the other hand, are much more stable and can have average half-lives up to several hours[302]. Moreover, as previously stated, protein synthesis is by far the more costly process for the cell, mainly due to the much higher number of proteins present per cell compared to the number of mRNA molecules[302]. Such high energy cost implies that growth rates are unavoidably strongly associated to TLRs[358]. In this case, however, even though the growth rate at 37 ^{o}C was very similar to that at 30 ^{o}C gene expression strategies changed. The reason for that will now have to be explored. RNA pol II transcription rates have also been recently discovered to be correlated to the growth

rate[364], but the transcription process overall is comparatively much less costly to the cell. These physiological dissimilarities are probably behind the different strategies regarding changes in growth temperature, as they will also probably be regarding other variables.

The fact that 37 ^{o}C is the highest growth temperature that this species tolerates without implying a decrease in viability[365] suggests that the changes observed may be depicting the mechanisms the cell starts using when adapting to adverse external conditions.

Finally, as explained at the beginning of this chapter, *C. albicans* has a different physiology than *S. cerevisiae* in many aspects. Therefore, it would now be very interesting to explore the possible differences between these two yeast in order to better understand the respective contributions of transcription and translation to their gene expression strategies. This will be particularly interesting in *C. albicans* as this organism grows in different niches whithin the human body that have different temperatures, such as its growth on skin in mucocutaneous infections or the growth whithin organs during systemic ones.

Conclusions

CHAPTER 1

All the conclusions in Chapter 1 refer exclusively to *C. albicans*:

1. In response to oxidative stress conditions caused by the addition of 0.1 mM t-BOOH, the levels of a total number of 29 proteins were found to change, 19 being up-regulated and 10 downregulated.

2. The differentially expressed genes listed in this study were regulated at the post-transcriptional level, with the exception of *TSA1*, *CIP1* and *SSC1*.

3. The oxidative stress response in *C. albicans* is stressor-specific.

CHAPTER 2

1. The application *TilingScan* allows the location of differentially expressed regions all over the genome of any given organism in both tiling array and RNAseq data without the need of any previous genome annotation.

2. In response to hypoxia, 30 min and 180 min under 0.02% O_2 conditions, *C. albicans* regulates the expression of 16 and 138 lncRNAs respectively, including antisense and intergenic transcripts.

3. In response to oxidative stress, 40 min and 180 min after the treatment with 0.1 mM of t-BOOH, *C. albicans* regulates the expression of 141 and 18 lncRNAs respectively, including antisense and intergenic transcripts.

4. The list of ORFs containing antisense transcripts which expression changed under oxidative stress was enriched in genes coding for proteins located on the cell surface. The same occurred with some of the genes detected in hypoxia.

5. Twelve of the antisense transcripts detected in this study are also involved in the adaptation of *C. albicans* to environmental conditions during biofilm formation and serum-induced transition to hyphal state. Ten of them corresponded to genes coding for proteins located in the cell surface.

CHAPTER 3

1. The radioactive approach of the genomic run-on technique (GRO) works in *C. albicans* as well as it does in *S. cerevisiae*, being the labeling efficiency in *C. albicans* 18% lower than the one in *S. cerevisiae*.

2. In both *S. cerevisiae* and *C. albicans*, the GRO labeling efficiency decreases by 40% when the cells are grown at 37 ^{o}C compared to 30 ^{o}C. This is caused by the reduction in RNA pol II density caused by the increase in elongation speed.

3. GRO assays can be performed using frozen cells as starting material. When doing so, in both *S. cerevisiae* and *C. albicans* the labeling efficiency increases by 60% on averagecompared to when the assay is done in freshly collected cells. Freezing the cells prior to performing the experiment does not significantly change the transcriptional profile observed in freshly collected cells.

4. The run-on reaction can work in the absence of Sarkosyl, so long as the cells subjected to the pulse have been previously made permeable by other means, such as freezing.

5. The absence of Sarkosyl in frozen *C. albicans* cells during a GRO experiment yields higher labeling efficiency than when the detergent is added. Regardless, the use of Sarkosyl in a GRO experiment is recommended to prevent new RNA pol initiation events.

6. The sequencing approach of the BioGRO assay (BioGROseq) has been implemented for its use in both *S. cerevisiae* and *C. albicans* providing reliable results for highly transcribed genes.

CHAPTER 4

7. RNA pol II density has a negative and linear correlation with cell growth temperature in *S. cerevisiae*. The same pattern is observed in *C. albicans*.

8. The decrease in RNApol II elongation speed is reflected in an increase of RNA pol II density.

9. In response to changes in growth temperature, changes in nascent transcription rate levels are controlled at the RNApol II initiation level in *S. cerevisiae*.

10. Within the optimal range of growth temperatures in *S. cerevisiae* (26-34 ^{o}C) total RNA and mRNA amounts are kept at homeostatic levels within a range of ± 50% of their steady state value.

11. In *S. cerevisiae*, mRNA stability decreases by 50% on average at 37 ^{o}C compared to 28 ^{o}C growth temperature.

12. Translation rate increases with growth temperature in spite of the reduction in active ribosomes.

Appendix

APPENDIX I

Description of the search algorithm used in *TilingScan*

For the detection of differential expression, the application has been implemented with a search algorithm that we define as the "sliding window search algorithm". Firstly, two main parameters have to be selected. The first one is the window size (V), which will define the minimum number of probes that will be considered a region of interest. For example, in the case of a data set obtained from a microarray that contains 25mer probes, if regions of 75 or more nucleotides want to be detected, a V=3 will be selected. The second one is the fold-change threshold (U), which will define the minimum fold-change that is required for a region to be considered of significant change. Once these two parameters are defined, two adjacent windows of the desired V size (A and B) are defined at the beginning of the data set signal. For each of them, the average intensity value of the probes contained within them is calculated. These two windows will slide along the data set, until the difference between the average value of A and the average value of B surpasses the fold-change threshold U. When this happens, the start point of a region will be defined from the first point of B. To determine the end point of the region of change, a new window (C), of fixed size = V will be created adjacent to the end of B. The comparison between B and C will now be repeated in the same way it was done for A and B, this time extending window B until the difference between B and C is equal to the selected threshold (U). This will determine the length of the detected region, which will be determined by the start and end point of the extended window B. To prevent false positives from appearing, we established two criteria that all detected regions must meet: First, when the start point of a detected region is defined, the average value of window B must be at least a certain percentage value (P) above 0. And second and complementarily to this, if during the extension of the window B the average intensity value within it drops below (1+U*P) at any time, the end point of the region is then defined. All the defined above will serve for the detection of up-regulated regions. In order to detect those regions that are down-regulated, the inverse of the signal is calculated, and the same algorithm with the same permutation criteria is applied.

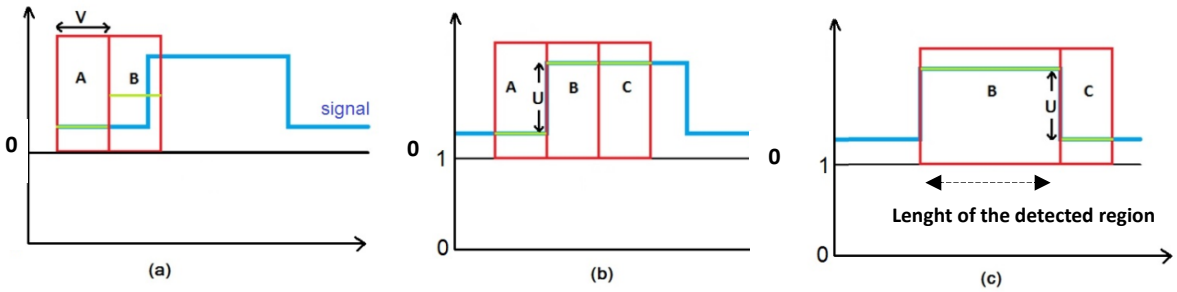


Figure A1. Description of the search algorithm. A) At the beginning of the data set, two windows are defined, "A" and "B", both of the selected size "V". **B)** For each window, the average intensity value of the probes contained within them is calculated. These two windows will slide along the data set, until the difference between the average value of A and the average value of B surpasses the fold-change threshold "U". **C)** When this happens, the start point of a region will be defined from the first point of B. To determine the end point of the region of change, a new window (C), of fixed size = V will be created adjacent to the end of B. The comparison between B and C will be repeated in the same way, extending window B until the difference between B and C is equal to the selected threshold (U). This will determine the length of the detected region.

APPENDIX II

Lists of detected transcripts in each data set that met the selection criteria.

Table A1. List of Up-regulated AS and intergenic transcripts found in the data set Hypoxia t30/t0

Region #	Chr	Length (nt)	Start	End	Watson Y-mean	Crick Y-mean	Overlapping ORFs
1	1	336	1784412	1784748	1.869	0.8771	NO ORF
2	2	864	595500	596364	1.6414	1.0873	orf19.5805(C)
3	3	216	1797168	1797384	1.6043	2.2149	NO ORF
4	4	372	844044	844416	1.8515	0.8052	NO ORF
5	R	264	1983084	1983348	0.9266	2.0723	orf19.7331(W)
6	R	228	2155248	2155476	0.9029	2.0799	NO ORF

Table A2. List of Down-regulated AS and intergenic transcripts found in the data set Hypoxia t30/t0

Region #	Chr	Length (nt)	Start	End	Watson Y-mean	Crick Y-mean	Overlapping ORFs
1	1	204	991080	991284	0.4822	0.953	NO ORF
2	1	912	1915164	1916076	0.4021	0.6034	orf19.4737(W)
3	1	204	2706648	2706852	1.0141	0.4595	orf19.5221(W)
4	3	228	80460	80688	0.6992	0.4	NO ORF
5	3	204	1279488	1279692	1.1433	0.4743	orf19.7350(W)
6	4	288	1363920	1364208	0.3687	0.4868	orf19.4716(W)
7	R	1284	574884	576168	0.453	0.8417	orf19.164(C)
8	R	360	1073652	1074012	0.9448	0.3946	orf19.1808(W)
9	R	264	1259496	1259760	0.996	0.4625	NO ORF
10	R	252	2202156	2202408	0.7915	0.4967	NO ORF

Table A3. List of Up-regulated AS and intergenic transcripts found in the data set Hypoxia t180/t0

Region #	Chr	Length (nt)	Start	End	Watson Y-mean	Crick Y-mean	Overlapping ORFs
1	1	1224	61308	62532	0.8699	1.9213	orf19.6070(W)
2	1	720	88608	89328	0.7982	1.8511	orf19.6053(W)
3	1	624	233304	233928	2.3232	1.0274	orf19.3303(C)
4	1	348	275580	275928	1.85	1.1	orf19.3329(C)
5	1	516	348264	348780	0.9737	2.143	orf19.3359(W)

6	1	348	844404	844752	1.8151	1.319	NO ORF
7	1	312	897828	898140	1.8427	1.8421	orf19.1069(W)
8	1	252	921516	921768	2.0139	1.0617	NO ORF
9	1	336	980400	980736	3.2179	1.0769	orf19.773(C)
10	1	1032	995532	996564	1.813	1.0783	orf19.764(C)
11	1	360	1019220	1019580	0.8025	1.8432	orf19.753(W)
12	1	636	1116312	1116948	4.1776	1.0645	orf19.433(C)
13	1	456	1784364	1784820	2.0685	0.7729	NO ORF
14	1	336	1815612	1815948	2.0831	0.9338	NO ORF
15	1	324	1831368	1831692	1.7439	1.1517	NO ORF
16	1	1176	2044848	2046024	2.4364	1.1353	orf19.4789(C)
17	1	372	2178276	2178648	0.9523	1.809	NO ORF
18	1	396	2281404	2281800	2.2944	0.9545	NO ORF
19	1	984	2289636	2290620	2.0263	2.8911	NO ORF
20	1	1104	2383536	2384640	2.9315	4.2202	orf19.2344(C)
21	1	384	2388852	2389236	1.2629	1.7665	orf19.2342(W)
22	1	864	2454660	2455524	0.8994	2.9479	orf19.2297(W)
23	1	648	2460960	2461608	2.642	2.1988	orf19.2296(W)
24	1	216	2775564	2775780	1.2021	2.0182	orf19.6349(W)
25	1	348	2963004	2963352	1.9104	1.0945	NO ORF
26	1	228	3010092	3010320	0.8413	1.7746	NO ORF
27	1	2568	3116220	3118788	1.8193	2.9308	orf19.7221(C)
28	1	276	3120144	3120420	2.8199	0.8611	orf19.7222(C)
29	2	684	291276	291960	1.3631	3.6391	orf19.1473(W)
30	2	540	576672	577212	1.2333	2.2623	NO ORF
31	2	1920	595248	597168	6.9984	0.6618	orf19.5805(C)
32	2	1296	733356	734652	1.6303	1.0746	NO ORF
33	2	552	838200	838752	2.5001	4.1094	orf19.822(C)
34	2	816	839724	840540	2.9333	1.0274	orf19.820(C)
35	2	216	864624	864840	1.1327	2.067	orf19.810(W)
36	2	348	879708	880056	1.9555	2.0165	orf19.802(C)
37	2	288	1425480	1425768	2.0012	0.9827	NO ORF
38	2	252	1455912	1456164	0.987	2.1565	NO ORF
39	2	276	1656780	1657056	2.0906	1.0165	NO ORF
40	2	276	1693608	1693884	2.4628	0.855	orf19.1434(C)
41	2	732	1890108	1890840	4.2026	1.9126	orf19.4082(W)
42	2	360	2057280	2057640	3.2186	1.0875	NO ORF
43	2	516	2120400	2120916	2.3242	0.8113	NO ORF
44	2	336	2169504	2169840	1.7992	0.8706	orf19.5325(C)
45	2	468	2185896	2186364	0.9299	2.8809	NO ORF
46	2	600	2197476	2198076	0.8586	2.4692	orf19.5346(W)

47	3	420	229692	230112	2.071	4.4168	orf19.2515(C)
48	3	612	333288	333900	1.0569	1.6587	NO ORF
49	3	1164	712896	714060	1.842	1.0198	NO ORF
50	3	1152	733236	734388	1.7328	7.7884	orf19.344(C)
51	3	528	924564	925092	1.9687	1.0045	orf19.5894(C) orf19.5894(C)
52	3	1332	1109748	1111080	0.9424	1.453	NO ORF
53	3	1176	1531584	1532760	1.7825	1.2689	NO ORF
54	4	276	17820	18096	2.5745	2.2327	orf19.5635(W)
55	4	240	20172	20412	3.3721	2.1653	orf19.5636(W)
56	4	228	119412	119640	2.2338	1.0076	orf19.4172(C)
57	4	288	578688	578976	2.1968	1.2564	orf19.2722(C)
58	4	300	775272	775572	2.5232	0.8115	NO ORF
59	4	888	851496	852384	2.5552	0.8047	NO ORF
60	4	228	976092	976320	0.9613	2.2403	NO ORF
61	4	360	1244400	1244760	3.1441	0.8909	NO ORF
62	4	540	1286808	1287348	1.9321	1.0739	NO ORF
63	4	2412	1462068	1464480	1.8433	0.9167	orf19.2879(C)
64	4	588	1557396	1557984	1.7397	1.9986	orf19.3104(W) orf19.3103(W)
65	5	228	138864	139092	2.3412	1.111	orf19.929(C)
66	5	228	345924	346152	1.8685	1.1655	orf19.4151(C)
67	5	228	369792	370020	1.8991	0.9134	NO ORF
68	5	312	419940	420252	1.6373	2.4828	NO ORF
69	5	252	706416	706668	2.5878	0.9059	NO ORF
70	6	864	127464	128328	6.7891	1.8435	orf19.3655(W)
71	6	336	273888	274224	0.9562	1.8657	orf19.3447(W)
72	6	1140	409932	411072	1.4448	2.1805	orf19.690(W)
73	6	648	460116	460764	2.7711	2.1434	orf19.3461(W)
74	6	240	525780	526020	1.8256	0.8423	orf19.5526(C)
75	6	1176	637188	638364	4.2622	1.1295	orf19.5586(C)
76	6	1056	694524	695580	0.8768	1.7977	orf19.5619(W)
77	6	216	703476	703692	2.4002	0.9332	orf19.5623(C)
78	7	396	96336	96732	2.5147	0.9946	orf19.7061(C)
79	7	240	142164	142404	1.463	2.4413	orf19.7044(W)
80	7	444	469752	470196	0.8003	2.2906	orf19.6501(W)
81	R	804	209484	210288	1.0488	2.223	orf19.3263(W)
82	R	240	258228	258468	1.1052	2.0417	orf19.3241(W)
83	R	744	268452	269196	1.2425	2.5543	orf19.3238(W)
84	R	1380	372024	373404	2.1251	1.0675	orf19.2552(C)
85	R	1008	373716	374724	2.8073	1.4239	orf19.2552(C) orf19.2553(C)

86	R	1740	441648	443388	1.8683	0.319	orf19.2602(C)
87	R	912	1983060	1983972	0.9366	2.3751	orf19.7331(W)
88	R	252	2011884	2012136	0.7456	1.7789	NO ORF

Table A4. List of Down-regulated AS and intergenic transcripts found in the data set Hypoxia t180/t0

Region #	Chr	Length (nt)	Start	End	Watson Y-mean	Crick Y-mean	Overlapping ORFs
1	1	504	46272	46776	0.2819	0.7348	NO ORF
2	1	564	536004	536568	0.6739	0.4167	NO ORF
3	1	228	668988	669216	0.4994	1.1196	orf19.3012(C)
4	1	264	991044	991308	0.3642	0.9559	NO ORF
5	1	360	1084344	1084704	0.8081	0.5106	NO ORF
6	1	252	1534848	1535100	1.0571	0.3763	orf19.4450.2(W)
7	1	1452	1657476	1658928	0.7198	0.4569	NO ORF
8	1	216	1885260	1885476	0.468	0.9391	NO ORF
9	1	276	2695488	2695764	0.5116	0.9345	NO ORF
10	1	264	2706612	2706876	1.065	0.3793	orf19.5221(W)
11	2	396	162288	162684	0.6491	0.4665	NO ORF
12	2	636	1030344	1030980	0.5517	0.9657	orf19.3526(C)
13	2	696	1377768	1378464	0.752	0.473	NO ORF
14	2	336	1545084	1545420	0.7913	0.5002	NO ORF
15	2	228	1755228	1755456	1.3614	0.5258	orf19.3615(W)
16	2	204	2105292	2105496	0.8619	0.4296	NO ORF
17	3	240	80472	80712	0.6897	0.294	NO ORF
18	3	456	456984	457440	1.0189	0.4797	orf19.1636(W)
19	3	228	952140	952368	0.7741	0.442	NO ORF
20	3	312	1121820	1122132	0.6311	0.518	NO ORF
21	3	1044	1252680	1253724	0.9668	0.5434	orf19.6995(W)
22	3	336	1279380	1279716	1.2651	0.4246	orf19.7350(W)
23	3	792	1722384	1723176	0.8444	0.5248	NO ORF
24	4	612	16524	17136	0.3989	4.0179	orf19.5634(C)
25	4	1488	253368	254856	1.0047	0.4982	orf19.4662(W)
26	4	720	475308	476028	0.5697	0.8048	NO ORF
27	4	264	867876	868140	0.4352	0.9003	orf19.5305(C)
28	4	300	991920	992220	0.5125	1.1193	NO ORF
29	4	312	995928	996240	0.8485	0.4897	CEN4(W)
30	4	276	1021812	1022088	0.8695	0.4048	NO ORF
31	4	288	1363932	1364220	0.4692	0.4843	orf19.4716(W)
32	4	204	1470720	1470924	0.7111	0.4616	NO ORF
33	5	444	150504	150948	0.3896	0.7765	NO ORF

34	5	276	215208	215484	0.5018	0.7717	orf19.1973(C)
35	5	276	262284	262560	1.0406	0.5593	orf19.1957(W)
36	5	216	638436	638652	0.4062	1.5835	NO ORF
37	5	264	776964	777228	0.4596	0.8775	NO ORF
38	5	204	956148	956352	0.4747	0.8697	NO ORF
39	5	1260	1155564	1156824	1.0096	0.5156	orf19.4029(W)
40	6	204	24408	24612	1.12	0.4478	orf19.1180(W)
41	6	264	336468	336732	0.9036	0.5385	orf19.3425(W)
42	6	264	356160	356424	1.0461	0.4956	orf19.3417(W)
43	6	312	486648	486960	0.2577	0.5012	orf19.3475(W)
44	7	312	465060	465372	0.915	0.5014	orf19.6506(W)
45	R	216	573396	573612	0.794	0.4676	NO ORF
46	R	528	1150752	1151280	0.514	0.8725	NO ORF
47	R	420	1163796	1164216	1.1634	0.4307	orf19.3517(W)
48	R	372	1259448	1259820	0.7843	0.3589	NO ORF
49	R	252	1738164	1738416	0.5044	0.8045	orf19.600(C)
50	R	288	1954092	1954380	0.5118	2.1436	orf19.7313(C)

Table A5. List of Up-regulated AS and Intergenic transcripts found in the data set OxStresst40/t0

Region #	Chr	Length (nt)	Start	End	Watson Y-mean	Crick Y-mean	Overlapping ORFs
1	1	216	52560	52776	2.2054	0.8595	NO ORF
2	1	444	67332	67776	2.9067	1.0903	NO ORF
3	1	348	104928	105276	1.1137	1.8749	orf19.6043(W)
4	1	528	208956	209484	3.8919	1.1923	orf19.3295(C)
5	1	312	244728	245040	1.1716	3.0389	NO ORF
6	1	756	275508	276264	2.7033	1.2089	orf19.3329(C)
7	1	312	297432	297744	1.8539	0.972	NO ORF
8	1	948	388308	389256	2.5261	0.89	orf19.4535(C)
9	1	552	404328	404880	1.153	3.4596	NO ORF
10	1	444	404964	405408	1.8308	1.169	NO ORF
11	1	240	515244	515484	1.899	0.9238	NO ORF
12	1	564	519756	520320	2.2915	1.0465	orf19.2934(C)
13	1	336	528768	529104	0.6895	1.3564	orf19.2940(W)
14	1	240	565476	565716	2.557	1.1837	NO ORF
15	1	360	579216	579576	1.0448	1.9299	NO ORF
16	1	588	614796	615384	2.2324	1.2855	orf19.2985(C)
17	1	240	800268	800508	2.5862	1.1349	NO ORF
18	1	288	842964	843252	2.4278	5.6921	orf19.4476(C)
19	1	468	844332	844800	3.8236	3.0332	NO ORF

20	1	324	895404	895728	0.6685	2.1202	orf19.1067(W)
21	1	600	921852	922452	2.2038	1.4793	orf19.6842(C)
22	1	312	920424	920736	1.8716	1.7873	orf19.6840(C)
23	1	468	1155048	1155516	1.0574	2.1362	orf19.414(W)
24	1	324	1190364	1190688	2.6201	1.3469	orf19.2481(C)
25	1	996	1261764	1262760	1.2283	2.7823	orf19.2445(W)
26	1	336	1335504	1335840	2.1726	1.0136	NO ORF
27	1	300	1363320	1363620	0.9859	2.1476	NO ORF
28	1	492	1395228	1395720	2.1598	0.7856	orf19.6259(C)
29	1	468	1421508	1421976	0.8776	2.1119	orf19.6246(W)
30	1	216	1445772	1445988	0.9488	2.167	orf19.6233(W)
31	1	1296	1488036	1489332	2.7976	0.8828	orf19.6209(C)
32	1	420	1599948	1600368	2.3393	1.0402	NO ORF
33	1	240	1719168	1719408	2.0879	0.6001	orf19.5061(C)
34	1	552	1738884	1739436	2.1558	1.0438	NO ORF
35	1	360	1784424	1784784	2.4273	0.7613	NO ORF
36	1	204	1785012	1785216	2.237	0.9645	NO ORF
37	1	624	1800600	1801224	1.9367	1.0674	NO ORF
38	1	312	1815636	1815948	1.9072	0.7456	NO ORF
39	1	696	1994928	1995624	2.2053	1.0476	NO ORF
40	1	456	2001708	2002164	1.0046	2.2448	NO ORF
41	1	780	2007744	2008524	1.2707	2.4112	orf19.4774(W)
42	1	264	2024796	2025060	0.9108	2.5216	NO ORF
43	1	1260	2044752	2046012	3.5956	1.212	orf19.4789(C)
44	1	312	2093256	2093568	0.5068	2.1146	orf19.4815(W)
45	1	240	2099052	2099292	0.8106	1.9277	NO ORF
46	1	528	2101152	2101680	0.8717	2.4153	NO ORF
47	1	276	2132628	2132904	3.5015	2.0392	orf19.4833(W)
48	1	384	2146116	2146500	2.0964	0.8208	orf19.4844(C)
49	1	348	2222844	2223192	2.849	0.9491	NO ORF
50	1	792	2244588	2245380	1.8692	1.0628	NO ORF
51	1	204	2254608	2254812	1.0069	2.2618	NO ORF
52	1	324	2264640	2264964	2.0715	0.9233	orf19.4894(C)
53	1	648	2313972	2314620	0.9269	2.4059	NO ORF
54	1	324	2315700	2316024	0.676	1.8964	orf19.986(W)
55	1	1044	2325612	2326656	1.1649	2.0351	orf19.994(W)
56	1	336	2336592	2336928	2.5703	1.2625	NO ORF
57	1	432	2473476	2473908	1.0036	4.1572	orf19.676(W)
58	1	396	2649924	2650320	1.5956	2.101	orf19.5253(W)
59	1	252	2775564	2775816	0.9845	1.5944	orf19.6349(W)
60	1	1068	2866896	2867964	1.7018	0.7139	orf19.4940(C)

61	1	1476	3003192	3004668	1.2152	2.0544	orf19.5004(W)
62	1	660	3049956	3050616	1.1563	3.1535	NO ORF
63	1	300	3059484	3059784	0.85	2.0069	NO ORF
64	1	264	3127200	3127464	1.0904	3.5755	orf19.7227(C)
65	1	312	3132504	3132816	2.1826	0.6842	NO ORF
66	1	252	3135144	3135396	0.7539	2.0666	NO ORF
67	1	456	3161820	3162276	2.8946	0.9066	NO ORF

Table A6. List of Down-regulated AS and intergenic transcripts found in the data set OxStresst40/t0

Region #	Chr	Length	Start	End	Watson Y-mean	Crick Y-mean	Overlapping ORFs
1	1	804	30696	31500	1.0056	0.511	orf19.6086(W)
2	1	288	32724	33012	0.4155	1.173	orf19.6084(C)
3	1	1260	761400	762660	0.2589	0.4754	orf19.3066(W)
4	1	276	1204308	1204584	0.3655	0.7126	orf19.2475(C)
5	1	300	1534800	1535100	0.8879	0.3665	orf19.4450.2(W)
6	1	360	1832868	1833228	0.544	1.2135	orf19.5114.1(C)
7	1	276	1860684	1860960	0.5495	0.7693	NO ORF
8	1	636	2301600	2302236	0.3778	0.3449	orf19.4910(C)
9	1	1020	2430876	2431896	2.4704	0.5076	orf19.2312(W)
10	1	228	2567748	2567976	0.4828	0.8247	NO ORF
11	1	516	2623776	2624292	0.3661	0.4477	orf19.5267(W)
12	2	216	252	468	0.4802	1.0243	NO ORF
13	2	492	28980	29472	0.4933	1.11	orf19.2111(C)
14	2	288	49800	50088	0.7398	0.4749	NO ORF
15	2	1488	61344	62832	0.5023	0.5764	orf19.2090(C)
16	2	264	178032	178296	0.4714	1.8	NO ORF
17	2	492	250968	251460	0.7946	0.4687	NO ORF
18	2	444	253512	253956	0.5258	0.7837	NO ORF
19	2	288	571464	571752	0.7114	0.4772	orf19.5820(W)
20	2	552	647856	648408	0.4866	0.8855	orf19.5776(C)
21	2	252	676716	676968	0.3306	0.9653	NO ORF
22	2	204	962256	962460	0.8196	0.4654	NO ORF
23	2	276	1120308	1120584	0.2913	0.7462	NO ORF
24	2	660	1331076	1331736	1.4145	0.2594	orf19.31(W)
25	2	240	1425876	1426116	0.4302	0.7101	NO ORF
26	2	384	1545036	1545420	0.8666	0.4289	NO ORF
27	2	828	1747500	1748328	0.3428	0.4315	orf19.3618(W)
28	2	228	2008752	2008980	0.4507	0.6844	NO ORF
29	2	432	2128224	2128656	0.4061	0.7882	orf19.1746(C)

30	2	552	2169852	2170404	0.4689	0.789	orf19.5325(C)
31	3	228	80496	80724	0.7147	0.3731	NO ORF
32	3	240	130440	130680	0.4384	0.5611	NO ORF
33	3	468	456996	457464	0.8854	0.4459	orf19.1636(W)
34	3	360	1279380	1279740	1.2784	0.3984	orf19.7350(W)
35	3	456	1567920	1568376	0.4207	1.2312	NO ORF
36	3	360	1701180	1701540	1.0608	0.4496	orf19.6747(W) orf19.6747(W)
37	4	348	180204	180552	0.7262	0.4074	orf19.4697(C)
38	4	1296	493836	495132	0.3233	0.1423	orf19.2758(C)
39	4	696	827844	828540	1.4607	0.5029	orf19.5045(W)
40	4	300	1468236	1468536	0.2959	0.7141	orf19.2877(C)
41	5	444	153588	154032	0.7831	0.4868	NO ORF
42	5	996	673560	674556	0.8153	0.4987	orf19.4330(W) orf19.4332(W)
43	5	252	808188	808440	1.9719	0.4693	orf19.6674(W)
44	5	228	809808	810036	0.5255	0.4789	orf19.6676(W)
45	5	1284	916752	918036	0.2699	0.445	orf19.3893(W)
46	5	1224	1039248	1040472	0.7572	0.4833	orf19.3966(W)
47	5	384	1107420	1107804	0.2419	0.5886	NO ORF
48	6	840	132276	133116	0.5164	0.3482	orf19.3651(C)
49	6	216	486744	486960	0.4298	0.454	orf19.3475(W)
50	6	900	896424	897324	0.5429	0.8203	orf19.4557(C)
51	6	204	899916	900120	0.4969	0.9284	orf19.4555(C)
52	6	552	975852	976404	0.9816	0.5282	orf19.1097(W)
53	6	264	1018800	1019064	0.9682	0.5197	orf19.2151(W)
54	7	372	26400	26772	0.441	0.4242	orf19.7104(W)
55	7	972	637152	638124	0.4734	0.8866	orf19.5167(C)
56	7	480	935364	935844	0.5844	0.4527	orf19.7136(W)
57	R	696	15672	16368	0.4924	0.633	orf19.7538(C)
58	R	312	79428	79740	0.4859	0.6806	NO ORF
59	R	240	461772	462012	0.9185	0.4792	orf19.2611(W)
60	R	912	575028	575940	0.5265	1.8025	orf19.164(C)
61	R	2124	675204	677328	0.9102	0.5357	orf19.2850(W)
62	R	432	686028	686460	0.4766	0.5301	orf19.2859(C)
63	R	468	941280	941748	0.4212	0.2472	orf19.508(C)
64	R	444	1163784	1164228	1.0771	0.5093	orf19.3517(W)
65	R	372	1259460	1259832	1.0431	0.4712	NO ORF
66	R	324	1460316	1460640	0.9712	0.4141	NO ORF
67	R	336	1474992	1475328	0.3426	1.2174	NO ORF
68	R	348	1655748	1656096	1.0422	0.4967	orf19.6287(W)
69	R	288	1739376	1739664	0.488	0.691	orf19.600(C)

70	R	288	1752240	1752528	1.0561	0.4731	orf19.596.2(W) orf19.596.2(W)
71	R	528	1907316	1907844	0.4618	0.7498	orf19.7284(C)
72	R	336	2009556	2009892	0.9009	0.3913	NO ORF
73	R	384	2149248	2149632	0.9355	0.4071	orf19.7579(W)
74	R	1596	2161476	2163072	0.1859	0.3276	orf19.7586(W)

Table A7. List of Up-regulated AS and intergenic transcripts found in the data set OxStresst180/t0

Region #	Chr	Length (nt)	Start	End	Watson Y-mean	Crick Y-mean	Overlapping ORFs
1	1	264	2473584	2473848	0.9417	1.9701	orf19.676(W)
2	2	588	109644	110232	1.1612	1.9545	NO ORF
3	2	456	449640	450096	2.0205	0.9643	orf19.1542(C)
4	2	252	454608	454860	1.5103	0.9397	NO ORF
5	2	216	495588	495804	0.975	2.1176	NO ORF
6	2	732	496908	497640	1.8469	0.7864	orf19.1573(C)
7	2	1068	670464	671532	1.0514	1.9127	NO ORF
8	3	324	1647540	1647864	3.2189	1.0488	NO ORF
9	3	216	1687728	1687944	1.5499	1.8962	orf19.6758(W)
10	4	1428	851448	852876	1.3465	0.8767	NO ORF
11	4	312	1106772	1107084	1.0429	2.2339	NO ORF
12	4	504	1107504	1108008	1.7672	1.0749	NO ORF
13	6	1140	279168	280308	1.0672	1.5115	orf19.3444(W)
14	7	312	170208	170520	0.9983	1.6996	NO ORF
15	R	252	1560180	1560432	0.9556	2.093	orf19.734(W)
16	R	660	1677756	1678416	0.9577	1.6222	orf19.2633(W)
17	R	600	1938756	1939356	0.8448	2.7046	NO ORF

Table A8. List of Down-regulated AS and Intergenic transcripts found in the data set OxStresst180/t0

Region #	Chr	Length	Start	End	Watson Y-mean	Crick Y-mean	Overlapping Orfs
1	3	252	1279476	1279728	1.1028	0.5416	orf19.7350(W)

Related publications

1. Miguel A, Montón F, Li T, Gómez-Herreros F, Chávez S, Alepuz P, Pérez-Ortín JE. (2013). External conditions inversely change the RNA polymerase II elongation rate and density in yeast. *Biochim Biophys Acta*. 1829(11):1248-55. doi: 10.1016/j.bbagrm.2013.09.008

2. Juanes JM, Miguel A, Morales LJ, Pérez-Ortín JE, Arnau V. (2015) A web application for the unspecific detection of differentially expressed DNA regions in strand-specific expression data. Bioinformatics. First published online June 2, 2015 doi:10.1093/bioinformatics/btv343

3. Jordán-Pla, A, Miguel A, Serna E, Pelechano V, Pérez-Ortín JE.(2016) Biotin-Genomic Run-on (Bio-GRO): a high-resolution method for the analysis of nascent transcription in yeast. *Meth. Mol. Biol. series*. "Yeast functional genomics" volume. Frederic Devaux (ed.)

4. Benet M, Miguel A, Carrasco F, Li T, Planells J, Alepuz P, Tordera V, Pérez-Ortín JE. (2017). Modulation of protein synthesis and degradation maintains proteostasis during yeast growth at different temperatures. Biochim Biophys Acta- Gene regulatory mechanisms. (*In press*)

References

1 Pretorius, I. S. Tailoring wine yeast for the new millennium: novel approaches to the ancient art of winemaking. *Yeast* 16, 675-729 (2000).
2 Goffeau, A. *et al.* Life with 6000 genes. *Science* 274, 546, 563-547 (1996).
3 Cherry, J. M. *et al.* Saccharomyces Genome Database: the genomics resource of budding yeast. *Nucleic acids research* 40, D700-705 (2012).
4 Turnbaugh, P. J. *et al.* The human microbiome project. *Nature* 449, 804-810 (2007).
5 RA, C. *Candida and Candidiasis*. (ASM Press 2002).
6 Pfaller MA, D. D., Rinaldi MG, Barnes R, Hu B, Veselov AV, Tiraboschi N, & Nagy E, G. D. T. G. A. S. G. Results from the ARTEMIS DISK Global Antifungal Surveillance study: a 6.5-year analysis of susceptibilities of Candida and other yeast species to fluconazole and voriconazole by standarized disk diffusion testing. *Journal of Clinical Microbiololgy* 43, 5848-5859. (2005).
7 Odds, F. *The ecology of Candida and epidemiology of Candidosis*. 68-92 (Balliere Tindall, 1998).
8 Parker, J. C., Jr., McCloskey, J. J. & Knauer, K. A. Pathobiologic features of human candidiasis. A common deep mycosis of the brain, heart and kidney in the altered host. *American journal of clinical pathology* 65, 991-1000 (1976).
9 Pfaller, M. A. Nosocomial candidiasis: emerging species, reservoirs, and modes of transmission. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* 22 Suppl 2, S89-94 (1996).
10 Wisplinghoff, H. *et al.* Nosocomial bloodstream infections in US hospitals: analysis of 24,179 cases from a prospective nationwide surveillance study. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* 39, 309-317 (2004).
11 Todeschini, G. T. Treatment of candidiasis: a perspective on recent advances and future challenges. *Int. J. Infect. Dis.*, 1S37-S41 (1997).
12 Espinel-Ingroff, A. Mechanisms of resistance to antifungal agents: yeasts and filamentous fungi. *Revista iberoamericana de micologia* 25, 101-106 (2008).
13 Heckman, D. S. *et al.* Molecular evidence for the early colonization of land by fungi and plants. *Science* 293, 1129-1133 (2001).
14 Scannell, D. R., Butler, G. & Wolfe, K. H. Yeast genome evolution--the origin of the species. *Yeast* 24, 929-942 (2007).
15 Seoighe, C. *et al.* Prevalence of small inversions in yeast gene order evolution. *Proceedings of the National Academy of Sciences of the United States of America* 97, 14433-14437 (2000).
16 Hagman, A. & Piskur, J. A study on the fundamental mechanism and the evolutionary driving forces behind aerobic fermentation in yeast. *PloS one* 10, e0116942 (2015).
17 Bennett, R. J. & Johnson, A. D. Mating in Candida albicans and the search for a sexual cycle. *Annual review of microbiology* 59, 233-255 (2005).
18 Heitman, J. Sexual reproduction and the evolution of microbial pathogens. *Current biology : CB* 16, R711-725 (2006).
19 Thompson, D. A. *et al.* Evolutionary principles of modular gene regulation in yeasts. *eLife* 2, e00603 (2013).
20 Duina, A. A., Miller, M. E. & Keeney, J. B. Budding yeast for budding geneticists: a primer on the Saccharomyces cerevisiae model system. *Genetics* 197, 33-48 (2014).

21 Merson-Davies, L. A. & Odds, F. C. A morphology index for characterization of cell shape in Candida albicans. *Journal of general microbiology* 135, 3143-3152 (1989).
22 Odds, F. C. Morphogenesis in Candida albicans. *Critical reviews in microbiology* 12, 45-93 (1985).
23 Gow, N. A. Germ tube growth of Candida albicans. *Current topics in medical mycology* 8, 43-55 (1997).
24 Zaragoza, O. & Gancedo, J. M. Pseudohyphal growth is induced in Saccharomyces cerevisiae by a combination of stress and cAMP signalling. *Antonie van Leeuwenhoek* 78, 187-194 (2000).
25 Lo, H. J. *et al.* Nonfilamentous C. albicans mutants are avirulent. *Cell* 90, 939-949 (1997).
26 Gow, N. A., van de Veerdonk, F. L., Brown, A. J. & Netea, M. G. Candida albicans morphogenesis and host defence: discriminating invasion from colonization. *Nature reviews. Microbiology* 10, 112-122 (2012).
27 Sudbery, P., Gow, N. & Berman, J. The distinct morphogenic states of Candida albicans. *Trends in microbiology* 12, 317-324 (2004).
28 Wickes, B. *et al.* Physical and genetic mapping of Candida albicans: several genes previously assigned to chromosome 1 map to chromosome R, the rDNA-containing linkage group. *Infection and immunity* 59, 2480-2484 (1991).
29 Herrero, E., de la Torre, M. A. & Valentin, E. Comparative genomics of yeast species: new insights into their biology. *International microbiology : the official journal of the Spanish Society for Microbiology* 6, 183-190 (2003).
30 Berman, J. & Sudbery, P. E. Candida Albicans: a molecular revolution built on lessons from budding yeast. *Nature reviews. Genetics* 3, 918-930 (2002).
31 Enjalbert, B. *et al.* Role of the Hog1 stress-activated protein kinase in the global transcriptional response to stress in the fungal pathogen Candida albicans. *Molecular biology of the cell* 17, 1018-1032 (2006).
32 Noble, S. M. & Johnson, A. D. Genetics of Candida albicans, a diploid human fungal pathogen. *Annual review of genetics* 41, 193-211 (2007).
33 Sentenac, A. Eukaryotic RNA polymerases. *CRC critical reviews in biochemistry* 18, 31-90 (1985).
34 Smale, S. T. & Kadonaga, J. T. The RNA polymerase II core promoter. *Annual review of biochemistry* 72, 449-479 (2003).
35 Tsukiyama, T. The in vivo functions of ATP-dependent chromatin-remodelling factors. *Nature reviews. Molecular cell biology* 3, 422-429 (2002).
36 Butler, J. E. & Kadonaga, J. T. The RNA polymerase II core promoter: a key component in the regulation of gene expression. *Genes & development* 16, 2583-2592 (2002).
37 Basehoar, A. D., Zanton, S. J. & Pugh, B. F. Identification and distinct regulation of yeast TATA box-containing genes. *Cell* 116, 699-709 (2004).
38 Rhee, H. S. & Pugh, B. F. Genome-wide structure and organization of eukaryotic pre-initiation complexes. *Nature* 483, 295-301, doi:10.1038/nature10799 (2012).
39 Perez-Ortin, J. E., Matallana, E. & Franco, L. Chromatin structure of yeast genes. *Yeast* 5, 219-238 (1989).
40 Yuan, G. C. *et al.* Genome-scale identification of nucleosome positions in S. cerevisiae. *Science* 309, 626-630 (2005).
41 Lee, W. *et al.* A high-resolution atlas of nucleosome occupancy in yeast. *Nature genetics* 39, 1235-1244 (2007).
42 Mavrich, T. N. *et al.* A barrier nucleosome model for statistical positioning of nucleosomes throughout the yeast genome. *Genome research* 18, 1073-1083 (2008).

43 Yassour, M., Kaplan, T., Jaimovich, A. & Friedman, N. Nucleosome positioning from tiling microarray data. *Bioinformatics* 24, i139-146 (2008).
44 Brogaard, K., Xi, L., Wang, J. P. & Widom, J. A map of nucleosome positions in yeast at base-pair resolution. *Nature* 486, 496-501 (2012).
45 Jiang, C. & Pugh, B. F. Nucleosome positioning and gene regulation: advances through genomics. *Nature reviews. Genetics* 10, 161-172 (2009).
46 Lorch, Y., LaPointe, J. W. & Kornberg, R. D. Nucleosomes inhibit the initiation of transcription but allow chain elongation with the displacement of histones. *Cell* 49, 203-210 (1987).
47 Han, M. & Grunstein, M. Nucleosome loss activates yeast downstream promoters in vivo. *Cell* 55, 1137-1145 (1988).
48 Boeger, H., Griesenbeck, J., Strattan, J. S. & Kornberg, R. D. Nucleosomes unfold completely at a transcriptionally active promoter. *Molecular cell* 11, 1587-1598 (2003).
49 Cramer, P. RNA polymerase II structure: from core to functional complexes. *Current opinion in genetics & development* 14, 218-226 (2004).
50 Woychik, N. A. & Young, R. A. RNA polymerase II subunit RPB4 is essential for high- and low-temperature yeast cell growth. *Molecular and cellular biology* 9, 2854-2859 (1989).
51 Woychik, N. A., Lane, W. S. & Young, R. A. Yeast RNA polymerase II subunit RPB9 is essential for growth at temperature extremes. *The Journal of biological chemistry* 266, 19053-19055 (1991).
52 Edwards, A. M., Kane, C. M., Young, R. A. & Kornberg, R. D. Two dissociable subunits of yeast RNA polymerase II stimulate the initiation of transcription at a promoter in vitro. *The Journal of biological chemistry* 266, 71-75 (1991).
53 Gnatt, A. L., Cramer, P., Fu, J., Bushnell, D. A. & Kornberg, R. D. Structural basis of transcription: an RNA polymerase II elongation complex at 3.3 A resolution. *Science* 292, 1876-1882 (2001).
54 Armache, K. J., Mitterweger, S., Meinhart, A. & Cramer, P. Structures of complete RNA polymerase II and its subcomplex, Rpb4/7. *The Journal of biological chemistry* 280, 7131-7134 (2005).
55 Meinhart, A., Kamenski, T., Hoeppner, S., Baumli, S. & Cramer, P. A structural perspective of CTD function. *Genes & development* 19, 1401-1415 (2005).
56 Spahr, H., Calero, G., Bushnell, D. A. & Kornberg, R. D. Schizosacharomyces pombe RNA polymerase II at 3.6-A resolution. *Proceedings of the National Academy of Sciences of the United States of America* 106, 9185-9190 (2009).
57 Conaway, R. C. & Conaway, J. W. General transcription factors for RNA polymerase II. *Progress in nucleic acid research and molecular biology* 56, 327-346 (1997).
58 Kelleher, R. J., 3rd, Flanagan, P. M. & Kornberg, R. D. A novel mediator between activator proteins and the RNA polymerase II transcription apparatus. *Cell* 61, 1209-1215 (1990).
59 Kim, Y. J., Bjorklund, S., Li, Y., Sayre, M. H. & Kornberg, R. D. A multiprotein mediator of transcriptional activation and its interaction with the C-terminal repeat domain of RNA polymerase II. *Cell* 77, 599-608 (1994).
60 Gustafsson, C. M. *et al.* Identification of new mediator subunits in the RNA polymerase II holoenzyme from Saccharomyces cerevisiae. *The Journal of biological chemistry* 273, 30851-30854 (1998).
61 Mannervik, M., Nibu, Y., Zhang, H. & Levine, M. Transcriptional coregulators in development. *Science* 284, 606-609 (1999).
62 Malik, S. & Roeder, R. G. Dynamic regulation of pol II transcription by the mammalian Mediator complex. *Trends in biochemical sciences* 30, 256-263 (2005).

63 Kornberg, R. D. Mediator and the mechanism of transcriptional activation. *Trends in biochemical sciences* 30, 235-239 (2005).
64 Thompson, C. M. & Young, R. A. General requirement for RNA polymerase II holoenzymes in vivo. *Proceedings of the National Academy of Sciences of the United States of America* 92, 4587-4590 (1995).
65 Takagi, Y. & Kornberg, R. D. Mediator as a general transcription factor. *The Journal of biological chemistry* 281, 80-89 (2006).
66 Guo, J. & Price, D. H. RNA polymerase II transcription elongation control. *Chemical reviews* 113, 8583-8603 (2013).
67 Mischo, H. E. & Proudfoot, N. J. Disengaging polymerase: terminating RNA polymerase II transcription in budding yeast. *Biochimica et biophysica acta* 1829, 174-185 (2013).
68 Cheung, A. C. & Cramer, P. A movie of RNA polymerase II transcription. *Cell* 149, 1431-1437 (2012).
69 Roeder, R. G. The role of general initiation factors in transcription by RNA polymerase II. *Trends in biochemical sciences* 21, 327-335 (1996).
70 van Helden, J., del Olmo, M. & Perez-Ortin, J. E. Statistical analysis of yeast genomic downstream sequences reveals putative polyadenylation signals. *Nucleic acids research* 28, 1000-1010 (2000).
71 Lee, T. I. *et al.* Redundant roles for the TFIID and SAGA complexes in global transcription. *Nature* 405, 701-704 (2000).
72 Huisinga, K. L. & Pugh, B. F. A genome-wide housekeeping role for TFIID and a highly regulated stress-related role for SAGA in Saccharomyces cerevisiae. *Molecular cell* 13, 573-585 (2004).
73 Grant, P. A. *et al.* A subset of TAF(II)s are integral components of the SAGA complex required for nucleosome acetylation and transcriptional stimulation. *Cell* 94, 45-53 (1998).
74 Hahn, S. & Young, E. T. Transcriptional regulation in Saccharomyces cerevisiae: transcription factor regulation and function, mechanisms of initiation, and roles of activators and coactivators. *Genetics* 189, 705-736 (2011).
75 Shandilya, J. & Roberts, S. G. The transcription cycle in eukaryotes: from productive initiation to RNA polymerase II recycling. *Biochimica et biophysica acta* 1819, 391-400 (2012).
76 Zawel, L., Kumar, K. P. & Reinberg, D. Recycling of the general transcription factors during RNA polymerase II transcription. *Genes & development* 9, 1479-1490 (1995).
77 Pokholok, D. K., Hannett, N. M. & Young, R. A. Exchange of RNA polymerase II initiation and elongation factors during gene expression in vivo. *Molecular cell* 9, 799-809 (2002).
78 Hafner, M. *et al.* Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell* 141, 129-141 (2010).
79 Nudler, E. RNA polymerase backtracking in gene regulation and genome instability. *Cell* 149, 1438-1445 (2012).
80 Gilchrist, D. A. *et al.* NELF-mediated stalling of Pol II can enhance gene expression by blocking promoter-proximal nucleosome assembly. *Genes & development* 22, 1921-1933 (2008).
81 Lagha, M. *et al.* Paused Pol II coordinates tissue morphogenesis in the Drosophila embryo. *Cell* 153, 976-987 (2013).
82 Gomez-Herreros, F. *et al.* One step back before moving forward: regulation of transcription elongation by arrest and backtracking. *FEBS letters* 586, 2820-2825 (2012).

83 Martinez-Rucobo, F. W., Sainsbury, S., Cheung, A. C. & Cramer, P. Architecture of the RNA polymerase-Spt4/5 complex and basis of universal transcription processivity. *The EMBO journal* 30, 1302-1310 (2011).
84 Core, L. J., Waterfall, J. J. & Lis, J. T. Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science* 322, 1845-1848 (2008).
85 Sigova, A. A. *et al.* Divergent transcription of long noncoding RNA/mRNA gene pairs in embryonic stem cells. *Proceedings of the National Academy of Sciences of the United States of America* 110, 2876-2881 (2013).
86 Miguel, A. *et al.* External conditions inversely change the RNA polymerase II elongation rate and density in yeast. *Biochim Biophys Acta* 1829, 1248-1255 (2013).
87 Krogan, N. J. & Greenblatt, J. F. Characterization of a six-subunit holo-elongator complex required for the regulated expression of a group of genes in Saccharomyces cerevisiae. *Molecular and cellular biology* 21, 8203-8212 (2001).
88 Govind, C. K., Zhang, F., Qiu, H., Hofmeyer, K. & Hinnebusch, A. G. Gcn5 promotes acetylation, eviction, and methylation of nucleosomes in transcribed coding regions. *Molecular cell* 25, 31-42 (2007).
89 Kim, J. H., Saraf, A., Florens, L., Washburn, M. & Workman, J. L. Gcn5 regulates the dissociation of SWI/SNF from chromatin by acetylation of Swi2/Snf2. *Genes & development* 24, 2766-2771 (2010).
90 Strahl, B. D. *et al.* Set2 is a nucleosomal histone H3-selective methyltransferase that mediates transcriptional repression. *Molecular and cellular biology* 22, 1298-1306 (2002).
91 Formosa, T. *et al.* Spt16-Pob3 and the HMG protein Nhp6 combine to form the nucleosome-binding factor SPN. *The EMBO journal* 20, 3506-3517 (2001).
92 Kaplan, C. D., Laprade, L. & Winston, F. Transcription elongation factors repress transcription initiation from cryptic sites. *Science* 301, 1096-1099 (2003).
93 Graber, J. H., Cantor, C. R., Mohr, S. C. & Smith, T. F. Genomic detection of new yeast pre-mRNA 3'-end-processing signals. *Nucleic acids research* 27, 888-894 (1999).
94 Pelechano, V., Wei, W. & Steinmetz, L. M. Extensive transcriptional heterogeneity revealed by isoform profiling. *Nature* 497, 127-131 (2013).
95 Moqtaderi, Z., Geisberg, J. V., Jin, Y., Fan, X. & Struhl, K. Species-specific factors mediate extensive heterogeneity of mRNA 3' ends in yeasts. *Proceedings of the National Academy of Sciences of the United States of America* 110, 11073-11078 (2013).
96 Gudipati, R. K., Villa, T., Boulay, J. & Libri, D. Phosphorylation of the RNA polymerase II C-terminal domain dictates transcription termination choice. *Nature structural & molecular biology* 15, 786-794 (2008).
97 Steinmetz, E. J. *et al.* Genome-wide distribution of yeast RNA polymerase II and its control by Sen1 helicase. *Molecular cell* 24, 735-746 (2006).
98 Jenks, M. H., O'Rourke, T. W. & Reines, D. Properties of an intergenic terminator and start site switch that regulate IMD2 transcription in yeast. *Molecular and cellular biology* 28, 3883-3893 (2008).
99 Logan, J., Falck-Pedersen, E., Darnell, J. E., Jr. & Shenk, T. A poly(A) addition site and a downstream termination region are required for efficient cessation of transcription by RNA polymerase II in the mouse beta maj-globin gene. *Proceedings of the National Academy of Sciences of the United States of America* 84, 8306-8310 (1987).
100 Whitelaw, E. & Proudfoot, N. Alpha-thalassaemia caused by a poly(A) site mutation reveals that transcriptional termination is linked to 3' end processing in the human alpha 2 globin gene. *The EMBO journal* 5, 2915-2922 (1986).

101 Park, N. J., Tsao, D. C. & Martinson, H. G. The two steps of poly(A)-dependent termination, pausing and release, can be uncoupled by truncation of the RNA polymerase II carboxyl-terminal repeat domain. *Molecular and cellular biology* 24, 4092-4103 (2004).
102 Nag, A., Narsinh, K. & Martinson, H. G. The poly(A)-dependent transcriptional pause is mediated by CPSF acting on the body of the polymerase. *Nature structural & molecular biology* 14, 662-669 (2007).
103 Kazerouninia, A., Ngo, B. & Martinson, H. G. Poly(A) signal-dependent degradation of unprocessed nascent transcripts accompanies poly(A) signal-dependent transcriptional pausing in vitro. *Rna* 16, 197-210 (2010).
104 Millevoi, S. & Vagner, S. Molecular mechanisms of eukaryotic pre-mRNA 3' end processing regulation. *Nucleic acids research* 38, 2757-2774 (2010).
105 Connelly, S. & Manley, J. L. A functional mRNA polyadenylation signal is required for transcription termination by RNA polymerase II. *Genes & development* 2, 440-452 (1988).
106 Kim, M. *et al.* The yeast Rat1 exonuclease promotes transcription termination by RNA polymerase II. *Nature* 432, 517-522 (2004).
107 West, S., Gromak, N. & Proudfoot, N. J. Human 5' --> 3' exonuclease Xrn2 promotes transcription termination at co-transcriptional cleavage sites. *Nature* 432, 522-525 (2004).
108 Kuehner, J. N., Pearson, E. L. & Moore, C. Unravelling the means to an end: RNA polymerase II transcription termination. *Nature reviews. Molecular cell biology* 12, 283-294 (2011).
109 Houseley, J. & Tollervey, D. The many pathways of RNA degradation. *Cell* 136, 763-776 (2009).
110 Kim, H. D., Choe, J. & Seo, Y. S. The sen1(+) gene of Schizosaccharomyces pombe, a homologue of budding yeast SEN1, encodes an RNA and DNA helicase. *Biochemistry* 38, 14697-14710 (1999).
111 Steinmetz, E. J. & Brow, D. A. Repression of gene expression by an exogenous sequence element acting in concert with a heterogeneous nuclear ribonucleoprotein-like protein, Nrd1, and the putative helicase Sen1. *Molecular and cellular biology* 16, 6993-7003 (1996).
112 Steinmetz, E. J., Conrad, N. K., Brow, D. A. & Corden, J. L. RNA-binding protein Nrd1 directs poly(A)-independent 3'-end formation of RNA polymerase II transcripts. *Nature* 413, 327-331 (2001).
113 Ursic, D., Chinchilla, K., Finkel, J. S. & Culbertson, M. R. Multiple protein/protein and protein/RNA interactions suggest roles for yeast DNA/RNA helicase Sen1p in transcription, transcription-coupled DNA repair and RNA processing. *Nucleic acids research* 32, 2441-2452 (2004).
114 Finkel, J. S., Chinchilla, K., Ursic, D. & Culbertson, M. R. Sen1p performs two genetically separable functions in transcription and processing of U5 small nuclear RNA in Saccharomyces cerevisiae. *Genetics* 184, 107-118 (2010).
115 Arigo, J. T., Eyler, D. E., Carroll, K. L. & Corden, J. L. Termination of cryptic unstable transcripts is directed by yeast RNA-binding proteins Nrd1 and Nab3. *Molecular cell* 23, 841-851 (2006).
116 Thiebaut, M., Kisseleva-Romanova, E., Rougemaille, M., Boulay, J. & Libri, D. Transcription termination and nuclear degradation of cryptic unstable transcripts: a role for the nrd1-nab3 pathway in genome surveillance. *Molecular cell* 23, 853-864 (2006).
117 Kawauchi, J., Mischo, H., Braglia, P., Rondon, A. & Proudfoot, N. J. Budding yeast RNA polymerases I and II employ parallel mechanisms of transcriptional termination. *Genes & development* 22, 1082-1092 (2008).
118 Buratowski, S. Progression through the RNA polymerase II CTD cycle. *Molecular cell* 36, 541-54 (2009).

119 Moteki, S. & Price, D. Functional coupling of capping and transcription of mRNA. *Molecular cell* 10, 599-609 (2002).
120 Rasmussen, E. B. & Lis, J. T. In vivo transcriptional pausing and cap formation on three Drosophila heat shock genes. *Proceedings of the National Academy of Sciences of the United States of America* 90, 7923-7927 (1993).
121 Shatkin, A. J. & Manley, J. L. The ends of the affair: capping and polyadenylation. *Nature structural biology* 7, 838-842 (2000).
122 Ho, C. K. & Shuman, S. Distinct roles for CTD Ser-2 and Ser-5 phosphorylation in the recruitment and allosteric activation of mammalian mRNA capping enzyme. *Molecular cell* 3, 405-411 (1999).
123 McCracken, S. *et al.* 5'-Capping enzymes are targeted to pre-mRNA by binding to the phosphorylated carboxy-terminal domain of RNA polymerase II. *Genes & development* 11, 3306-3318 (1997).
124 Cho, E. J., Takagi, T., Moore, C. R. & Buratowski, S. mRNA capping enzyme is recruited to the transcription complex by phosphorylation of the RNA polymerase II carboxy-terminal domain. *Genes & development* 11, 3319-3326 (1997).
125 Affymetrix, E. T. P. & Cold Spring Harbor Laboratory, E. T. P. Post-transcriptional processing generates a diversity of 5'-modified long and short RNAs. *Nature* 457, 1028-1032 (2009).
126 Otsuka, Y., Kedersha, N. L. & Schoenberg, D. R. Identification of a cytoplasmic complex that adds a cap onto 5'-monophosphate RNA. *Molecular and cellular biology* 29, 2155-2167 (2009).
127 Topisirovic, I., Svitkin, Y. V., Sonenberg, N. & Shatkin, A. J. Cap and cap-binding proteins in the control of gene expression. *Wiley interdisciplinary reviews. RNA* 2, 277-298 (2011).
128 Rodriguez-Navarro, S., Igual, J. C. & Perez-Ortin, J. E. SRC1: an intron-containing yeast gene involved in sister chromatid segregation. *Yeast* 19, 43-54 (2002).
129 Jurica, M. S. & Moore, M. J. Pre-mRNA splicing: awash in a sea of proteins. *Molecular cell* 12, 5-14 (2003).
130 Carrillo Oesterreich, F., Preibisch, S. & Neugebauer, K. M. Global analysis of nascent RNA reveals transcriptional pausing in terminal exons. *Molecular cell* 40, 571-581 (2010).
131 Alexander, R. D. *et al.* RiboSys, a high-resolution, quantitative approach to measure the in vivo kinetics of pre-mRNA splicing and 3'-end processing in Saccharomyces cerevisiae. *Rna* 16, 2570-2580 (2010).
132 Churchman, L. S. & Weissman, J. S. Nascent transcript sequencing visualizes transcription at nucleotide resolution. *Nature* 469, 368-373 (2011).
133 Carrillo Oesterreich, F., Bieberstein, N. & Neugebauer, K. M. Pause locally, splice globally. *Trends in cell biology* 21, 328-335 (2011).
134 Braberg, H. *et al.* From structure to systems: high-resolution, quantitative genetic analysis of RNA polymerase II. *Cell* 154, 775-788 (2013).
135 Braun, B. R. *et al.* A human-curated annotation of the Candida albicans genome. *PLoS genetics* 1, 36-57 (2005).
136 Kersanach, R. *et al.* Five identical intron positions in ancient duplicated genes of eubacterial origin. *Nature* 367, 387-389 (1994).
137 Beiter, T., Reich, E., Williams, R. W. & Simon, P. Antisense transcription: a critical look in both directions. *Cellular and molecular life sciences : CMLS* 66, 94-112 (2009).
138 Berretta, J. & Morillon, A. Pervasive transcription constitutes a new level of eukaryotic genome regulation. *EMBO reports* 10, 973-982 (2009).

139 Khalil, A. M. *et al.* Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. *Proceedings of the National Academy of Sciences of the United States of America* 106, 11667-11672, (2009).
140 Guttman, M. *et al.* Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature* 458, 223-227 (2009).
141 Moazed, D. Small RNAs in transcriptional gene silencing and genome defence. *Nature* 457, 413-420 (2009).
142 Ozsolak, F. *et al.* Comprehensive polyadenylation site maps in yeast and human reveal pervasive alternative polyadenylation. *Cell* 143, 1018-1029 (2010).
143 Dutrow, N. *et al.* Dynamic transcriptome of Schizosaccharomyces pombe shown by RNA-DNA hybrid mapping. *Nature genetics* 40, 977-986 (2008).
144 Wilhelm, B. T. *et al.* Dynamic repertoire of a eukaryotic transcriptome surveyed at single-nucleotide resolution. *Nature* 453, 1239-1243 (2008).
145 Ni, T. *et al.* The prevalence and regulation of antisense transcripts in Schizosaccharomyces pombe. *PloS one* 5, e15271 (2010).
146 Neil, H. *et al.* Widespread bidirectional promoters are the major source of cryptic transcripts in yeast. *Nature* 457, 1038-10427 (2009).
147 Xu, Z. *et al.* Bidirectional promoters generate pervasive transcription in yeast. *Nature* 457, 1033-1037 (2009).
148 Yassour, M. *et al.* Strand-specific RNA sequencing reveals extensive regulated long antisense transcripts that are conserved across yeast species. *Genome biology* 11, R87 (2010).
149 Venkatesh, S., Li, H., Gogol, M. M. & Workman, J. L. Selective suppression of antisense transcription by Set2-mediated H3K36 methylation. *Nature communications* 7, 13610 (2016).
150 Gomez-Navarro, N., Jordan-Pla, A., Estruch, F. & J, E. P.-O. Defects in the NC2 repressor affect both canonical and non-coding RNA polymerase II transcription initiation in yeast. *BMC genomics* 17, 183 (2016).
151 Sellam, A. *et al.* Experimental annotation of the human pathogen Candida albicans coding and noncoding transcribed regions using high-resolution tiling arrays. *Genome Biol* 11, R71 (2010).
152 Cleary, I. A., Lazzell, A. L., Monteagudo, C., Thomas, D. P. & Saville, S. P. BRG1 and NRG1 form a novel feedback circuit regulating Candida albicans hypha formation and virulence. *Molecular microbiology* 85, 557-573 (2012).
153 Wang, Z., Gerstein, M. & Snyder, M. RNA-Seq: a revolutionary tool for transcriptomics. *Nature reviews. Genetics* 10, 57-63 (2009).
154 Perez-Ortin, J. E., Medina, D. A., Chavez, S. & Moreno, J. What do you mean by transcription rate?: the conceptual difference between nascent transcription rate and mRNA synthesis rate is essential for the proper understanding of transcriptomic analyses. *BioEssays : news and reviews in molecular, cellular and developmental biology* 35, 1056-1062 (2013).
155 Perez-Ortin, J. E., Alepuz, P. M. & Moreno, J. Genomics and gene transcription kinetics in yeast. *Trends in genetics : TIG* 23, 250-257 (2007).
156 Hirayoshi, K. & Lis, J. T. Nuclear run-on assays: assessing transcription by measuring density of engaged RNA polymerases. *Methods in enzymology* 304, 351-362 (1999).
157 Haimovich, G. *et al.* Gene expression is circular: factors for mRNA degradation also foster mRNA synthesis. *Cell* 153, 1000-1011 (2013).
158 Garcia-Martinez, J., Aranda, A. & Perez-Ortin, J. E. Genomic run-on evaluates transcription rates for all yeast genes and identifies gene regulatory mechanisms. *Molecular cell* 15, 303-313 (2004).

159 Jordan-Pla, A. *et al.* Chromatin-dependent regulation of RNA polymerases II and III activity throughout the transcription cycle. *Nucleic acids research* 43, 787-802 (2015).
160 Sandoval, J. *et al.* RNAPol-ChIP: a novel application of chromatin immunoprecipitation to the analysis of real-time gene transcription. *Nucleic acids research* 32, e88 (2004).
161 Alepuz, P. M., de Nadal, E., Zapater, M., Ammerer, G. & Posas, F. Osmostress-induced transcription by Hot1 depends on a Hog1-mediated recruitment of the RNA Pol II. *The EMBO journal* 22, 2433-2442 (2003).
162 Pelechano, V., Chavez, S. & Perez-Ortin, J. E. A complete set of nascent transcription rates for yeast genes. *PloS one* 5, e15442 (2010).
163 Perez-Ortin, J. E., de Miguel-Jimenez, L. & Chavez, S. Genome-wide studies of mRNA synthesis and degradation in eukaryotes. *Biochimica et biophysica acta* 1819, 604-615 (2012).
164 Venters, B. J. & Pugh, B. F. A canonical promoter organization of the transcription machinery and its regulators in the Saccharomyces genome. *Genome research* 19, 360-371 (2009).
165 Pelechano, V. & Perez-Ortin, J. E. There is a steady-state transcriptome in exponentially growing yeast cells. *Yeast* 27, 413-422 (2010).
166 Castells-Roca, L. *et al.* Heat shock response in yeast involves changes in both transcription rates and mRNA stabilities. *PloS one* 6, e17272 (2011).
167 Jordan-Pla, A., Miguel, A., Serna, E., Pelechano, V. & Perez-Ortin, J. E. Biotin-Genomic Run-On (Bio-GRO): A High-Resolution Method for the Analysis of Nascent Transcription in Yeast. *Methods in molecular biology* 1361, 125-139 (2016).
168 McKinlay, A., Araya, C. L. & Fields, S. Genome-Wide Analysis of Nascent Transcription in Saccharomyces cerevisiae. *G3* 1, 549-558 (2011).
169 Jackson, D. A., Iborra, F. J., Manders, E. M. & Cook, P. R. Numbers and organization of RNA polymerases, nascent transcripts, and transcription units in HeLa nuclei. *Molecular biology of the cell* 9, 1523-1536 (1998).
170 Bobard, A., Mellouk, N. & Enninga, J. Spotting the right location- imaging approaches to resolve the intracellular localization of invasive pathogens. *Biochimica et biophysica acta* 1810, 297-307 (2011).
171 Scandalios, J. G. Oxidative stress responses--what have genome-scale studies taught us? *Genome biology* 3 (2002).
172 Wilson, D. *et al.* Identifying infection-associated genes of Candida albicans in the postgenomic era. *FEMS yeast research* 9, 688-700 (2009).
173 Csardi, G., Franks, A., Choi, D. S., Airoldi, E. M. & Drummond, D. A. Accounting for experimental noise reveals that mRNA levels, amplified by post-transcriptional processes, largely determine steady-state protein levels in yeast. *PLoS genetics* 11, e1005206 (2015).
174 Li, J. J. & Biggin, M. D. Gene expression. Statistics requantitates the central dogma. *Science* 347, 1066-1067 (2015).
175 Goldfine, H. The evolution of oxygen as a biosynthetic reagent. *The Journal of general physiology* 49, Suppl:253-274 (1965).
176 Raymond, J. & Segre, D. The effect of oxygen on biochemical networks and the evolution of complex life. *Science* 311, 1764-1767 (2006).
177 Summons, R. E., Bradley, A. S., Jahnke, L. L. & Waldbauer, J. R. Steroids, triterpenoids and molecular oxygen. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 361, 951-968 (2006).
178 Grahl, N., Shepardson, K. M., Chung, D. & Cramer, R. A. Hypoxia and fungal pathogenesis: to air or not to air? *Eukaryotic cell* 11, 560-570 (2012).

179 Cramer, T. *et al.* HIF-1alpha is essential for myeloid cell-mediated inflammation. *Cell* 112, 645-657 (2003).
180 Nizet, V. & Johnson, R. S. Interdependence of hypoxic and innate immune responses. *Nature reviews. Immunology* 9, 609-617 (2009).
181 Oill, P. A. *et al.* Infectious disease emergencies. Part III: Patients presenting with respiratory distress syndromes. *The Western journal of medicine* 125, 452-478 (1976).
182 Peyssonaux, C. & Johnson, R. S. An unexpected role for hypoxic response: oxygenation and inflammation. *Cell cycle* 3, 168-171 (2004).
183 Arnold, F., West, D. & Kumar, S. Wound healing: the effect of macrophage and tumour derived angiogenesis factors on skin graft vascularization. *British journal of experimental pathology* 68, 569-574 (1987).
184 Dewhirst, M. W. Concepts of oxygen transport at the microcirculatory level. *Seminars in radiation oncology* 8, 143-150 (1998).
185 Matherne, G. P., Headrick, J. P., Coleman, S. D. & Berne, R. M. Interstitial transudate purines in normoxic and hypoxic immature and mature rabbit hearts. *Pediatric research* 28, 348-353 (1990).
186 Simmen, H. P., Battaglia, H., Giovanoli, P. & Blaser, J. Analysis of pH, pO2 and pCO2 in drainage fluid allows for rapid detection of infectious complications during the follow-up period after abdominal surgery. *Infection* 22, 386-389 (1994).
187 Van Belle, H., Goossens, F. & Wynants, J. Formation and release of purine catabolites during hypoperfusion, anoxia, and ischemia. *The American journal of physiology* 252, H886-893 (1987).
188 He, G. *et al.* Noninvasive measurement of anatomic structure and intraluminal oxygenation in the gastrointestinal tract of living mice with spatial and spectral EPR imaging. *Proceedings of the National Academy of Sciences of the United States of America* 96, 4586-4591 (1999).
189 Karhausen, J. *et al.* Epithelial hypoxia-inducible factor-1 is protective in murine experimental colitis. *The Journal of clinical investigation* 114, 1098-1106 (2004).
190 Johnston, M. Feasting, fasting and fermenting. Glucose sensing in yeast and other cells. *Trends in genetics : TIG* 15, 29-33 (1999).
191 Bard, M. *et al.* Sterol uptake in Candida glabrata: rescue of sterol auxotrophic strains. *Diagnostic microbiology and infectious disease* 52, 285-293 (2005).
192 Sellam, A. *et al.* Modeling the transcriptional regulatory network that controls the early hypoxic response in Candida albicans. *Eukaryot Cell* 13, 675-690, (2014).
193 Setiadi, E. R., Doedt, T., Cottier, F., Noffz, C. & Ernst, J. F. Transcriptional response of Candida albicans to hypoxia: linkage of oxygen sensing and Efg1p-regulatory networks. *Journal of molecular biology* 361, 399-411 (2006).
194 Synnott, J. M., Guida, A., Mulhern-Haughey, S., Higgins, D. G. & Butler, G. Regulation of the hypoxic response in Candida albicans. *Eukaryotic cell* 9, 1734-1746 (2010).
195 Askew, C. *et al.* Transcriptional regulation of carbohydrate metabolism in the human pathogen Candida albicans. *PLoS pathogens* 5, e1000612 (2009).
196 Hon, T. *et al.* A mechanism of oxygen sensing in yeast. Multiple oxygen-responsive steps in the heme biosynthetic pathway affect Hap1 activity. *The Journal of biological chemistry* 278, 50771-50780 (2003).
197 Mennella, T. A., Klinkenberg, L. G. & Zitomer, R. S. Recruitment of Tup1-Ssn6 by yeast hypoxic genes and chromatin-independent exclusion of TATA binding protein. *Eukaryotic cell* 2, 1288-1303 (2003).
198 Becerra, M. *et al.* The yeast transcriptome in aerobic and hypoxic conditions: effects of hap1, rox1, rox3 and srb10 deletions. *Molecular microbiology* 43, 545-555 (2002).

199 Rape, M. *et al.* Mobilization of processed, membrane-tethered SPT23 transcription factor by CDC48(UFD1/NPL4), a ubiquitin-selective chaperone. *Cell* 107, 667-677 (2001).
200 Vasconcelles, M. J. *et al.* Identification and characterization of a low oxygen response element involved in the hypoxic induction of a family of Saccharomyces cerevisiae genes. Implications for the conservation of oxygen sensing in eukaryotes. *The Journal of biological chemistry* 276, 14374-14384 (2001).
201 Hughes, A. L., Todd, B. L. & Espenshade, P. J. SREBP pathway responds to sterols and functions as an oxygen sensor in fission yeast. *Cell* 120, 831-842 (2005).
202 Davies, B. S. & Rine, J. A role for sterol levels in oxygen sensing in Saccharomyces cerevisiae. *Genetics* 174, 191-201 (2006).
203 Dirmeier, R. *et al.* Exposure of yeast cells to anoxia induces transient oxidative stress. Implications for the induction of hypoxic genes. *The Journal of biological chemistry* 277 (2002).
204 Castello, P. R., David, P. S., McClure, T., Crook, Z. & Poyton, R. O. Mitochondrial cytochrome oxidase produces nitric oxide under hypoxic conditions: implications for oxygen sensing and hypoxic signaling in eukaryotes. *Cell metabolism* 3, 277-287 (2006).
205 Kadosh, D. & Johnson, A. D. Rfg1, a protein related to the Saccharomyces cerevisiae hypoxic regulator Rox1, controls filamentous growth and virulence in Candida albicans. *Molecular and cellular biology* 21, 2496-2505 (2001).
206 Khalaf, R. A. & Zitomer, R. S. The DNA binding protein Rfg1 is a repressor of filamentation in Candida albicans. *Genetics* 157, 1503-1512 (2001).
207 Brown, D. H., Jr., Giusani, A. D., Chen, X. & Kumamoto, C. A. Filamentous growth of Candida albicans in response to physical environmental cues and its regulation by the unique CZF1 gene. *Molecular microbiology* 34, 651-662 (1999).
208 Sonneborn, A., Bockmuhl, D. P. & Ernst, J. F. Chlamydospore formation in Candida albicans requires the Efg1p morphogenetic regulator. *Infection and immunity* 67, 5514-5517 (1999).
209 Cao, F. *et al.* The Flo8 transcription factor is essential for hyphal development and virulence in Candida albicans. *Molecular biology of the cell* 17, 295-307 (2006).
210 Giusani, A. D., Vinces, M. & Kumamoto, C. A. Invasive filamentous growth of Candida albicans is promoted by Czf1p-dependent relief of Efg1p-mediated repression. *Genetics* 160, 1749-1753 (2002).
211 Mulhern, S. M., Logue, M. E. & Butler, G. Candida albicans transcription factor Ace2 regulates metabolism and is required for filamentation in hypoxic conditions. *Eukaryotic cell* 5, 2001-2013 (2006).
212 Stichternoth, C. & Ernst, J. F. Hypoxic adaptation by Efg1 regulates biofilm formation by Candida albicans. *Applied and environmental microbiology* 75, 3663-3672 (2009).
213 MacPherson, S. *et al.* Candida albicans zinc cluster protein Upc2p confers resistance to antifungal drugs and is an activator of ergosterol biosynthetic genes. *Antimicrobial agents and chemotherapy* 49, 1745-1752 (2005).
214 Scandalios, J. G. Oxidative stress: molecular perception and transduction of signals triggering antioxidant gene defenses. *Brazilian journal of medical and biological research* 38, 995-1014 (2005).
215 Fridovich, I. Superoxide radical and superoxide dismutases. *Annual review of biochemistry* 64, 97-112 (1995).
216 Reeves, E. P. *et al.* Killing activity of neutrophils is mediated through activation of proteases by K+ flux. *Nature* 416, 291-297 (2002).
217 Fang, F. C. Antimicrobial reactive oxygen and nitrogen species: concepts and controversies. *Nature reviews. Microbiology* 2, 820-832 (2004).

218 Frohner, I. E., Bourgeois, C., Yatsyk, K., Majer, O. & Kuchler, K. Candida albicans cell surface superoxide dismutases degrade host-derived reactive oxygen species to escape innate immune surveillance. *Molecular microbiology* 71, 240-252 (2009).
219 Ferrari, C. K., Souto, P. C., Franca, E. L. & Honorio-Franca, A. C. Oxidative and nitrosative stress on phagocytes' function: from effective defense to immunity evasion mechanisms. *Archivum immunologiae et therapiae experimentalis* 59, 441-448 (2011).
220 Garcera, A., Casas, C. & Herrero, E. Expression of Candida albicans glutathione transferases is induced inside phagocytes and upon diverse environmental stresses. *FEMS yeast research* 10, 422-431 (2010).
221 Kaloriti, D. *et al.* Combinatorial stresses kill pathogenic Candida species. *Medical mycology* 50, 699-709 (2012).
222 Enjalbert, B., MacCallum, D. M., Odds, F. C. & Brown, A. J. Niche-specific activation of the oxidative stress response by the pathogenic fungus Candida albicans. *Infection and immunity* 75, 2143-2151 (2007).
223 Chauhan, N., Latge, J. P. & Calderone, R. Signalling and oxidant adaptation in Candida albicans and Aspergillus fumigatus. *Nature reviews. Microbiology* 4, 435-444 (2006).
224 Westwater, C., Balish, E. & Schofield, D. A. Candida albicans-conditioned medium protects yeast cells from oxidative stress: a possible link between quorum sensing and oxidative stress resistance. *Eukaryotic cell* 4, 1654-1661 (2005).
225 Abegg, M. A. *et al.* Response to oxidative stress in eight pathogenic yeast species of the genus Candida. *Mycopathologia* 170, 11-20 (2010).
226 Hwang, C. S. *et al.* Copper- and zinc-containing superoxide dismutase (Cu/ZnSOD) is required for the protection of Candida albicans against oxidative stresses and the expression of its full virulence. *Microbiology* 148, 3705-3713 (2002).
227 Martchenko, M., Alarco, A. M., Harcus, D. & Whiteway, M. Superoxide dismutases in Candida albicans: transcriptional regulation and functional characterization of the hyphal-induced SOD5 gene. *Molecular biology of the cell* 15, 456-467 (2004).
228 Fradin, C. *et al.* Granulocytes govern the transcriptional response, morphology and proliferation of Candida albicans in human blood. *Molecular microbiology* 56, 397-415 (2005).
229 Wysong, D. R., Christin, L., Sugar, A. M., Robbins, P. W. & Diamond, R. D. Cloning and sequencing of a Candida albicans catalase gene and effects of disruption of this gene. *Infection and immunity* 66, 1953-1961 (1998).
230 Nakagawa, Y., Kanbe, T. & Mizuguchi, I. Disruption of the human pathogenic yeast Candida albicans catalase gene decreases survival in mouse-model infection and elevates susceptibility to higher temperature and to detergents. *Microbiology and immunology* 47, 395-403 (2003).
231 Ramon, A. M. & Fonzi, W. A. Diverged binding specificity of Rim101p, the Candida albicans ortholog of PacC. *Eukaryotic cell* 2, 718-728 (2003).
232 Srinivasa, K. *et al.* Characterization of a putative thioredoxin peroxidase prx1 of Candida albicans. *Molecules and cells* 33, 301-307 (2012).
233 Urban, C., Sohn, K., Lottspeich, F., Brunner, H. & Rupp, S. Identification of cell surface determinants in Candida albicans reveals Tsa1p, a protein differentially localized in the cell. *FEBS letters* 544, 228-235 (2003).
234 Shin, D. H. *et al.* Characterization of thiol-specific antioxidant 1 (TSA1) of Candida albicans. *Yeast* 22, 907-918 (2005).
235 Chaves, G. M. & da Silva, W. P. Superoxide dismutases and glutaredoxins have a distinct role in the response of Candida albicans to oxidative stress generated by the chemical

compounds menadione and diamide. *Memorias do Instituto Oswaldo Cruz* 107, 998-1005 (2012).

236 Chaves, G. M., Bates, S., Maccallum, D. M. & Odds, F. C. Candida albicans GRX2, encoding a putative glutaredoxin, is required for virulence in a murine model. *Genetics and molecular research* 6, 1051-1063 (2007).

237 Sen, C. K. Redox signaling and the emerging therapeutic potential of thiol antioxidants. *Biochemical pharmacology* 55, 1747-1758 (1998).

238 da Silva Dantas, A. *et al.* Thioredoxin regulates multiple hydrogen peroxide-induced signaling pathways in Candida albicans. *Molecular and cellular biology* 30, 4550-4563 (2010).

239 Mayra Cuéllar-Cruz, G. G.-S., Everardo López-Romero, Estela Ruiz-Baca, Julio C. Villagómez-Castro, Lucio Rodríguez-Sifuentes. Identification of Candida albicans heat shock proteins and Candida glabrata and Candida krusei enolases involved in the response to oxidative stress. *Central European Journal of Biology* 8, 337-345 (2013).

240 Jong, A. Y. *et al.* Binding of Candida albicans enolase to plasmin(ogen) results in enhanced invasion of human brain microvascular endothelial cells. *Journal of medical microbiology* 52, 615-622 (2003).

241 Lain, A. *et al.* Use of recombinant antigens for the diagnosis of invasive candidiasis. *Clinical & developmental immunology* 2008, 721950 (2008).

242 Fernandez-Arenas, E. *et al.* Integrated proteomics and genomics strategies bring new insight into Candida albicans response upon macrophage interaction. *Molecular & cellular proteomics : MCP* 6, 460-478 (2007).

243 Kultz, D. & Burg, M. Evolution of osmotic stress signaling via MAP kinase cascades. *The Journal of experimental biology* 201, 3015-3021 (1998).

244 Banuett, F. Signalling in the yeasts: an informational cascade with links to the filamentous fungi. *Microbiology and molecular biology reviews* 62, 249-274 (1998).

245 Monge, R. A., Roman, E., Nombela, C. & Pla, J. The MAP kinase signal transduction network in Candida albicans. *Microbiology* 152, 905-912 (2006).

246 Alonso-Monge, R. *et al.* The Hog1 mitogen-activated protein kinase is essential in the oxidative stress response and chlamydospore formation in Candida albicans. *Eukaryotic cell* 2, 351-361 (2003).

247 Navarro-Garcia, F., Eisman, B., Fiuza, S. M., Nombela, C. & Pla, J. The MAP kinase Mkc1p is activated under different stress conditions in Candida albicans. *Microbiology* 151, 2737-2749 (2005).

248 de Dios, C. H., Roman, E., Monge, R. A. & Pla, J. The role of MAPK signal transduction pathways in the response to oxidative stress in the fungal pathogen Candida albicans: implications in virulence. *Current protein & peptide science* 11, 693-703 (2010).

249 Chauhan, N. *et al.* Candida albicans response regulator gene SSK1 regulates a subset of genes whose functions are associated with cell wall biosynthesis and adaptation to oxidative stress. *Eukaryotic cell* 2, 1018-1024 (2003).

250 Smith, D. A., Nicholls, S., Morgan, B. A., Brown, A. J. & Quinn, J. A conserved stress-activated protein kinase regulates a core stress response in the human pathogen Candida albicans. *Molecular biology of the cell* 15, 4179-4190 (2004).

251 Kruppa, M. & Calderone, R. Two-component signal transduction in human fungal pathogens. *FEMS yeast research* 6, 149-159 (2006).

252 Smith, D. A., Morgan, B. A. & Quinn, J. Stress signalling to fungal stress-activated protein kinase pathways. *FEMS microbiology letters* 306, 1-8 (2010).

253 Calera, J. A., Choi, G. H. & Calderone, R. A. Identification of a putative histidine kinase two-component phosphorelay gene (CaHK1) in Candida albicans. *Yeast* 14, 665-674 (1998).
254 Calera, J. A. & Calderone, R. A. Identification of a putative response regulator two-component phosphorelay gene (CaSSK1) from Candida albicans. *Yeast* 15, 1243-1254 (1999).
255 Calera, J. A., Herman, D. & Calderone, R. Identification of YPD1, a gene of Candida albicans which encodes a two-component phosphohistidine intermediate protein. *Yeast* 16, 1053-1059 (2000).
256 Cheetham, J. *et al.* A single MAPKKK regulates the Hog1 MAPK pathway in the pathogenic fungus Candida albicans. *Molecular biology of the cell* 18, 4603-4614 (2007).
257 Arana, D. M., Nombela, C., Alonso-Monge, R. & Pla, J. The Pbs2 MAP kinase kinase is essential for the oxidative-stress response in the fungal pathogen Candida albicans. *Microbiology* 151, 1033-1049 (2005).
258 San Jose, C., Monge, R. A., Perez-Diaz, R., Pla, J. & Nombela, C. The mitogen-activated protein kinase homolog HOG1 gene controls glycerol accumulation in the pathogenic fungus Candida albicans. *Journal of bacteriology* 178, 5850-5852 (1996).
259 Alarco, A. M. & Raymond, M. The bZip transcription factor Cap1p is involved in multidrug resistance and oxidative stress response in Candida albicans. *Journal of bacteriology* 181, 700-708 (1999).
260 Zhang, X., De Micheli, M., Coleman, S. T., Sanglard, D. & Moye-Rowley, W. S. Analysis of the oxidative stress regulation of the Candida albicans transcription factor, Cap1p. *Molecular microbiology* 36, 618-629 (2000).
261 Znaidi, S. *et al.* Identification of the Candida albicans Cap1p regulon. *Eukaryotic cell* 8, 806-820 (2009).
262 Singh, P., Chauhan, N., Ghosh, A., Dixon, F. & Calderone, R. SKN7 of Candida albicans: mutant construction and phenotype analysis. *Infection and immunity* 72, 2390-2394 (2004).
263 Bruce, C. R. *et al.* Identification of a novel response regulator, Crr1, that is required for hydrogen peroxide resistance in Candida albicans. *PloS one* 6, e27979 (2011).
264 de Sousa Abreu, R., Penalva, L. O., Marcotte, E. M. & Vogel, C. Global signatures of protein and mRNA expression levels. *Molecular bioSystems* 5, 1512-1526 (2009).
265 Schwanhausser, B. *et al.* Global quantification of mammalian gene expression control. *Nature* 473, 337-342 (2011).
266 Maier, T., Guell, M. & Serrano, L. Correlation of mRNA and protein in complex biological samples. *FEBS letters* 583, 3966-3973 (2009).
267 Li, G. W., Burkhardt, D., Gross, C. & Weissman, J. S. Quantifying absolute protein synthesis rates reveals principles underlying allocation of cellular resources. *Cell* 157, 624-635 (2014).
268 Kafri, M., Metzl-Raz, E., Jona, G. & Barkai, N. The Cost of Protein Production. *Cell reports* 14, 22-31 (2016).
269 Dennis, P. P. In vivo stability, maturation and relative differential synthesis rates of individual ribosomal proteins in Escherichia coli B/r. *Journal of molecular biology* 88, 25-41 (1974).
270 Lemaux, P. G., Herendeen, S. L., Bloch, P. L. & Neidhardt, F. C. Transient rates of synthesis of individual polypeptides in E. coli following temperature shifts. *Cell* 13, 427-434 (1978).
271 Schwanhausser, B., Gossen, M., Dittmar, G. & Selbach, M. Global analysis of cellular protein translation by pulsed SILAC. *Proteomics* 9, 205-209 (2009).
272 Barbara Deracinois, C. F., Sophie Duban-Deweer, and Yannis Karamanos. Comparative and Quantitative Global Proteomics Approaches: An Overview. *Proteomes* 1, 180-218 (2013).

273 Klose, J. & Kobalz, U. Two-dimensional electrophoresis of proteins: an updated protocol and implications for a functional analysis of the genome. *Electrophoresis* 16, 1034-1059 (1995).
274 O'Farrell, P. H. High resolution two-dimensional electrophoresis of proteins. *The Journal of biological chemistry* 250, 4007-4021 (1975).
275 Gevaert, K. & Vandekerckhove, J. Protein identification methods in proteomics. *Electrophoresis* 21, 1145-1154 (2000).
276 Apweiler, R., Bairoch, A. & Wu, C. H. Protein sequence databases. *Current opinion in chemical biology* 8, 76-80 (2004).
277 Chasse, H., Boulben, S., Costache, V., Cormier, P. & Morales, J. Analysis of translation using polysome profiling. *Nucleic acids research* 45, e15 (2017).
278 Melamed, D., Pnueli, L. & Arava, Y. Yeast translational response to high salinity: global analysis reveals regulation at multiple levels. *Rna* 14, 1337-1351 (2008).
279 Esposito, A. M. *et al.* Eukaryotic polyribosome profile analysis. *Journal of visualized experiments* (2010).
280 Shenton, D. *et al.* Global translational responses to oxidative stress impact upon multiple levels of protein synthesis. *The Journal of biological chemistry* 281, 29011-29021 (2006).
281 Rajasekhar, V. K. *et al.* Oncogenic Ras and Akt signaling contribute to glioblastoma formation by differential recruitment of existing mRNAs to polysomes. *Molecular cell* 12, 889-901 (2003).
282 Kawaguchi, R., Girke, T., Bray, E. A. & Bailey-Serres, J. Differential mRNA translation contributes to gene regulation under non-stress and dehydration stress conditions in Arabidopsis thaliana. *The Plant journal : for cell and molecular biology* 38, 823-839 (2004).
283 MacKay, V. L. *et al.* Gene expression analyzed by high-resolution state array analysis and quantitative proteomics: response of yeast to mating pheromone. *Molecular & cellular proteomics : MCP* 3, 478-489 (2004).
284 Lindemann, S. W. *et al.* Neutrophils alter the inflammatory milieu by signal-dependent translation of constitutive messenger RNAs. *Proceedings of the National Academy of Sciences of the United States of America* 101, 7076-7081 (2004).
285 Huber, M. *et al.* Comparison of proteomic and genomic analyses of the human breast cancer cell line T47D and the antiestrogen-resistant derivative T47D-r. *Molecular & cellular proteomics : MCP* 3, 43-55 (2004).
286 Bonifacino, J. S. Metabolic labeling with amino acids. *Current protocols in protein science* Chapter 3, Unit 3 7 (2001).
287 Esposito, A. M. & Kinzy, T. G. In vivo [35S]-methionine incorporation. *Methods in enzymology* 536, 55-64 (2014).
288 Martinez-Pastor, M. T. & Estruch, F. Sudden depletion of carbon source blocks translation, but not transcription, in the yeast Saccharomyces cerevisiae. *FEBS letters* 390, 319-322 (1996).
289 Jordán-Pla A., M. A., Serna E., Pelechano V., Pérez-Ortín J.E. in *Methods in molecular biology* (ed Frederic Devaux) (2015).
290 Alberola, T. M. *et al.* A new set of DNA macrochips for the yeast Saccharomyces cerevisiae: features and uses. *International microbiology : the official journal of the Spanish Society for Microbiology* 7, 199-206 (2004).
291 Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nature methods* 9, 357-359 (2012).
292 Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078-2079 (2009).

293 Anders, S., Pyl, P. T. & Huber, W. HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics* 31, 166-169 (2015).
294 Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L. & Wold, B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature methods* 5, 621-628 (2008).
295 Shen, L., Shao, N., Liu, X. & Nestler, E. ngs.plot: Quick mining and visualization of next-generation sequencing data by integrating genomic databases. *BMC genomics* 15, 284 (2014).
296 Russell, J. B. & Cook, G. M. Energetics of bacterial growth: balance of anabolic and catabolic reactions. *Microbiological reviews* 59, 48-62 (1995).
297 Schimmel, P. GTP hydrolysis in protein synthesis: two for Tu? *Science* 259, 1264-1265 (1993).
298 Liu, Y. & Aebersold, R. The interdependence of transcript and protein abundance: new data--new complexities. *Molecular systems biology* 12, 856 (2016).
299 Hauf, J., Zimmermann, F. K. & Muller, S. Simultaneous genomic overexpression of seven glycolytic enzymes in the yeast Saccharomyces cerevisiae. *Enzyme and microbial technology* 26, 688-698 (2000).
300 Lang, G. I., Murray, A. W. & Botstein, D. The cost of gene expression underlies a fitness trade-off in yeast. *Proceedings of the National Academy of Sciences of the United States of America* 106, 5755-5760 (2009).
301 MacLean, R. C. Pleiotropy and GAL pathway degeneration in yeast. *Journal of evolutionary biology* 20, 1333-1338 (2007).
302 Garcia-Martinez, J., Gonzalez-Candelas, F. & Perez-Ortin, J. E. Common gene expression strategies revealed by genome-wide analysis in yeast. *Genome biology* 8, R222 (2007).
303 Vogel, C., Silva, G. M. & Marcotte, E. M. Protein expression regulation under oxidative stress. *Molecular & cellular proteomics* 10, M111 009217 (2011).
304 de Nobel, H. *et al.* Parallel and comparative analysis of the proteome and transcriptome of sorbic acid-stressed Saccharomyces cerevisiae. *Yeast* 18, 1413-1428 (2001).
305 Lackner, D. H., Schmidt, M. W., Wu, S., Wolf, D. A. & Bahler, J. Regulation of transcriptome, translation, and proteome in response to environmental stress in fission yeast. *Genome biology* 13, R25 (2012).
306 Cuéllar-Cruz, M., , G. G.-S., López-Romero E., Ruiz-Baca E., Villagómez-Castro J.C., Rodríguez-Sifuentes L. . Identification of Candida albicans heat shock proteins and Candida glabrata and Candida krusei enolases involved in the response to oxidative stress. *Central European Journal of Biology* 8, 337-345 (2013).
307 Kusch, H., Engelmann, S., Albrecht, D., Morschhauser, J. & Hecker, M. Proteomic analysis of the oxidative stress response in Candida albicans. *Proteomics* 7, 686-697 (2007).
308 Vogel, C. & Marcotte, E. M. Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nature reviews. Genetics* 13, 227-232 (2012).
309 Fournier, M. L. *et al.* Delayed correlation of mRNA and protein expression in rapamycin-treated cells and a role for Ggc1 in cellular sensitivity to rapamycin. *Molecular & cellular proteomics : MCP* 9, 271-284 (2010).
310 Maier, T. *et al.* Quantification of mRNA and protein and integration with protein turnover in a bacterium. *Molecular systems biology* 7, 511 (2011).
311 Jayapal, K. P. *et al.* Uncovering genes with divergent mRNA-protein dynamics in Streptomyces coelicolor. *PloS one* 3, e2097 (2008).
312 Thorvaldsdottir, H., Robinson, J. T. & Mesirov, J. P. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Briefings in bioinformatics* 14, 178-192 (2013).

313 Nicol, J. W., Helt, G. A., Blanchard, S. G., Jr., Raja, A. & Loraine, A. E. The Integrated Genome Browser: free software for distribution and exploration of genome-scale datasets. *Bioinformatics* 25, 2730-2731 (2009).
314 Suarez, E., Burguete, A. & McLachlan, G. J. Microarray data analysis for differential expression: a tutorial. *Puerto Rico health sciences journal* 28, 89-104 (2009).
315 Rapaport, F. *et al.* Comprehensive evaluation of differential gene expression analysis methods for RNA-seq data. *Genome biology* 14, R95 (2013).
316 Medina, I. *et al.* Babelomics: an integrative platform for the analysis of transcriptomics, proteomics and genomic data with advanced functional profiling. *Nucleic acids research* 38, W210-213 (2010).
317 Juanes, J. M., Miguel, A., Morales, L. J., Perez-Ortin, J. E. & Arnau, V. A web application for the unspecific detection of differentially expressed DNA regions in strand-specific expression data. *Bioinformatics* 31, 3228-3230 (2015).
318 M. Basseville, I. V. N. *Detection of abrupt changes: theory and application*. Vol. 15 28-32 (Prentice Hall Englewood Clifss, 1993).
319 Nadal-Ribelles, M. *et al.* Control of Cdc28 CDK1 by a stress-induced lncRNA. *Molecular cell* 53, 549-561 (2014).
320 Geisler, S., Lojek, L., Khalil, A. M., Baker, K. E. & Coller, J. Decapping of long noncoding RNAs regulates inducible genes. *Molecular cell* 45, 279-291 (2012).
321 Swamy, K. B., Lin, C. H., Yen, M. R., Wang, C. Y. & Wang, D. Examining the condition-specific antisense transcription in S. cerevisiae and S. paradoxus. *BMC genomics* 15, 521 (2014).
322 Leong, H. S. *et al.* A global non-coding RNA system modulates fission yeast protein levels in response to stress. *Nature communications* 5, 3947 (2014).
323 Bruno, V. M. *et al.* Comprehensive annotation of the transcriptome of the human fungal pathogen Candida albicans using RNA-seq. *Genome research* 20, 1451-1458 (2010).
324 Enjalbert, B., Nantel, A. & Whiteway, M. Stress-induced gene expression in Candida albicans: absence of a general stress response. *Molecular biology of the cell* 14, 1460-1467 (2003).
325 Gasch, A. P. *et al.* Genomic expression programs in the response of yeast cells to environmental changes. *Molecular biology of the cell* 11, 4241-4257 (2000).
326 Morris, K. V. & Mattick, J. S. The rise of regulatory RNA. *Nature reviews. Genetics* 15, 423-437 (2014).
327 Koziol, M. J. & Rinn, J. L. RNA traffic control of chromatin complexes. *Current opinion in genetics & development* 20, 142-148 (2010).
328 Mercer, T. R. & Mattick, J. S. Structure and function of long noncoding RNAs in epigenetic regulation. *Nature structural & molecular biology* 20, 300-307 (2013).
329 Tudek, A., Candelli, T. & Libri, D. Non-coding transcription by RNA polymerase II in yeast: Hasard or necessite? *Biochimie* 117, 28-36 (2015).
330 Gascoigne, D. K. *et al.* Pinstripe: a suite of programs for integrating transcriptomic and proteomic datasets identifies novel proteins and improves differentiation of protein-coding and non-coding genes. *Bioinformatics* 28, 3042-3050 (2012).
331 Guttman, M., Russell, P., Ingolia, N. T., Weissman, J. S. & Lander, E. S. Ribosome profiling provides evidence that large noncoding RNAs do not encode proteins. *Cell* 154, 240-251 (2013).
332 Rinn, J. & Guttman, M. RNA Function. RNA and dynamic nuclear organization. *Science* 345, 1240-1241 (2014).
333 Rinn, J. L. & Chang, H. Y. Genome regulation by long noncoding RNAs. *Annual review of biochemistry* 81, 145-166 (2012).

334 Xu, Z. *et al.* Antisense expression increases gene expression variability and locus interdependency. *Molecular systems biology* 7, 468 (2011).
335 Garcia-Martinez, J. *et al.* The relative importance of transcription rate, cryptic transcription and mRNA stability on shaping stress responses in yeast. *Transcription* 3, 39-44 (2012).
336 Gardini, A. Global Run-On Sequencing (GRO-Seq). *Methods in molecular biology* 1468, 111-120 (2017).
337 Kharchenko, P. V. *et al.* Comprehensive analysis of the chromatin landscape in Drosophila melanogaster. *Nature* 471, 480-485 (2011).
338 Hetzel, J., Duttke, S. H., Benner, C. & Chory, J. Nascent RNA sequencing reveals distinct features in plant transcription. *Proceedings of the National Academy of Sciences of the United States of America* 113, 12316-12321 (2016).
339 Szentirmay, M. N. & Sawadogo, M. Sarkosyl block of transcription reinitiation by RNA polymerase II as visualized by the colliding polymerases reinitiation assay. *Nucleic acids research* 22, 5341-5346 (1994).
340 Kovelman, R. & Roeder, R. G. Sarkosyl defines three intermediate steps in transcription initiation by RNA polymerase III: application to stimulation of transcription by E1A. *Genes & development* 4, 646-658 (1990).
341 Kwak, H., Fuda, N. J., Core, L. J. & Lis, J. T. Precise maps of RNA polymerase reveal how promoters direct initiation and pausing. *Science* 339, 950-953 (2013).
342 Dangkulwanich, M., Ishibashi, T., Bintu, L. & Bustamante, C. Molecular mechanisms of transcription through single-molecule experiments. *Chemical reviews* 114, 3203-3223 (2014).
343 Kireeva, M. L. *et al.* Nature of the nucleosomal barrier to RNA polymerase II. *Molecular cell* 18, 97-108 (2005).
344 Cottier, F. *et al.* The transcriptional stress response of Candida albicans to weak organic acids. *G3* 5, 497-505 (2015).
345 Mason, P. B. & Struhl, K. Distinction and relationship between elongation rate and processivity of RNA polymerase II in vivo. *Molecular cell* 17, 831-840 (2005).
346 Jonkers, I. & Lis, J. T. Getting up to speed with transcription elongation by RNA polymerase II. *Nature reviews. Molecular cell biology* 16, 167-177 (2015).
347 Barranco, F. M. Estudio de los cambios de velocidad de la RNA polimerasa II de *Saccharomyces cerevisiae* en función de la temperatura (2012).
348 Connors, K. A. *Chemical Kinetics: The Study of Reaction Rates in Solution* (VCH Publishers, 1998).
349 Li, Z. *et al.* Systematic exploration of essential yeast gene function with temperature-sensitive mutants. *Nature biotechnology* 29, 361-367, doi:10.1038/nbt.1832 (2011).
350 Exinger, F. & Lacroute, F. 6-Azauracil inhibition of GTP biosynthesis in Saccharomyces cerevisiae. *Current genetics* 22, 9-11 (1992).
351 Shaw, R. J. & Reines, D. Saccharomyces cerevisiae transcription elongation mutants are defective in PUR5 induction in response to nucleotide depletion. *Molecular and cellular biology* 20, 7427-7437 (2000).
352 José E. Pérez-Ortín, D. A. M., and Antonio Jordán-Pla. Genomic Insights into the Different Layers of Gene Regulation in Yeast. *Genetics Research International* 2011, 12 pages (2011).
353 Pérez-Ortín, J., Medina D, Chávez S, Moreno J. What do you mean by transcription rate? *BioEssays* 35, 1056–1062 (2013).

354 Warner, J. R. The economics of ribosome biosynthesis in yeast. *Trends in biochemical sciences* 24, 437-440 (1999).

355 Perez-Ortin, J. E., Alepuz, P., Chavez, S. & Choder, M. Eukaryotic mRNA decay: methodologies, pathways, and links to other stages of gene expression. *Journal of molecular biology* 425, 3750-3775 (2013).

356 Salas, D. A. M. Análisis genómico de la interacción entre la transcripción y la degradación durante el recambio del mRNA en *S. cerevisiae* (2015).

357 Sun, M. *et al.* Global analysis of eukaryotic mRNA degradation reveals Xrn1-dependent buffering of transcript levels. *Molecular cell* 52, 52-62 (2013).

358 von der Haar, T. A quantitative estimation of the global translational activity in logarithmically growing yeast cells. *BMC systems biology* 2, 87 (2008).

359 Ratkowsky, D. A., Olley, J. & Ross, T. Unifying temperature effects on the growth rate of bacteria and the stability of globular proteins. *Journal of theoretical biology* 233, 351-362 (2005).

360 Farewell, A. & Neidhardt, F. C. Effect of temperature on in vivo protein synthetic capacity in Escherichia coli. *Journal of bacteriology* 180, 4704-4710 (1998).

361 Waldron, C., Jund, R. & Lacroute, F. Evidence for a high proportion of inactive ribosomes in slow-growing yeast cells. *The Biochemical journal* 168, 409-415 (1977).

362 Petersen, N. S., McLaughlin, C. S. & Nierlich, D. P. Half life of yeast messenger RNA. *Nature* 260, 70-72 (1976).

363 Chia, L. L. & McLaughlin, C. The half-life of mRNA in Saccharomyces cerevisiae. *Molecular & general genetics* 170, 137-144 (1979).

364 García-Martínez J, D.-R. L., Ayala G, Pelechano V, Medina DA, Carrasco F, González R, Andrés-León E, Steinmetz L, Warringer J, Chávez S, Pérez-Ortín JE. The cellular growth rate controls overall mRNA turnover, and modulates either transcription or degradation rates of particular gene regulons. *Nucleic Acids Research* Dec 29 (2015).

365 Caspeta, L. *et al.* Biofuels. Altered sterol composition renders yeast thermotolerant. *Science* 346, 75-78 (2014).