

Multimic investigation on the relationship between metabolism and gene expression



Shidong Xi

Wadham College

Hilary Term 2019

A thesis submitted to the University of Oxford for the degree of Doctor of Philosophy
Word count: 49900

Abstract

Cellular metabolism and gene expression are proposed to be highly integrated but a detailed mechanistic understanding of this relationship is lacking. The central metabolite, acetyl-CoA, is the substrate for histone acetylation and increases in levels of acetyl-CoA are associated with increased histone acetylation and active gene expression. How histone acetylation leads to upregulated gene expression and how the altered patterns of gene expression subsequently influence cellular metabolism is not fully understood. This work addresses the 'acetylation-recruitment' hypothesis: that the upregulated histone acetylation in response to the increased acetyl-CoA can recruit a bromodomain containing chromatin remodelling complex to activate gene expression, which can further regulate cellular metabolism. To facilitate this study, a growth phenomenon in budding yeast, *S.cerevisiae*, known as the YMC (yeast metabolic cycle) is exploited. In the YMC, yeast cells exhibit highly synchronized and periodic oscillations between the high and low oxygen consumption states, associated with cycling levels of acetyl-CoA, histone acetylation and transcripts. To link histone acetylation to cycling transcripts, the role of various bromodomain containing proteins in maintaining the YMC is tested and the Rsc2 isoform of the RSC complex is identified as important. By comparing the temporal profiles of transcripts between the wild-type and the *rsc2Δ* strain, it is found that the expression of genes having histone H3 acetylation at their promoters is significantly more likely to be dependent on the RSC complex. Furthermore, *ACO2* is identified and verified as a gene associated with the upregulation of oxygen consumption and also a downstream target of the acetylation-RSC-recruitment pathway. Finally, two new techniques have been developed with the aim of addressing the question about the relationship between transcription and transcripts during the YMC and how this relates to metabolism. One is ultra-low background NET-seq, which reduces the percentage of background signals in a yeast NET-seq library from 32% to 0.7%. The other is size-fractionated NET-seq, which can map transcription start sites for both coding and noncoding transcription units.

Acknowledgements

First of all, I would like to give special thanks to my supervisor, Professor Jane Mellor, for giving me the opportunity to join her lab in 2016 and for her encouragement, guidance and support throughout my DPhil studies. Jane creates an enjoyable and supportive research environment for everybody in the lab and also provides freedom for people in the lab to explore various research questions. Under her guidance, I have learnt how to think critically, how to manage time efficiently and how to carry out research independently.

Secondly, I would like to thank all my colleagues. I have a lot of constructive discussions with Jack Feltham. The idea of the size fractionated NET-seq was created during the discussion with him. I also want to thank Harry Fischl, Françoise Howe and Anitha Nair for teaching me all the essential lab techniques and Ronja Woloszczuk and Struan Murray for their support on the bioinformatics.

Abbreviations

Abbreviation	Term in full
2-HG	2-hydroxyglutarate
ac	Acetylation
acetyl-CoA	Acetyl coenzyme A
ACL	ATP-citrate lyase
ACO2	Aconitase 2
ACS1	Acetyl-coenzyme A synthetase 1
ACS2	Acetyl-coenzyme A synthetase 2
BAH	Bromo adjacent homology domain
BD	Bromodomain
bps	Base pairs
CDR	Complementarity determining region
ChIP-seq	Chromatin Immunoprecipitation Sequencing
CLN3	Cyclin 3
co-IP	coimmunoprecipitation
CT	c-terminal domain
CTD	c-terminal repeat domain
DO	Dissolved oxygen
FAD	Flavin adenosine dinucleotide
HAT	Histone acetyltransferases
HDAC	Histone deacetylases
HOC-1	High oxygen consumption phase-1
HOC-2	High oxygen consumption phase-2
HOC-3	High oxygen consumption phase-3
IDH1	Isocitrate dehydrogenase 1
IGV	Integrative genome viewer
IP	Immunoprecipitation
K	Lysine
KAT	Lysine acetyltransferases
K _d	Dissociation constant
KDAC	Lysine deacetylases
K _m	Michaelis constant
LOC	Low oxygen consumption phase
LSD1	Lysine specific demethylase 1
me	Methylation
MS	Mass spectrometry
NAD ⁺	Nicotinamide Adenine Dinucleotide
NAM	Nicotinamide
NET-seq	Native elongating transcript sequencing
NTH1	Neutral trehalase 1

Abbreviation	Term in full
nts	Nucleotides
OD ₆₀₀	Optical density at 600nm
PAS	Polyadenylation sites
PCR	Polymerase chain reaction
RNAPII	RNA polymerase II
RNA-seq	RNA sequencing
rRNA	Ribosomal RNA
RSC	Remodels the structure of chromatin
SAGA	Spt-Ada-Gcn5 acetyltransferase
SAH	S-adenosyl homocysteine
SAM	S-adenosylmethionine
snRNA	Small nuclear RNA
TCA cycle	Tricarboxylic acid cycle
TEF-seq	Transcription elongation factor sequencing
TIF-seq	Transcript isoform sequencing
tRNA	Transfer RNA
TSS	Transcription start sites
TW1	Time window 1
TW2	Time window 2
YMC	Yeast metabolic cycle
ub	Ubiquitination

Table of Contents

1	Introduction	2
1.1	There is an interplay between epigenetics and cellular metabolism	2
1.2	The scope of this study	8
1.3	The nucleocytosolic and mitochondrial pools of acetyl-CoA	8
1.4	An overview of protein acetylation	12
1.5	The functions of lysine acetylation on non-histone proteins.....	13
1.6	Histone (lysine) acetylation	14
1.6.1	The classification, specificity and metabolic regulation of histone acetyltransferases (HATs)	15
1.6.2	The classification, specificity and metabolic regulation of histone deacetylases (HDACs)	18
1.6.3	Bromodomains: histone acetylation reader	21
1.6.4	The cellular concentration of acetyl-CoA can fluctuate and it is often associated with changes in protein acetylation	22
1.6.5	Histone acetylation is associated with active gene expression	23
1.7	Chromatin remodelers.....	25
1.7.1	The three main functions of chromatin remodelers	25
1.7.2	The four families of chromatin remodelers	26
1.7.3	The chromatin remodelers in the SWI/SNF family contain one or more bromodomains.....	29
1.7.4	The SWI/SNF family and gene expression	30
1.8	An introduction to the yeast metabolic cycle.....	31
1.9	Using the YMC to study the relationship between transcription and transcripts	34
1.10	The two aims of this thesis.....	37
2	Materials and methods	39
2.1	Strain construction.....	39
2.2	Operating the fermenter	43
2.3	Next generation sequencing.....	45
2.3.1	Nascent elongating transcripts sequencing (conventional protocol)	45
2.3.2	Nascent elongating transcripts sequencing (ultra-low background protocol)	49
2.3.3	Size fractionated NET-seq.....	50
2.3.4	mRNA 3'end sequencing.....	54
2.4	Proteome.....	55

2.5	Biochemical assays	56
2.5.1	Western blot	56
2.5.2	Non-phenol genomic DNA extraction	57
2.6	Bioinformatics	57
2.6.1	The processing of the raw sequencing data	57
2.6.2	Pearson correlation, K-means clustering, statistical tests and plotting metagene	58
2.6.3	Mapping time points to the wild-type time axis	58
2.6.4	The dry lab part of size-fractionated NET-seq	59
2.6.5	Motif enrichment analysis.....	60
2.6.6	Gene ontology (GO) analysis	60
3	Investigating the role of the RSC complex during the YMC	62
3.1	Introduction	62
3.1.1	An introduction to the YMC system.....	62
3.1.1.1	The basics of the fermenter.....	62
3.1.1.2	The changes in the DO levels of the medium during a typical fermenter run	64
3.1.1.3	The DO level measured by the DO sensor can be converted to the cellular oxygen consumption rate by performing calibration.....	65
3.1.1.4	The typical oxygen consumption profile of the YMC	67
3.1.1.5	Cells in the LOC phase can be induced to enter the HOC phase	69
3.1.2	An introduction to the research question.....	70
3.1.2.1	The current understanding of the relationship between acetyl-CoA, histone acetylation and gene expression during the HOC phase of the YMC	70
3.1.2.2	The RSC complex.....	73
3.1.2.3	The structures and functions of Rsc1, Rsc2 and Rsc4	74
3.1.2.4	The hypothesized acetylation-recruitment pathway and plans for testing the hypothesis.....	76
3.2	Results	79
3.2.1	The <i>rsc2Δ</i> strain has a fluctuating oxygen consumption profile instead of the normal YMC, while no obvious defects can be seen in the <i>rsc1Δ</i> strain	79
3.2.2	The Rsc2 isoform of the RSC complex is required for the YMC.....	81
3.2.3	The loss of Rsc2 bromodomain 1 results in defects in the HOC-2 phase.....	83
3.2.4	The <i>rsc2-BD2Δ</i> strain cannot exhibit the YMC at the high dilution rate.....	84
3.2.5	The <i>rsc2-BAHA</i> strain has defects in the HOC-2 phase of the YMC as well ..	85
3.2.6	No obvious YMC defects can be seen in either <i>sth1-BDΔ</i> or <i>rsc4-Y225AY226A</i> 86	

3.2.7	A summary for the answer to the first question	87
3.2.8	RNA-seq samples are taken from the induced YMC-like response of the <i>rsc2Δ</i> strain and the normal YMC of the wild-type	88
3.2.9	The matrix of Pearson correlation coefficient maps the sampling points of the <i>rsc2Δ</i> strain to the time axis of the wild-type and also largely agrees with the phases of the YMC	90
3.2.10	There are two particular time windows in the HOC phase where the peaks of the transcripts of certain genes are not seen in the <i>rsc2Δ</i> strain	93
3.2.11	Genes losing their peaks of transcripts in TW1 in the <i>rsc2Δ</i> strain have a strong positive association with a peak of H3K9Ac, H3K14Ac, H3K18Ac or H4K5Ac at point B	97
3.2.12	The expression of CLN3 is Rsc2 dependent during the HOC-2 phase	101
3.2.13	A few genes have their expression patterns changed in the <i>rsc2Δ</i> strain; <i>ACO2</i> is most likely to be the downstream of the hypothesized acetylation-recruitment pathway	103
3.2.14	The expression of <i>ACO2</i> is sensitive to the deletion of genes required for histone H3 acetylation	111
3.2.15	The protein levels of Aco2p is increased after passing point B and returned to the ground level after point C	113
3.2.16	The <i>aco2Δ</i> strain is unable to further increase cellular oxygen consumption rates after passing point B	115
3.2.17	The defects seen in the <i>aco2Δ</i> strain can be partially rescued by adding excess ferrous ions	115
3.2.18	Chelating extracellular $\text{Fe}^{2+}/\text{Fe}^{3+}$ ions in the wild-type strain can mimic the effect of deleting <i>ACO2</i>	116
3.2.19	Excessive ferrous ions can rescue <i>rsc2-BD1Δ</i> and <i>rsc2-BAHA</i> as well	117
3.3	Discussion	120
3.3.1	The possible mechanisms of the RSC complex in bridging acetylation and gene expression	120
3.3.2	The RSC complex may activate genes in three different modes	124
3.3.3	The possible explanations for the RSC mutants showing defects in the YMC	126
3.3.4	The ‘Sozu’ model and the molecular mechanism behind the YMC	127
3.3.5	Other experiments that may help understand the mechanism of the YMC	134
3.3.5.1	Checking other bromodomain containing chromatin remodelling complexes	134
3.3.5.2	High throughput screening of mutants that show defects in the YMC ..	135
3.3.5.3	C^{13} metabolic flux analysis	138
3.3.5.4	Phosphoproteome and Acetylome	139

4	Technical development of yeast NET-seq	141
4.1	Ultra-low background NET-seq, a new protocol for yeast NET-seq	141
4.1.1	Introduction	141
4.1.1.1	The high background is a major technical limitation for yeast NET-seq.	141
4.1.1.2	The possible causes of the high background and the potential solutions	144
4.1.1.3	Urea can potentially be used in NET-seq to increase the stringency of the wash step	147
4.1.1.4	Aims and plans	148
4.1.2	Results	149
4.1.2.1	Both the FLAG-epitope-antibody interaction and RNAPII can tolerate the stringency of the new wash buffer	149
4.1.2.2	The new protocol can greatly reduce the background of NET-seq	150
4.1.2.3	The metagene profile of the library prepared by the new protocol is similar to that of the background subtracted library prepared by the conventional protocol	155
4.1.3	Discussion	156
4.1.3.1	The advantages and possible applications of the new protocol	156
4.1.3.2	The limitations and possible improvements of the new protocol	158
4.2	Size-fractionated NET-seq, a new method for mapping transcription start sites	159
4.2.1	Introduction	159
4.2.1.1	The current methods for mapping TSS	159
4.2.1.2	The limitation for the current TSS mapping methods and the potential solution	164
4.2.1.3	Aims and plans	166
4.2.2	Results	166
4.2.2.1	The workflow of size-fractionated NET-seq for annotating TSS	166
4.2.2.2	Size fractionated NET-seq can identify different type of TSS accurately, specifically and efficiently	169
4.2.2.3	The distributions of TSS around protein coding genes can be revealed by size fractionated NET-seq	176
4.2.2.4	In the region flanking PAS, an upstream noncoding antisense TSS and a downstream coding sense TSS have a strong tendency to co-occur	179
4.2.2.5	The DNA sequences of divergent promoters are significantly enriched for the motif of the Reb1 binding site	180
4.2.2.6	The divergent promoters are associated with Reb1 binding	182
4.2.2.7	Coding and noncoding TSS are associated with different genomic features	184

4.2.2.8	Coding and noncoding transcription units have a similar distribution of RNAPII in their promoter proximal regions	186
4.2.3	Discussion.....	188
4.2.3.1	The advantages of the size fractionated NET-seq.....	188
4.2.3.2	The limitations and future improvements.....	189
4.2.3.3	The possibility of using the same size selection method to annotate the mammalian genome	190
5	Discussion.....	193
5.1	Future plan	193
5.1.1	Developing precise run-on NET-seq/TEF-seq.....	193
5.2	Concluding remarks	196
Appendix		197
S1	At the global level, there is no obvious change in the distribution of RNA polymerase II on genes	197
S2	The levels of transcripts are dominantly regulated by transcription.....	198
S3	Genes with introns have a slightly longer time delay between transcription and transcripts than genes without introns	200
References		211

Chapter 1.

Introduction

1 Introduction

1.1 There is an interplay between epigenetics and cellular metabolism

Epigenetics is defined by the study of heritable phenotypes that are not associated with alterations in DNA sequences. Particularly, it mainly includes the investigation of the causes, (both temporal and spatial) patterns and consequences of DNA methylation and various chemical modifications on histone proteins. DNA methylation commonly occurs by adding a methyl group to an adenine or cytosine base. It normally causes the repression of gene expression and plays vital roles in normal embryonic development (Li et al. 1992; Okano et al. 1999; Siegfried et al. 1999). Altered DNA methylation patterns are associated with the development of cancer (Esteller 2007). Histone proteins have five major types, H1, H2A, H2B, H3 and H4. The latter 4 types are often referred to as the core histones. Two copies of each of the core histones comprise the globular structure of a histone octamer. Each histone octamer is wrapped with 147bps of DNA, which forms a nucleosome. Eukaryotic chromatin is often analogous to 'beads on a string' (Khorasanizadeh 2004). That is, the genomic DNA of eukaryotes is wrapped around numerous nucleosomes. Every two adjacent nucleosomes are connected with the linker DNA which is about 80bps long. The N-terminal tails of the core histones are flexible and unstructured. They often protrude from the nucleosomes and their amino acid residues are subject to various post-translational modifications, including but not limited to acetylation, methylation, phosphorylation, ubiquitination and sumoylation (Zhao & Garcia 2015). These histone modifications can modulate gene expression via two major ways (Kouzarides 2007). First, they may regulate the interactions between histones and DNA, which can make the chromatin structures become more permissive or restrictive for transcription. Second, they can regulate the recruitment of various chromatin associated factors which have transcriptional activation or repression effects. Among all the modifications, acetylation and methylation are the two most extensively studied ones. Generally, histone acetylation is

associated with the activation of gene expression, while the effects of histone methylation on gene expression are variable (Kaelin & McKnight 2013). The wide-spread transcriptional regulatory roles of histone modifications allows them to be implicated in the regulation of the majority of biological processes.

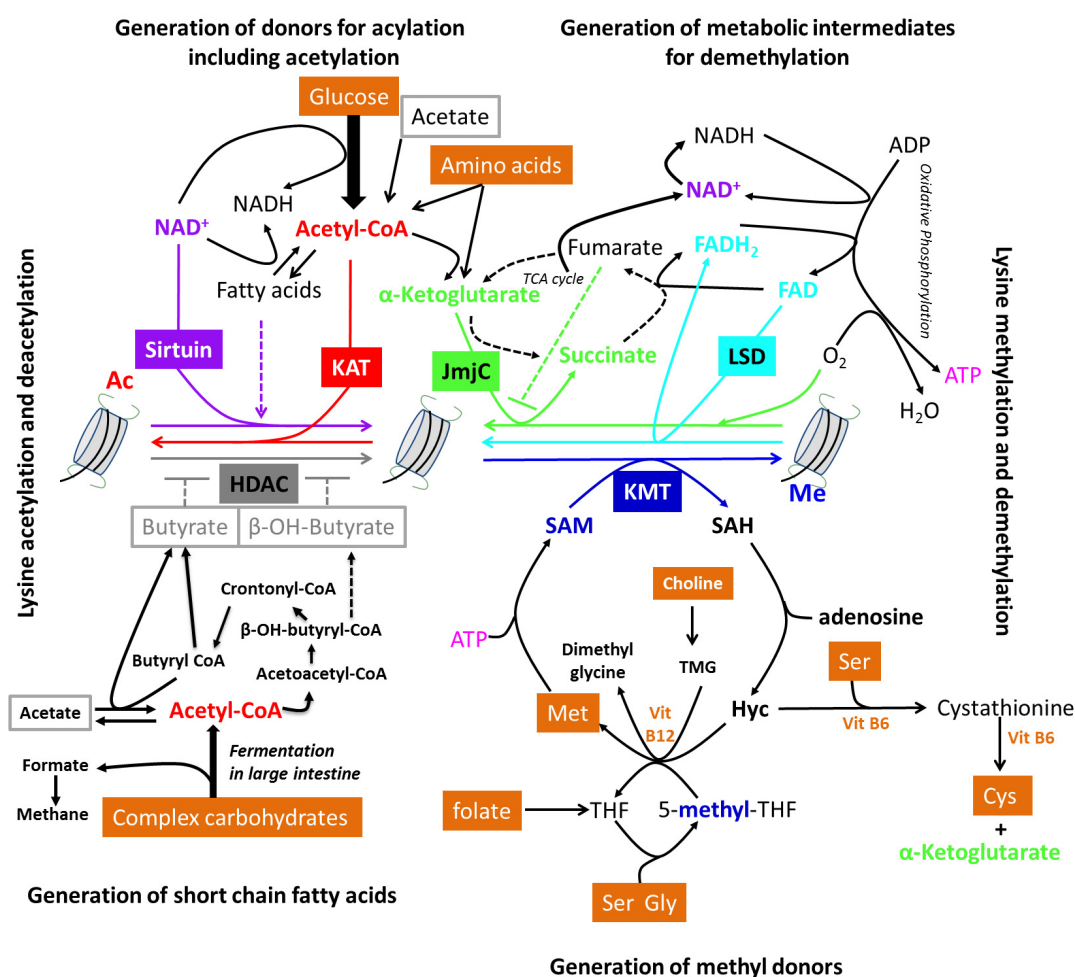


Figure 1.1 A schematic diagram showing the interplay between histone modifications and cellular metabolism. Metabolites can be either the chemical donors or inhibitors for histone acetylation and methylation as well as their reverse reactions. (Modified from Prof. Jane Mellor's lecture series)

Most of the changes in epigenetics involve the addition or removal of chemical groups on DNA or histone proteins. They are enzymatic reactions and require chemical donors or acceptors. In many cases, these chemical donors and acceptors are important cellular metabolites (Janke

et al. 2015). Here, it will be briefly exemplified by acetylation and methylation reactions (Figure 1.1). Histone acetylation is the addition of the acetyl group to histone lysine residues. It is catalysed by lysine acetyltransferases and uses acetyl-CoA (acetyl coenzyme A) as the chemical donor for the acetyl group (Menzies et al. 2016). Acetyl-CoA is a key metabolic intermediate participating in various pathways, such as the TCA cycle (tricarboxylic acid cycle) and the synthesis of fatty acids and steroids (Bravo-San Pedro et al. 2015). The reverse process, histone deacetylation, can be catalysed by lysine deacetylases. There is a class of histone deacetylases called Sirtuins. They can deacetylate histones in an NAD^+ (Nicotinamide Adenine Dinucleotide) dependent manner (Menzies et al. 2016). NAD^+ is a critical electron carrier in oxidative phosphorylation and also participates in other processes, such as gluconeogenesis (Cantó et al. 2015). Histone acetylation/deacetylation and their associations with metabolism as well as gene expression will be discussed in more details in the later sections of this chapter. The methylation of DNA and histones both requires SAM (S-adenosylmethionine) as the chemical donor of the methyl group and methyltransferases as the corresponding enzymes for the reaction (Suganuma & Workman 2018). After donating its methyl group, SAM is converted into SAH (S-adenosyl homocysteine), which can regenerate SAM via the one-carbon cycle. It is also referred to as the folate cycle. It can take various metabolites as inputs and synthesise nucleotides as well as amino acids (glycine, serine and methionine) (Mato et al. 2017). Hence, the cellular SAM level is implicated in a complex network of cellular metabolism. Multiple metabolites can be the chemical receptors of the demethylation reaction. The demethylation of H3K4me and H3K4me₂ can be performed by LSD1 (lysine specific demethylase 1) in a FAD (Flavin adenosine dinucleotide) dependent manner (Shi et al. 2004; Shi & Whetstone 2007). FAD also participates in a variety of redox reactions in the cell, such as the oxidation of succinate to fumarate during the TCA cycle and the beta-oxidation of fatty acids (Houten et al. 2016). Histone and DNA demethylation can also be carried out by alpha-ketoglutarate-dependent dioxygenases, which include JmJc (Jumonji C) domain-containing

histone demethylases and TET (ten-eleven translocation) proteins. Apart from alpha-ketoglutarate, oxygen and Fe(II) are also required (Dimitrova et al. 2015). During the demethylation reaction, oxygen and α -ketoglutarate are converted into CO₂ and succinate. Both α -ketoglutarate and succinate are critical players in the TCA cycle (Janke et al. 2015). To sum up, chromatin modifications are tightly related to cellular metabolites.

Metabolites can influence the activities of chromatin modifying enzymes in two major ways. First, they can act as the substrates and the changes in their cellular levels can directly regulate the enzymatic activities. The biochemical principle behind it is that the K_m (Michaelis constant) and K_d (dissociation constant) of many DNA and histone modifying enzymes fall within the normal physiological ranges of the concentrations of their corresponding substrates in the cell (Reid et al. 2017). For example, the K_m of various histone methyltransferases ranges from 1.2 μ M to 34.5 μ M, while the cellular concentrations of their substrate, SAM, normally fluctuate between 3.3 μ M and 59 μ M (Melnik et al. 2000; Oden & Clarke 1983; Lobet et al. 1989; Rathert et al. 2008; Tuck et al. 1985). Acetyl-CoA also influences the activity of histone acetyltransferases in a similar way. It will be demonstrated in details in later sections. For more examples, please refer to a recent review (Reid et al. 2017). Actually, metabolic enzymes can also form complexes with chromatin modifying enzymes to supply the substrates locally. For example, the NAD⁺ synthase, NMNAT-1, is associated with lysine deacetylase Sirt1 and promotes the deacetylation of H4K16Ac (Zhang et al. 2009). MAT (Methionine Adenosyltransferase) produces SAM from methionine. Experiments carried out by Katoh et al suggested it could be recruited to the target genes and promote histone methylation via interacting with transcription factor MafK (Katoh et al. 2011). It is worth to note that not all the enzymes in the cell can respond to the changes in its substrate levels. The typical examples are kinases involved in cell signalling pathways. The intracellular ATP levels are normally several magnitudes higher than the K_m for these enzymes (Reid et al. 2017). Hence, these kinase activities are not influenced by ATP levels. Second, some metabolites can regulate the

activities of chromatin modifying enzymes in a competitive or non-competitive manner. For example, NAM (Nicotinamide) can be a non-competitive inhibitor of histone deacetylase, Sir2 (Avalos et al. 2005). Fumarate and succinate can compete with α -ketoglutarate and inhibit the activities of many histone demethylases (Xiao et al. 2012). To conclude, the changes in cellular metabolites have the potential to regulate chromatin modifications.

Chromatin modifications can also be associated with the changes in the expression of metabolism related genes, which could be important for energy homeostasis. One example is the upregulation of glucose metabolism related genes via ACL (ATP-citrate lyase)-dependent histone acetylation after refeeding glucose containing serum. It was observed on immortalized mouse embryo fibroblasts by Wellen et al in 2009 (Wellen et al. 2009). Their results suggest that after refeeding, the increased glucose flux through glycolysis and the TCA cycle leads to the export of citrate from mitochondria. Citrate is further utilized to synthesize acetyl-CoA in the cytosol by ACL. The increased acetyl-CoA in the cytosolic pool further results in histone acetylation and the expression of a series of genes, including *GLUT4* (encoding a glucose transporter), three genes encoding enzymes for glycolysis, and genes for diverting glucose into amino acids. All of them can increase the consumption of glucose. The other example is the FAD-dependent LSD1-mediated repression of genes involved in active mitochondrial respiration (Hino et al. 2012). The low FAD levels normally indicate the limited cellular energy availability. Owing to the low level of FAD, the histone demethylation activity of LSD1 is reduced, which further leads to the increased histone methylation at the promoter of mitochondrial respiration related genes and the repression of these genes. Both examples show that there exists an interplay between epigenetic modifications and cellular metabolism.

The interplay between epigenetics and cellular metabolism is implicated in physiological and pathological processes. In terms of the physiological processes, one example is the regulation of cell differentiation (Moussaieff et al. 2015). Acetyl-CoA is accumulated via glycolysis, which

maintains histone acetylation and the pluripotency of embryonic stem cells (Shyh-Chang & Daley 2015). Upon cell differentiation, both the levels of acetyl-CoA and histone acetylation drop dramatically. Cell differentiation can be inhibited and promoted by modulating acetyl-CoA levels. Increasing acetyl-CoA by adding acetate can prevent histone deacetylation and cause a delay in cell differentiation. On the other hand, reducing acetyl-CoA by inhibiting its production can lead to histone deacetylation and cell differentiation (Barasch et al. 2015). The other example for the physiological process is the activation of T cells, which is proposed to be mediated by metabolically regulated epigenetic changes (Peng et al. 2016). T cell activation is associated with the metabolic switch from oxidative phosphorylation to aerobic glycolysis (Frauwirth et al. 2002). During this process, *LDHA* (lactate dehydrogenase A) is first expressed, which diverts pyruvate to lactate and elevates aerobic glycolysis. The data also suggest that the increased aerobic glycolysis promotes the export of citrate from mitochondria. ATP-citrate lyase then converts the cytosolic citrate to acetyl-CoA. It increases the cytosolic concentration of acetyl-CoA and further causes the upregulation of histone H3 acetylation, which can trigger the expression of genes for T cell activation including *IFN- γ* (Peng et al. 2016). In terms of the pathological processes, carcinogenesis is a typical example. *IDH1* (isocitrate dehydrogenase) and *IDH2* are enzymes catalysing the conversion between isocitrate and α -ketoglutarate during the TCA cycle. They are found to be mutated at a high frequency in multiple types of cancers (75% for grade 2-3 gliomas and 20% for AML)(Yang et al. 2012). The DNA and histone proteins in these cancers cells are often hypermethylated (Noushmehr et al. 2010). The point mutations on *IDH1* or *IDH2* can change their enzymatic activity leading to the production of 2-HG (2-hydroxyglutarate) (Wong et al. 2017). The normal cellular concentration of 2-HG is lower than 100 μ M (Losman & Kaelin 2013). It is raised to several mM by point mutations (Dang et al. 2009). 2-HG acts as a competitive inhibitor for several DNA and histone demethylases including JmjC domain-containing histone demethylases and TET proteins, which causes chromatin hypermethylation in these cancer cells bearing *IDH1/2* mutations (Lu

et al. 2012; Xu et al. 2011; Choudhary et al. 2009). Because the interplay between metabolism and epigenetics is implicated in cancers, it has become a therapeutic target. For example, ivosidenib is an IDH1 inhibitor for treating relapsed AML (Acute Myeloid Leukemia)(Dhillon 2018). By inhibiting DNA methyltransferases, decitabine has also been used for treating AML (Malik & Cashen 2014). To sum up, the interplay between epigenetics and cellular metabolism plays an important role in many physiological processes. The dysregulation of such an interplay can be associated with pathological conditions.

1.2 The scope of this study

The major focus of this thesis is the relationship between acetyl-CoA metabolism, chromatin acetylation and gene expression. In the following sections of this chapter, a series of topics on it will be introduced sequentially, including acetyl-CoA as a central metabolite (section 1.3), acetyl-CoA as the substrate for protein acetylation (section 1.4-1.5), histone acetylation and its related enzymes and readers (section 1.6) and chromatin remodelling complexes (section 1.7). It is followed by the introduction to the YMC (yeast metabolic cycle, section 1.8), which is the major experimental system used in this study. Some new techniques have also been developed based on NET-seq (native elongating transcripts sequencing) during the progression of this study, so their relevant background will be explained at the end (section 1.9).

1.3 The nucleocytosolic and mitochondrial pools of acetyl-CoA

Acetyl-CoA is a central metabolite as well as the indispensable substrate for histone and non-histone protein acetylation (Menzies et al. 2016; Bravo-San Pedro et al. 2015). Its molecular structure is shown in Figure 1.2. The acetyl group (the blue part) is connected to the coenzyme A (the black part) via a thioester bond. This bond is chemically reactive and allows the acetylation of proteins, which is the transfer of the acetyl group from acetyl-CoA to proteins. Acetyl-CoA is also a charged molecule. As a result, biological membranes are impermeable for

acetyl-CoA without regulated transportation systems. However, the presence of nuclear pores permits it to diffuse freely between the nucleus and the cytosol. Hence, in a cell, acetyl-CoA is divided into two major pools, which are the nucleocytosolic pool and the mitochondrial pool (Bravo-San Pedro et al. 2015; Galdieri et al. 2014). The metabolism of acetyl-CoA in both pools can be variable in different species. The discussion below is mainly based on budding yeast, *S.cerevisiae*, unless otherwise stated. The metabolic pathways involving acetyl-CoA are briefly summarized in Figure 1.3.

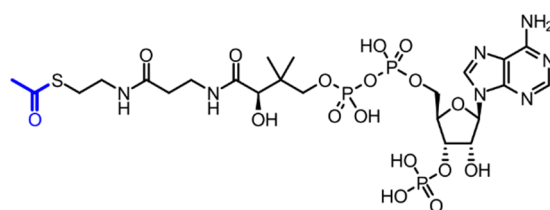


Figure 1.2 The chemical structure of acetyl-CoA. The reactive acetyl group is colored blue. The rest part is composed of a cysteine and a derivative of pantothenate (Vitamin B5).

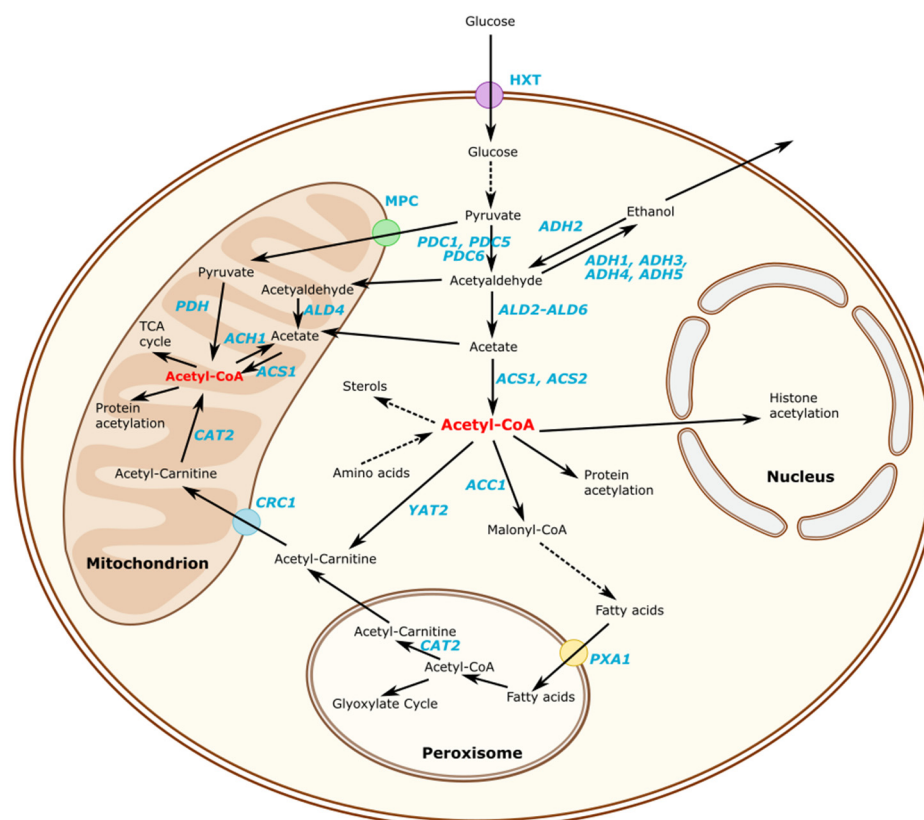


Figure 1.3 The metabolic pathways of acetyl-CoA in *S.cerevisiae*. The cellular acetyl-CoA is broadly separated into the nucleocytosolic pool and the mitochondrial pool. The acetyl-CoA in the peroxisomes belong to the nucleocytosolic pool. Metabolites are written in black, while the genes coding the enzymes for the metabolic reactions are in blue italics. Some metabolic pathways involving multiple steps are simplified by dashed arrows. (Modified from Galdieri L et al, 2014).

First of all, intracellular glucose is converted into pyruvate via the multistep glycolysis pathway in the cytosol. Acetyl-CoA in the nucleocytoplasmic pool can be synthesized from three major sources, pyruvate, fatty acids and amino acids. They will be discussed sequentially. First, pyruvate can be converted to acetaldehyde by pyruvate decarboxylases in the cytosol (Pronk et al. 1996). Acetaldehyde is then oxidised to acetate by multiple aldehyde dehydrogenases (Gombert et al. 2001). Acetate can be used to produce acetyl-CoA by the two acetyl-CoA synthetases, Acs1 and Acs2 (van den Berg et al. 1996). The former is present in both cytosol and mitochondria, while the latter is mainly present in cytosol and nucleus, which is important for histone acetylation (Takahashi et al. 2006; Kals et al. 2005; Huh et al. 2003). It is worth to note here that acetaldehyde and ethanol can be interconverted by different alcohol dehydrogenases (Galdieri et al. 2014). The free diffusion of ethanol between cells and the medium might allow the synchronization of the nucleocytoplasmic pools of acetyl-CoA between cells, which might further contribute to the synchronization of cells during the yeast metabolic cycle (Futcher 2006). This will be discussed in detail in section 3.3.4. In addition to pyruvate, acetyl-CoA can also be synthesized by fatty acids through beta-oxidation (Galdieri et al. 2014). For *S. cerevisiae*, beta-oxidation occurs in peroxisomes exclusively, because *S. cerevisiae* does not possess the enzymes for mitochondrial beta-oxidation. It is also that case for many other fungi and plants (van Roermund et al. 2003; Kretschmer et al. 2012). However, mammalian cells have both the mitochondrial and the peroxisomal beta-oxidation pathways (Wanders et al. 2010; Wanders et al. 2001). Acetyl-CoA is transported from peroxisomes to the cytosol by using acetyl-Carnitine as the intermediate (Franken et al. 2008). Lastly, branched chain amino acids, such as valine, leucine and isoleucine can be converted to acetyl-CoA as well via the transamination and decarboxylation reactions sequentially (Harris et al. 2005). The acetyl-CoA in the nucleocytoplasmic pool has three major functions. First, it can be used for the biosynthesis of various molecules, including fatty acids, steroids, certain amino acids and carbohydrates (via the glyoxylate cycle)(Bravo-San Pedro et al. 2015). Second, it can be sent to mitochondria

for energy production (van Roermund et al. 1999). Third, it can be used as the substrate for the acetylation of histone and non-histone proteins.

The acetyl-CoA in the mitochondrial pool can be synthesized by pyruvate or acetate or directly transported from the cytosol. In terms of the synthesis pathways, pyruvate generated from glycolysis is imported into the mitochondrial matrix by MPC (Mitochondrial Pyruvate Carrier) (Herzig et al. 2012). Pyruvate is further decarboxylated by the PDH complex (pyruvate dehydrogenase) in a CoA and NAD⁺-dependent manner, which produces acetyl-CoA, CO₂ and NADH (Patel et al. 2014). Apart from the pyruvate pathway, acetate from the cytosol can diffuse across the mitochondrial membrane. It can also be converted into acetyl-CoA by Acs1. Regarding its transport, cytosolic acetyl-CoA can also be transferred to mitochondria in the form of acetyl-Carnitine (Franken et al. 2008). Acetyl-CoA in the mitochondrial pool can be used to produce energy by joining the TCA cycle via the condensation reaction with oxaloacetate.

As was mentioned previously, acetyl-CoA in the nucleocytoplasmic pool can join the mitochondrial pool. Theoretically the reverse should happen, because the acetyl-CoA in mitochondria can be converted back into acetate by Ach1 (Acetyl-CoA hydrolase 1) and acetate can diffuse into the cytosol (Lee et al. 1996). However, there is evidence showing that it does not happen in a significant amount in budding yeast (Takahashi et al. 2006). In fact, for *S.cerevisiae*, the concentration of acetyl-CoA in the mitochondrial pool is more than 20 times higher than that in the nucleocytoplasmic pool (Weinert et al. 2014). Hence, the acetyl-CoA substrate for histone acetylation is mainly provided by the nucleocytoplasmic pool, rather than the mitochondrial pool. However, for mammalian cells, the reverse does happen. Apart from the Carnitine shuttle mechanism, mammalian cells also possess ACL (ATP citrate lyase), which budding yeast cells do not have (Galdieri et al. 2014). It allows the citrate exported from mitochondria to be converted back to acetyl-CoA in the cytosol (Zaidi et al. 2012). Actually,

this pathway is shown to be the major contributor to the nucleocytosolic pool of acetyl-CoA in mammalian cells (Mahmood et al. 2016; Wellen et al. 2009).

1.4 An overview of protein acetylation

Apart from being a critical metabolite in various biochemical pathways, acetyl-CoA is also the indispensable substrate for protein acetylation. According to the acetylation positions, protein acetylation can be divided into two types. One is the N-terminal acetylation. It happens on the alpha-amino group at the N-terminus of the nascent peptide mainly during the process of translation (Starheim et al. 2012). It is found that 84% and 57% of human and yeast proteins are acetylated at their N-terminus (Arnesen et al. 2009). This modification is catalysed by N-terminal acetyltransferases and is normally irreversible. It has been shown to influence various aspects of a protein, including synthesis, localization and stability (Starheim et al. 2012). The other type of protein acetylation is lysine acetylation. As the name suggests, an acetyl group is transferred from acetyl-CoA to the epsilon-amino group on the lysine residue of a histone or non-histone protein. This modification is normally added to the lysine residue in a certain context of amino acid sequences post-translationally by KATs (lysine acetyltransferases). Unlike the N-terminal acetylation, it can be reversed by various KDACs (lysine deacetylases). It is worth to note here that KATs and KDACs are sometimes called HATs (Histone acetyltransferases) and HDACs (Histone deacetylases). For more details about HATs and HDACs, please refer to section 1.6.1 and 1.6.2. Apart from being catalysed by enzymes, lysine acetylation can also happen automatically in a slightly alkaline solution ($\text{pH} > 7.5$), such as the matrix of mitochondria (Wagner & Payne 2013; Paik et al. 1970). Lysine acetylation can exert its impacts via three major mechanisms. First, due to the epsilon-amino group, lysine is a positively charged amino acid, so it normally has a relatively strong electrostatic interaction with negatively charged moieties. The acetylation can weaken the interaction by reducing the positive charge on lysine. Second, acetylated lysine residues can recruit other factors via the

physical interaction with certain protein domains, such as bromodomains. Third, lysine can also undergo other functional post-translational modifications, such as methylation and ubiquitination. Acetylation can occupy the lysine residue and prevent other modifications from happening. For example, acetylation can regulate NF- κ B by competing with ubiquitination (Weber et al. 2011). To sum up, acetyl-CoA is critical for the cells, because it is the substrate for both types of protein acetylation. Next, the functions of lysine acetylation on non-histone proteins will be briefly discussed first, followed by the detailed discussion of histone acetylation (section 1.6).

1.5 The functions of lysine acetylation on non-histone proteins

Lysine acetylation on non-histone proteins can regulate these proteins in multiple ways. First, it can influence the activity of enzymes. Acetylation at Lys661 residue of AceCS1 (mammalian cytoplasmic acetyl-CoA synthetase) or at Lys642 residue of AceCS2 (mammalian mitochondrial acetyl-CoA synthetase) can strongly inhibit their enzyme activities. In both cases, acetylation neutralizes the positive charge of the lysine residue at their active site. Furthermore, the inhibited AceCS1 and AceCS2 can restore their enzymatic activities through deacetylation by Sirtuins (Hallows et al. 2006). Second, it can regulate the functions of transcription factors. HNF-4 (hepatocyte nuclear factor-4) is a transcription factor that is crucial for the development of the liver (DeLaForest et al. 2011). It can be acetylated at several lysine residues located in its nuclear localization sequence. The acetylation has been shown to promote both its nuclear retention and its DNA binding (Soutoglou et al. 2000). Transcription factor CREB (cAMP-responsive element binding) is another example. It is involved in various physiological processes, including immune responses and neuronal development (Wen et al. 2010; Lonze & Ginty 2002). The acetylation of the lysine residues in its activation domain can enhance its transcription activation effects (Lu et al. 2003). Third, acetylation is able to prevent or induce the degradation of proteins. For instance, sphingosine kinase is a component of the

G-protein coupled signalling pathway. It can be acetylated by lysine acetyltransferase, p300. The acetylation blocks the lysine residue for the ubiquitination, which leads to the stabilization of sphingosine kinase (Yu et al. 2012). On the other hand, acetylation can also promote degradation. PCK1 (Phosphoenolpyruvate carboxykinase 1) is a metabolic enzyme involved in gluconeogenesis (Tang et al. 2018). It can also be acetylated by p300. However, the acetylated lysine allows the interaction between PCK1 and E3 ubiquitin ligase. The latter promotes the proteasome-mediated degradation of PCK1 (Jiang et al. 2011). To summarize this paragraph, lysine acetylation can regulate non-histone proteins in various ways.

1.6 Histone (lysine) acetylation

As was mentioned in the first section, eukaryotic DNA is wrapped around histone octamers. Each histone octamer is comprised of two copies of each of the four core histone proteins, H2A, H2B, H3 and H4. Histone acetylation, first characterized by Allfrey et al in 1964, mainly occurs at the lysine residues on the N-terminal tails of the four core histones (Allfrey et al. 1964). However, some acetylation marks, such as H3K56Ac, H4K79Ac are located at the globular cores of histones (Lawrence et al. 2016). Histone acetylation is associated with a lot of important molecular processes, such as activating gene expression, histone deposition and DNA damage repair (Shahbazian & Grunstein 2007). At the cellular level, it is becoming clear that histone acetylation can provide a linkage between metabolism and various cellular responses, such as growth, proliferation and differentiation (Cai et al. 2011; Shi & Tu 2013; Chisolm & Weinmann 2018). Many diseases, such as cancer and diabetes are associated with aberrant histone acetylation levels (Villeneuve & Natarajan 2010; Glozak & Seto 2007; Ropero & Esteller 2007). In a cell, histone acetylation needs to be “written, erased and read” properly by HATs (histone acetyltransferases), HDACs (histone deacetylases) and bromodomains respectively. In this section, the three players will be introduced first. It is followed by a discussion on the role of histone acetylation in activating gene expression.

1.6.1 The classification, specificity and metabolic regulation of histone

acetyltransferases (HATs)

HATs add acetylation marks to histones. They can be classified in two ways. First, based on the sequence and structural similarity, HATs can be grouped into 4 classes, including GNAT family, MYST family, p300/CBP family and SRC/p160 family (Table 1.1) (Steunou et al. 2014). The members in the GNAT (Gcn5-related acetyltransferases) family are homologous to Gcn5 in yeast (Neuwald & Landsman 1997). Interestingly, there is a bromodomain in the c-terminal region of most of the HATs in this family, including Gcn5 (Salah Ud-Din et al. 2016). HATs in the MYST family are homologous to yeast Esa1. The four classes are the canonical HATs. Apart from that, there also exists proteins that do not belong to these classes but possess the activity of HAT. For example, TFIID is a transcription initiation factor. In yeast, its TAFII250 subunit can acetylate histone H3 and H4 in vitro (Mizzen et al. 1996). Following synthesis, histones can be acetylated before or after being deposited onto the chromatin. The second classification is based on this, which divides HATs into two classes, HAT-A and HAT-B. HATs in HAT-A mainly reside in the nucleus and add acetyl groups to the histones on the chromatin. This class includes the majority of the HATs and is mainly involved in the regulation of gene expression. HATs in HAT-B are mainly localized in the cytosol and are normally responsible for the acetylation of newly synthesized histone proteins before chromatin deposition (Steunou et al. 2014). In yeast, only two HAT-B enzymes are found. They are Hat1 and Rtt109 (Masumoto et al. 2005; Parthun et al. 1996). The former acetylates histone H4 at K5 and K12 residues (Kleff et al. 1995; Sobel et al. 1995), while the latter is responsible for the acetylation of K56 at histone H3 (Masumoto et al. 2005). H3K56Ac is known to facilitate the deposition of newly synthesized onto the DNA during the S phase via enhancing the binding between histone H3 and histone chaperones, such as Caf1 and Rtt106 (Li et al. 2008). These acetylation marks are quickly removed after histone deposition.

	Name	<i>Homo sapiens</i> (Hs)	<i>Drosophila melanogaster</i> (Dm)	<i>Saccharomyces cerevisiae</i> (Sc)	<i>Schizosaccharomyces pombe</i> (Sp)
GNAT family	KAT1	HAT1	CG2051	Hat1	Hat1/Hag603
	KAT2		dGCN5/PCAF	Gcn5	Gcn5
	KAT2A	GCN5			
	KAT2B	PCAF			
	KAT9	ELP3	dELP3/CG15433	Elp3	Elp3
	KAT10			Hpa2	
	KAT14	CSRP2BP	ATAC2		
MYST family		HAT4	HAT4		
	KAT5	TIP60/PLIP	dTIP60	Esa1	Mst1
	KAT6		(CG1894)	Sas3	(Mst2)
	KAT6A	MOZ/MYST3	ENOK		
	KAT6B	MORF/MYST4			
	KAT7	HBO1/MYST2	CHM		(Mst2)
	KAT8	MOF/MYST1	dMOF/(CG1894)	Sas2	(Mst2)
p300/CBP family	KAT3		dCBP/NEJ		
	KAT3A	CBP			
	KAT3B	p300			
	KAT11			Rtt109	
SRC/p160 family	KAT13A	SRC1/NCOA1			
	KAT13B	ACTR/NCOA3			
	KAT13C	GRIP1/NCOA2			
	KAT13D	CLOCK			
Others	KAT4	TAF1	dTAF1	Taf1	Taf1
	KAT12	TFIIIC90			
				Nut1	

Table 1.1 The four main classes of lysine acetyltransferases. The homologs in human, fly, budding yeast and fission yeast are listed in separated columns. KAT4 and 12 are orphan KATs, so they are put into the category of ‘Others’ (Modified from Steunou et al, 2014)

Although HATs normally do not act on a specific single residue like histone methyltransferases, different HATs do have preferences for their substrates (Steunou et al. 2014). For example, Gcn5 is over 1000 fold more efficient in acetylating histone H3 than H4, while Esa1 mainly catalyse the acetylation of histone H4 and H2A (Marmorstein & Zhou 2014; Doyon et al. 2004; Allard et al. 1999). Additionally, many HATs are present in protein complexes, so that other subunits of the complex can also influence the substrate specificities. Gcn5 can be a part of the SAGA complex, while Esa1 is a subunit of the NuA4 complexes. Gcn5 on its own mainly acetylates histone H3 at lys14. However, in the presence of Ada2, which is another subunit of the SAGA complex, Gcn5 is able to acetylate lys18 of histone H3 (Balasubramanian et al. 2002). Within the whole SAGA complex, Gcn5 acetylates lys14 and lys18 better than lys9 and lys23 of histone H3 (Grant et al. 1997). Some subunits of these complexes can also interact with transcription factors, which allows HATs to be targeted to the promoters of specific genes. Finally, it is worth noting that many HATs are able to acetylate non-histone proteins as well, so HATs are more often referred to as KATs (lysine acetyltransferases). For instance, yeast

Gcn5 is able to acetylate several other members of the SAGA complex (Cai et al. 2011). Mammalian HAT p300 can acetylate p53 and HDAC1, in addition to histone H3 and H4 (Qiu et al. 2006; Ito et al. 2001).

The activity of HATs can be regulated by metabolites. There are two major mechanisms. First, the availability of the acetyl group donor, acetyl-CoA, can directly influence the enzymatic activities of HATs, due to the fact that the K_d (dissociation constants) and K_m (Michealis constants) of many HATs fall in the fluctuation range of the cellular concentration of acetyl-CoA. For example, the K_d and K_m of yeast Gcn5 are determined to be $8.5\mu\text{M}$ and $2.5\mu\text{M}$ respectively in vitro, while the intracellular concentration of acetyl-CoA can fluctuate between $3\mu\text{M}$ and $30\mu\text{M}$ (Langer et al. 2002; Cai et al. 2011; Berndsen & Denu 2008). For human cells, the concentration of acetyl-CoA can change between $2\mu\text{M}$ and $20\mu\text{M}$ (Lee, Carrer, et al. 2014). Human HAT, p300 has a K_d of $6.7\mu\text{M}$ (Liu et al. 2008), which is in that fluctuation range. In fact, some metabolic enzymes are also found to form complexes with HATs in the nucleus to increase the local concentration of acetyl-CoA. For instance, PDC (pyruvate decarboxylase) can form a complex with p300 and AhR (aryl hydrocarbon receptor). PDC provides acetyl-CoA to p300, which acetylates the histones and promote the transcription of AhR responsive genes (Matsuda et al. 2016). Second, some metabolites are able to inhibit the function of HATs in a competitive manner. During transfer of the acetyl group to proteins, the acetyl-CoA is converted into CoA. CoA has been shown to inhibit a variety of HATs competitively with different strength (Table 1.2 (Fan et al. 2015)). Unlike yeast Gcn5, human Gcn5 has an amino acid substitution in its active centre, which increases its affinity for acetyl-CoA and makes its K_d ($0.64\mu\text{M}$) (Kirk G. Tanner et al. 2000) fall below the physiological range of acetyl-CoA concentrations in human cells. However, human Gcn5 is also more sensitive to the inhibition caused by CoA (Tanner et al. 2000). The experimental results show that the activity of human Gcn5 correlates with the ratio between acetyl-CoA and CoA, rather than the absolute concentration of CoA (Lee, Haas, et al. 2014). Apart from CoA, fatty acyl-CoA can also compete

for the binding of acetyl-CoA to HATs. For example, palmitoyl-CoA and propionyl-CoA, which are both endogenous cellular metabolites, can inhibit human Gcn5 and pCAF respectively (Montgomery et al. 2015).

Enzymes	K_m	K_d	K_i (inhibition coefficient by CoA)
Gcn5 (yeast)	2.5 μ M	8.5 μ M	6.7 μ M
Gcn5 (human)	0.98 μ M	0.64 μ M	6.7 μ M
P/CAF (human)	0.64 μ M	No data	0.57 μ M
Esa1 (yeast)	0.9 μ M	No data	1.8 μ M
p300 (human)	6.7 μ M	No data	21 μ M

Table 1.2 The enzymatic properties of various lysine acetyltransferases. (Modified from Fan, J et al, 2015)

1.6.2 The classification, specificity and metabolic regulation of histone

deacetylases (HDACs)

HDACs remove acetylation marks from histones. Based on their enzymatic activities, HDACs can be categorized into two groups. One group are classical HDACs. In their active centre, there is a zinc ion chelated by histidine and aspartic acid residues (Lombardi et al. 2011). Their deacetylase activities are dependent on the zinc ion. However, apart from water, the classical HDACs do not require other cellular metabolites as the substrate of deacetylation reaction. These HDACs are normally present in protein complexes and other components of the protein complexes are also required for their enzymatic activity (Steunou et al. 2014). In *S.cerevisiae*, Hda1 forms a tetrameric complex with Hda2, Hda3 and another copy of Hda1 (Carmen et al. 2002). Rpd3 can be present in two complexes, Rpd3L and Rpd3S complexes. Other subunits of the protein complexes also target HDACs to specific regions on the genome to regulate transcription. Rpd3L can be recruited to the promoter regions by transcription factor Ume6, which represses gene expression (Carrozza, Florens, et al. 2005). Rpd3S is mobilized to the gene body via the interaction with H3K36me and the Ser2P and Ser5P of RNAPII CTD (c-

terminal repeat domain), which can suppress the transcription of cryptic unstable transcripts (Joshi & Struhl 2005; Carrozza, Li, et al. 2005; Keogh et al. 2005). Based on the sequence and structural homology, this group can be further divided into three classes, I, II and IV (Table 1.3 (Steunou et al. 2014)). Hda1 and Rpd3 belong to class I and class II respectively. The other group of HDACs is sirtuins. They all belong to class III HDACs, based on the sequence homology. Unlike the classical HDACs, recombinant sirtuins can exhibit deacetylase activities without other protein partners. However, they require NAD⁺ as the substrate. During the deacetylation of histones, NAD⁺ is first cleaved into NAM and ADPR (ADP-ribose). The latter accepts the acetyl group transferred from the lysine residue of histones and forms acetyl-ADPR. In budding yeast, Sir2 and the other four sirtuins are required for heterochromatin silencing (Dang 2014).

		<i>Homo sapiens</i> (Hs)	<i>Drosophila melano-</i> <i>gaster</i> (Dm)	<i>Saccharomyces</i> <i>cerevisiae</i> (Sc)	<i>Schizosaccharomyces</i> <i>pombe</i> (Sp)
Class I		HDAC1	Rpd3	Rpd3	Clr6
		HDAC2	HDAC3	Hos1	Phd1
		HDAC3		Hos2	
		HDAC8			
Class II	A	HDAC4	HDAC4	Hda1	Clr3
		HDAC5		Hos3	
		HDAC7			
		HDAC9			
	B	HDAC6	HDAC6		
		HDAC10			
Class III	I	SIRT1	Sir2	Sir2	Sir2
		SIRT2	Sirt2	Hst1	Hst2
		SIRT3		Hst2	Hst4
				Hst3	
	II	SIRT4	Sirt4	Hst4	
	III	SIRT5			
	IV	SIRT6	Sirt6		
		SIRT7	Sirt7		
	Class IV	HDAC11	HDACX		

Table 1.3 The four main classes of lysine deacetylases. The homologs in human, fly, budding yeast and fission yeast are listed in separated columns.

Generally, HDACs do not have a strong specificity for their targets. For instance, most histone acetylation marks on all the four core histone proteins can be removed by yeast Rpd3. Hda1 has a preference for both histone H3 and H2B, while Hos2 has its deacetylation activity on histone both H3 and H4 (Robyr et al. 2002). Actually, HDACs can also deacetylate non-histone

proteins. They are also referred to as KDACs (lysine deacetylases). However, there are some exceptions in the Sirtuin group. Yeast Hst2 has a strong preference for H4K16ac (Vaquero et al. 2006). Mammalian sirtuin, SIRT6 and SIRT7, can deacetylate H3K9ac and H3K18ac specifically (Barber et al. 2012; Michishita et al. 2008).

Like HATs, the enzymatic activities of HDACs can also be affected by the concentrations of cellular metabolites. Because of the different catalytic mechanisms, here the classical HDACs and the sirtuins are discussed separately. In terms of the classical HDACs, many of them can be strongly inhibited by butyrate and beta-hydroxybutyrate. The former is produced by bacteria in the gut via fermenting dietary fibre and is a normal energy source for the colonocytes. The latter is a type of ketone body which is produced by breaking down fatty acids during starvation (Etchegaray & Mostoslavsky 2016). Administering beta-hydroxybutyrate to mice can cause elevated H3K9ac and H3K14ac and changes in gene expression (Shimazu et al. 2013). In terms of the sirtuins, like HATs, their activities can be influenced by the availability of their substrate, NAD⁺, because the K_d of several sirtuins, such as SIRT1 (95μM (Cantó et al. 2015)) and SIRT2 (83μM (Cantó et al. 2015)) are close to the concentration of NAD⁺ in the nucleus (109μM (Cambronne et al. 2016)). NAD⁺ is a metabolite participating in the redox reactions in oxidative phosphorylation. It can be generated de novo or recycled via the salvage pathway. Hence, the cellular levels of NAD⁺ can be influenced by both the diet composition and the calorie intake. Both the high fat diet and diabetes can decrease cellular levels of NAD⁺. Reduced activity of sirtuins has been observed in these conditions (Yoon et al. 2015; Gomes et al. 2013; Yoshino et al. 2011). Furthermore, the activities of sirtuins can be inhibited allosterically by NAM (Menzies et al. 2016) (converted from NAD⁺ after protein deacetylation), while free fatty acids are shown to increase the activities of sirtuins (Feldman et al. 2013).

1.6.3 Bromodomains: histone acetylation reader

Histone acetylation marks can recruit other protein factors via interaction with bromodomains. The general structure of a bromodomain is a bundle structure comprised of 4 helices (α A, α B, α C and α Z). The BC and ZA loops constitute a hydrophobic binding pocket for the acetylated lysine residue, which has a neutralized charge (Brownlee et al. 2012). In human, 61 bromodomains have been found in a total of 46 proteins (Fujisawa & Filippakopoulos 2017), while budding yeast has 15 bromodomains found in 10 proteins (Zhang et al. 2010). For budding yeast, its bromodomains fall into 4 major groups. Rsc1, Rsc2 and Rsc4 (bromodomain 1) belong to the same group, while Bdf1, Bdf2 and Gcn5 belong to another group. Spt7 and Rsc4 (bromodomain 2) are in two separate groups as well (Zhang et al. 2010). Most of these bromodomain containing proteins are regulators of transcription, including subunits of chromatin remodelling complexes, chromatin modifiers and transcription factors. Different bromodomains can have narrow or wide specificities for acetylated lysine residues. For instance, yeast Snf2 is able to bind both acetylated H3 and H4 peptides, while the bromodomain of Spt7 only exhibits a weak interaction with acetylated H3 (Hassan et al. 2006). The cooperation with other domains can increase the binding specificity of bromodomains. A typical example is human protein TRIM33. It has a bromodomain connected to a PHD domain, which recognizes H3K9me3. As a result, its bromodomain can specifically recognize H3K18ac, which is close to H3K9me3. This arrangement also allows for a specific binding orientation of TRIM33 (Xi et al. 2011).

The binding affinity of bromodomains to acetylated lysine residues can be enhanced by the avidity effect. Specifically, tandem bromodomain can bind multiple acetylated lysine residues more strongly. For example, human TAF1 protein has a tandem bromodomain. It has a stronger interaction with multiple acetylated histone H4 than single acetylated H4 (Jacobson et al. 2000).

Most histone acetylation binders are bromodomains. However, there are few exceptions. For example, the DPF (double PHD finger) domain of Dpf3 can interact with both acetylated histone H3 and H4 (Lange et al. 2008). The PH (Pleckstrin Homology) domain of Rtt106 is able to bind H3K56ac (Su et al. 2012).

1.6.4 The cellular concentration of acetyl-CoA can fluctuate and it is often associated with changes in protein acetylation

Unlike cellular ATP levels, which are tightly regulated, the cellular concentration of acetyl-CoA can fluctuate in different physiological and pathological conditions. There is some evidence suggesting that the acetyl-CoA concentration of the nucleocytoplasmic pool is more subject to variations than that of the mitochondrial pool (Mariño et al. 2014). Ideally, the changes in the acetyl-CoA concentrations in the two pools should be discussed separately. However, due to the technical limitations, the majority of the measurements on acetyl-CoA concentrations cannot distinguish different subcellular compartments. Hence, the discussion in this section is mainly based on whole cells or whole organs.

Generally, the cellular level of acetyl-CoA is low in the starved/fast states and high in the fed states. For budding yeast cells, the acetyl-CoA level is maintained at a high level during the log phase of growth when carbon sources and nutrients are abundant, which is associated with high levels of histone acetylation (Cai et al. 2011). Once Acs2 (responsible for synthesizing acetyl-CoA in the nucleocytoplasmic pool) is inactivated, there is a dramatic decrease in histone acetylation (Takahashi et al. 2006). After the diauxic shift, the cells have consumed the majority of the carbon sources. The acetyl-CoA level is reduced to a low level, which is also associated with histone hypoacetylation (Galdieri et al. 2010). The YMC is an ideal system to demonstrate the fluctuation of acetyl-CoA in various cellular metabolic states. This will be discussed in details in section 1.9 and section 3.1.1. Mammalian cell lines also exhibit a similar phenomenon. After refeeding the immortalized mouse fibroblasts, an ACLY-mediated

increase in acetyl-CoA in the nucleocytosolic pool is associated with histone acetylation (Wellen et al. 2009). Marino et al in 2014 reported a rapid decrease in acetyl-CoA levels in cancer cell lines after switching the culture medium to the nutrient-free medium. Furthermore, this phenomenon can be suppressed by supplementing metabolites upstream of acetyl-CoA or reproduced in the normal rich medium by disturbing the synthesis of nucleocytosolic acetyl-CoA (Mariño et al. 2014). Apart from the experiments on cells, the fluctuations in acetyl-CoA levels are also observed on organs by studying the mouse model. After starving the mouse for 24 hours, the acetyl-CoA and protein acetylation levels in many organs experience a dramatic decrease (Mariño et al. 2014). The molecular mechanism behind it is suggested to be both the reduced cellular glucose uptake and the reduced ACLY phosphorylation (required for its enzymatic activity of producing acetyl-CoA) (Lee, Haas, et al. 2014). To sum up, the cellular concentrations of acetyl-CoA can respond to the nutrient states and influence protein acetylation, including histone acetylation.

1.6.5 Histone acetylation is associated with active gene expression

It is widely accepted that histone acetylation is associated with active gene expression. The early work done by Robert et al in 2004 showed that the genome wide localization of histone acetyltransferases, Gcn5 and Esa1 are strongly enriched for the promoters of actively expressed genes (Robert et al. 2004). Later, many experiments suggested that the SAGA and NuA4 complex, which harbours Gcn5 and Esa1 respectively, can be recruited to the promoters of genes by their interaction with transcription activators, such as Gcn4 and Tra1, which leads to the acetylation of histones at the promoter regions as well as the activation of transcription (Kuo et al. 2000; Brown et al. 2001; Knutson & Hahn 2011; Fishburn et al. 2005; Utley et al. 1998). Similarly, p300/CBP, a mammalian HAT, can be found at the promoter regions of nuclear hormone receptor-dependent genes. Its binding at these promoters correlates with both the acetylation of histones and the activation of gene expression (Chakravarti et al. 1996).

Tropberger et al in 2013 showed that one of the targeted residues of p300/CBP is the lysine 122 residue of histone H3, which is positioned at the lateral surface of histone octamers. The acetylation of this residue is found to be sufficient for transcription initiation (Tropberger et al. 2013). Interestingly, the deletion of *RPD3*, which is an HDAC, can lead to histone hyperacetylation in the gene bodies. This aberrant histone hyperacetylation is associated with the transcription initiation of cryptic unstable transcripts in the gene bodies (Carrozza, Li, et al. 2005; Smolle et al. 2013).

Although currently it is not strictly proven that histone acetylation leads to active gene expression, it is believed that the associated gene activation effects of histone acetylation are linked to the increased mobility of nucleosomes which can negatively regulate transcription by preventing the formation of the pre-initiation complex. Two mechanisms have been proposed to explain why histone acetylation can increase nucleosome mobility. One is the charge neutralization effect (Allfrey et al. 1964). The positive charges on the histone N-terminal tails conferred by multiple lysine residues can create a strong electrostatic interaction with the negatively charged DNA. The addition of the acetyl group reduces the positive charges on the lysine residue, which loosens the interaction and increases nucleosome mobility. The other mechanism is that histone acetylation may increase nucleosome mobility by recruiting other functional factors, such as histone chaperones and chromatin remodelling factors (Yang 2004; Hassan et al. 2002a; Dhalluin et al. 1999; Haynes et al. 1992). Several chromatin remodelling complexes, including the RSC complex, the SWI/SNF complex and the SWR1 complex, have subunits that contain bromodomains, which can target them to the histone acetylated regions. The chromatin remodelling factors can then consume ATP to provide energy for sliding or removing nucleosomes. They will be discussed in more detail in the next section.

1.7 Chromatin remodelers

As was mentioned before, the eukaryotic DNA is wound around numerous histone octamers to form a beads-on-a-string structure. The chromatin remodeler is a class of chromatin regulatory proteins or protein complexes that can utilize the energy provided by the hydrolysis of ATP to move, evict, deposit or edit nucleosomes on the DNA ('edit' means changing the histone compositions of nucleosomes). All the known chromatin remodelers have three shared features (Clapier & Cairns 2014). First, they have a higher affinity for histones than DNA. Second, they all have a catalytic subunit, which contains an ATPase domain for hydrolysing ATP that is converted into mechanical energy to disrupt the contacts between DNA and nucleosomes. The ATPase domain is further divided into two RecA-like lobes, called DExx and HELICc respectively. Between the two lobes, there is an insertion of variable size. Third, chromatin remodelers have various regulatory domains or subunits which can modulate the ATPase domain, bind certain histone modifications or interact with other functional factors, such as transcription factors and histone chaperones. These features allow chromatin remodelers to recognize various chromatin features and execute their functions.

1.7.1 The three main functions of chromatin remodelers

Chromatin remodelers are involved in three major functions: assembling and spacing nucleosomes, modulating chromatin accessibility and performing nucleosome editing. Chromatin remodelers realize these functions by depositing, moving, evicting and editing nucleosomes (Clapier et al. 2017). Each remodeler can be involved in one or more of these functions. Nucleosome assembly and spacing happens extensively during DNA replication, when the newly synthesized DNA is wound around histone octamers to form nucleosomes. Chromatin remodelers can facilitate the formation of nucleosomes and adjust their spacing to form a periodic array by moving nucleosomes. Chromatin accessibility has a great impact on transcription. Normally, nucleosomes limit the access of various protein factors required for

transcription to their binding sites on the DNA. Chromatin remodelers can activate or repress transcription by increasing or decreasing the accessibility of chromatin via moving and evicting nucleosomes. Finally, chromatin remodelers can edit nucleosomes. That is, the canonical histones (H3, H4, H2A and H2B) of the nucleosomes at certain locations can be replaced with certain noncanonical histones, such as H2A.Z, CENPA and H3.3. This can result in various consequences, including changes in nucleosome stability and interactions between nucleosomes and other factors. Although chromatin remodelers can perform various manipulations on nucleosomes, a lot of evidence shows that their molecular mechanisms have a shared prototype. Different manipulations are achieved by slight modifications to that prototype (Clapier et al. 2017).

1.7.2 The four families of chromatin remodelers

Based on the structures of the catalytic subunits as well as the compositions of the accessory subunits, chromatin remodelers can be divided into four families, ISWI, CHD, INO80 and SWI/SNF. Generally, assembling and spacing nucleosomes involves the ISWI and CHD family, while the SWI/SNF and INO80 families mainly regulate chromatin accessibility and perform nucleosome editing respectively (Clapier et al. 2017). The characteristic features of the catalytic subunits of the four families are shown in figure 1.4. Within each family, there are a lot of subtypes which may have altered ATPase domains or different subunit compositions. For example, BAF and PBAF are both subtypes of the SWI/SNF family. For simplicity, the following paragraphs will mainly focus on the most conserved features within each family.

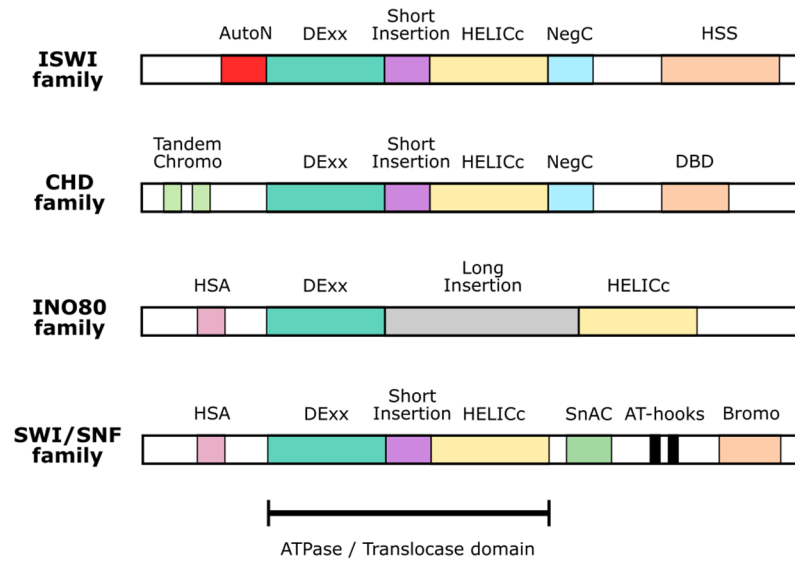


Figure 1.4 The domain arrangements of the catalytic subunits of the four families of chromatin remodelers. DExx and HELICc are the two RecA-like lobes of the ATPase domain. Some abbreviations: HSS (Hand-Sant-Slide) domain, Tandem Chromo (Chromodomain), DBD (DNA binding domain), HAS (helicase/SANT-associated) domain, SnAC (Snf2-ATP coupling) domain, Bromo (Bromodomain). (Modified from Clapier C. R. and Cairns B. R. 2014)

In terms of the ISWI family, the HSS domain near the c-terminus of its catalytic subunit can interact with both histones and DNA (Dang & Bartholomew 2007; Boyer et al. 2004). The ATPase domain is flanked by another two domains, AutoN and NegC. The two domains together with the HSS domain modulate the catalytic activity of the ATPase domain (Mueller-Planitz et al. 2013; Clapier & Cairns 2012; Grüne et al. 2003). Remodelers in the ISWI family are mainly involved in transcription repression. They can move nucleosomes and are essential for the establishment and the maintenance of a periodic nucleosome array, which can occlude the sites on the DNA required for the binding of transcription factors and the formation of the preinitiation complex (Yadon et al. 2010; Moreau et al. 2003; Alén et al. 2002; Whitehouse & Tsukiyama 2006). In fact, the ISWI remodelers often collaborate with HDACs to repress gene expression (Sugiyama et al. 2007; Burgio et al. 2008).

In terms of the CHD family, the catalytic subunit has tandem chromodomains near the N-terminus, which can interact with methylated histone H3 (Flanagan et al. 2005). There exist

some similarities between the ISWI and the CHD family, in terms of the domains in their catalytic subunits. The CHD family also has the NegC domain. The DBD (DNA-binding) domain in the CHD family is actually a part of the HSS domain in the ISWI family (Ryan et al. 2011). As its name suggests, the DBD domain can bind to DNA. The CHD family has diverse subtypes. They can participate in all the three functions of chromatin remodelers. Their influence on transcription can be promotive or repressive, depending on the subtypes. It is worth to note here that chromatin remodelers are not essentially protein complexes, because the CHD remodeler in budding yeast can perform its function with its single catalytic subunit (Flanagan et al. 2005).

In terms of the INO80 family, the long insertion between the two RecA lobes is the unique feature of this family (Eustermann et al. 2018). It can bind various proteins, including several ARPs (actin related proteins). Both this insert and the N-terminal HSA domain can interact with ARPs (Kapoor et al. 2013; Zhang et al. 2018). These ARPs can positively regulate the functions of remodelers in the INO80 family by enhancing the activity of the ATPase domain as well as the affinity for DNA and histones (Szerlong et al. 2008; Shen et al. 2003; Saravanan et al. 2012; Jónsson et al. 2004). Chromatin remodelers in this family are mainly implicated in nucleosome editing. It is partially mediated by these ARPs. The yeast SWR1 complex (a member of the INO80 family) can exchange the histone H3 in the +1 nucleosome for histone H2A.Z. One of the ARPs in the SWR1 complex is called Arp6. It can promote the physical interaction between the SWR1 complex and H2A.Z (Wu et al. 2005). This editing preferentially appears on inactive but inducible promoters (Zhang et al. 2005a; Guillemette et al. 2005; Santisteban et al. 2000; Adam et al. 2001).

The SWI/SNF family is similar to the INO80 family in that it has the HSA domain and can be regulated by ARPs (Szerlong et al. 2003; Mizuguchi et al. 2004a; Szerlong et al. 2008). Apart from that, the SWI/SNF family also have many unique features. The c-terminus of the catalytic

subunit in this family contains a bromodomain, which can bind to acetylated histones. The c-terminal AT-hook is a motif that can interact with DNA minor grooves (Clapier & Cairns 2014). The most typical function of remodelers in this family is regulating chromatin accessibility. They can induce transcription via sliding or removing nucleosomes to expose certain binding sites on the DNA. The composition of remodelers in this family can be highly variable, which allows them to modulate gene expression in a cell type specific and developmental stage specific manner (Clapier & Cairns 2014).

1.7.3 The chromatin remodelers in the SWI/SNF family contain one or more bromodomains

Most bromodomain containing remodelers are members of the SWI/SNF family. As was mentioned in the previous section, remodelers in this family at least have a bromodomain in the c-terminus of their catalytic subunit (Tang et al. 2010). This bromodomain is required for the optimal remodelling activities of the yeast SWI/SNF complex as well as its physical interaction with acetylated histone H3 and H4 (Awad & Hassan 2008b). In fact, depending on whether multiple bromodomains are present in the non-catalytic subunits of the remodelers, this family can be divided into two subfamilies. Most species have at least one member in each of the two subfamilies. For example, yeast possesses the RSC complex and the SWI/SNF complex. Both of them belong to the SWI/SNF family, however the former contains tandemly arranged bromodomains in many of its subunits. Their mammalian counterparts are the PBAF complex and the BAF complex respectively (Wang 2003). The BAF180 subunit of the PBAF complex contains six bromodomains arranged in tandem (Brownlee et al. 2012). Apart from the SWI/SNF family, some member of the INO80 family also contains bromodomains. The Bdf1 and Bdf2 subunits of yeast SWR1 complex have two bromodomains respectively. They can be recruited by acetylated histone H4, which targets the SWR1 complex to the nucleosomes for editing (Ladurner et al. 2003).

1.7.4 The SWI/SNF family and gene expression

The SWI/SNF family chromatin remodeler can be associated with active transcription. It was originally shown by Winston and Carlson that some subunits of the SWI/SNF complex are required for transcription (Winston & Carlson 1992). Their results also suggested that the complex can counteract the transcription suppressive effects of nucleosomes. Later in 1991, it was found that the yeast SWI/SNF complex can be recruited by transcription activators, such as Swi5 and Gcn4 (Yoon et al. 2003; Cosma et al. 1999). In 2002, Armstrong et al showed that the SWI/SNF complex is associated with the increased occupancy of RNAPII on genes (Armstrong et al. 2002). In 2008, Bryant G.O et al found that the SWI/SNF complex can promote the removal of nucleosomes in the promoter regions, which is important for the fast induction of inducible genes (Bryant et al. 2008).

There is both in vitro and in vivo evidence suggesting that the SWI/SNF family chromatin remodeling complexes, especially the RSC complex, is also associated with the transcription elongation step. Nucleosomes are one of the major barriers for RNAPII during the process of transcription elongation. It was originally found in 2006 by an in vitro transcription elongation assay that the RSC complex can help RNAPII overcome the mono-nucleosome barrier in a histone acetylation dependent manner (Carey et al. 2006). In a later study, the c-terminal residues of the Rsc4 subunit of the RSC complex were shown to be able to bind to the Rpb5 subunit of RNA polymerase II and the RSC complex can coimmunoprecipitate with RNAPII (Soularue et al. 2006). It was found that the RSC complex can be recruited to the gene bodies of actively transcribed genes (Spain et al. 2014). The recruitment requires the histone acetyltransferase activities of the SAGA and NuA4 complexes. Additionally, this process is also facilitated by kinases targeting the serine 2 position of the CTD of RNAPII (Spain et al. 2014).

1.8 An introduction to the yeast metabolic cycle

In nutrient-limited conditions, budding yeast (*S.cerevisiae*) cells, can exhibit a highly synchronized and periodic oscillation between two different metabolic states, which are named the high oxygen consumption (HOC) phase and the low oxygen consumption (LOC) phase (Satroutdinov et al. 1992). This phenomenon is referred to as the yeast metabolic cycle (YMC). In most cases, the YMC is achieved by growing cells in a chemostat in a fermenter (For more details about the fermenter, please refer to section 3.1.1.1). It was found that cellular metabolism and the cell cycle are synchronized and periodic during the YMC (Tu et al. 2005; Burnett et al. 2016). Moreover, many molecular events, including gene expression and histone modifications have highly dynamic temporal changes (Kuang et al. 2014). All these make the YMC an excellent system to study various research topics, including the relationship between metabolism, epigenetics and gene expression.

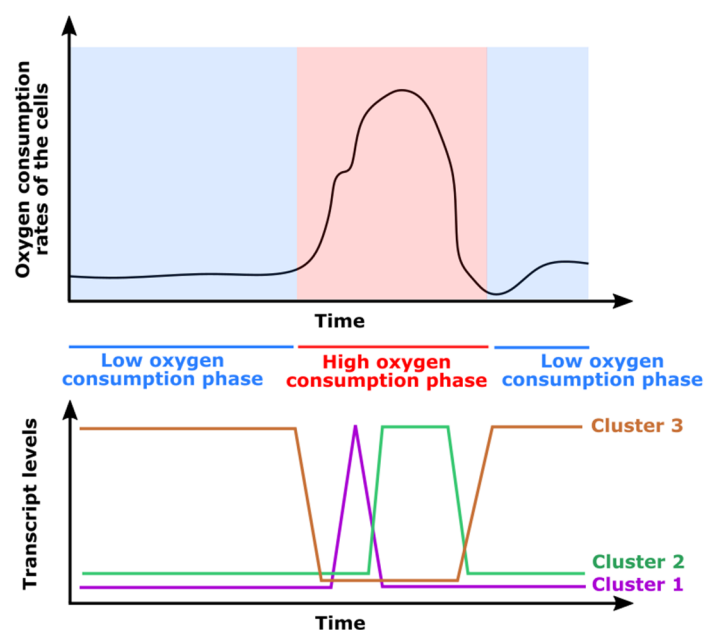


Figure 1.5 A schematic diagram showing the expression of genes during a typical cycle of the YMC. The upper panel shows the changes in the cellular oxygen consumption rates. The red and blue colours denote the high and low oxygen consumption phases respectively. The lower panel shows the typical changes in the transcript levels of the genes in the three clusters.

The YMC as a phenomenon was first documented in 1964 (CHANCE et al. 1964). However, it was not studied at the molecular level until 2005, when the temporal profile of gene expression during the YMC was measured genome-wide by microarray (Tu et al. 2005). It was found that at least 60% of all yeast genes have highly dynamic and periodic changes in their transcript levels. Furthermore, the temporal expression profiles of different genes also exhibit diverse shapes. Based on the different shapes, genes are roughly grouped into 3 clusters (Figure 1.5). The first cluster mainly contains the growth related genes, such as amino acid synthesis genes and ribosome biogenesis genes. Their transcript levels exhibit a sharp peak shortly after the cells enters the HOC phase. Notably, based on the normalized high temporal resolution RNA-seq data in this study, their transcript levels can increase by 3 folds within 10mins and then return to the original ground level in 10mins. Genes in the second cluster have a relatively broad peak of transcript levels in the late stage of the HOC phase. The second cluster is significantly enriched for genes related to the cell cycle and mitochondria. In fact, DNA replication also happens in a similar time. In the third cluster, the genes are mostly associated with autophagy, stress responses and anabolism. Typically, the expression of these genes seems to have two states. That is, they maintain high and low levels of transcripts during the LOC and HOC phase respectively. The clustering above is a rather general characterization of gene expression during the YMC. Actually, in terms of the expression profiles, genes in each individual cluster still have a lot of differences. In a later study, the YMC was studied with RNA-seq. The three clusters are further divided into several sub-clusters (Kuang et al. 2014). In this study, all the genes are divided into 16 clusters. A larger number of clusters decreases the granularity, which enables a finer characterization of gene expression during the YMC. To sum up, different genes have diverse temporal profiles of transcript levels during the YMC.

Apart from the transcript levels, more omics tools, such as mass spectrometry (MS) and ChIP-seq, were also applied to the YMC. The MS quantified the changes in the levels of 167 cellular metabolites during the YMC (Tu et al. 2007). Most of these metabolites, including acetyl-CoA

(the substrate for protein deacetylation), have periodic changes in their levels during the YMC. The ChIP-seq analysed the genome-wide levels of several histone acetylation and methylation marks at multiple time points during the YMC (Kuang et al. 2014). It reveals that some histone modifications are also highly dynamic during the YMC. For the same histone modification, different genes can have different temporal changes. For the same gene, different histone modifications also show different profiles. Notably, most of the genes in the first cluster (the growth related genes) mentioned above have a preceding peak of histone acetylation. In this study, both the MS data and the ChIP-seq will be utilized for bioinformatic analysis.

The YMC is an excellent experimental system to study the relationship between acetyl-CoA metabolism, chromatin acetylation and gene expression, for two major reasons. First, all the key factors of this research topic (including acetyl-CoA, histone acetylation and transcripts levels) show dynamic temporal profiles during the YMC and the relevant MS and ChIP-seq datasets are available in the literature. Second, the YMC is highly periodic and reproducible, which is indicated by the fact that the oxygen consumption curves of different YMC cycles can almost completely overlap (For details, please refer to section 3.1.1.4). Furthermore, the dissolved oxygen level measured by the sensor in real time can guide the sampling time accurately, which is hard to achieve by other ways. This feature of the YMC not only allows the biological repeats to be taken from different cycles, but also enables various datasets from different studies to be mapped to a uniform time axis (For details, please refer to section 3.2.9). As a result, it is possible to differentiate the temporal sequences of various changes in metabolism, epigenetics and gene expression, which may even give some indications of the potential causal relationship.

The YMC may also be a better system to study the impact of a gene deletion on the transcriptome than batch cultures, for two reasons. First, the cell cycle is largely synchronized in the YMC. It can reduce the averaging effect associated with the asynchronized cell cycle in

batch cultures. For example, if the transcriptomes of cells in different cell cycle stages respond to a gene deletion differently, the sequencing data from batch cultures will become an average of the transcriptomes in different cell cycle stages (Dolatabadi et al. 2017). Furthermore, if certain genes are only expressed at a particular cell cycle stage and a gene deletion can change the length of different cell cycle stages, the increased or decreased expression levels seen in batch cultures can be an artefact of changing the length of cell cycle stages (Marguerat et al. 2014). Because the cell cycle and cellular metabolic states are largely synchronized during the YMC, the averaging effect is relatively small, comparing to batch cultures. Although there are other methods to synchronize cells, such as using drugs to arrest the cell cycle or changing the medium, these treatments can cause cellular stress responses, which may be undesirable in many experiments. Second, by taking multiple samples across the YMC and comparing the wild-type and the deletion mutant, the changes in the transcript levels of individual genes can be located to specific time windows during the YMC. The time window may be associated with certain changes in cellular metabolites and histone modifications. Due to the availability of various other YMC datasets (metabolome, ChIP-seq, MNase-seq (Nocetti & Whitehouse 2016) and NET-seq), the impact of a gene deletion can be studied in relation to various molecular events.

1.9 Using the YMC to study the relationship between transcription and transcripts

The YMC could potentially also be a system to study the relationship between transcription and the levels of transcripts. It is already known that the transcript levels during the YMC are highly dynamic and can change dramatically in a short period of time (Kuang et al. 2014). The cellular mRNA levels can be regulated by various mechanisms, including transcription and mRNA turnover. A study on the gene expression changes in dendritic cells after LPS (lipopolysaccharide) stimulation showed that transcription is the dominant factor that influences the transcript levels of the majority of genes (Rabani et al. 2014). Another study on

S.pombe suggested that the mRNA turnover can work in concert with transcription to achieve steeper changes in mRNA levels (Marguerat et al. 2014). Recently, it was also found that RNA degradation by the exosome plays an important role in the downregulation of transcripts during the differentiation process of embryonic stem cells (Lloret-Llinares et al. 2018). Currently, it is still not known how different mechanisms work together to shape the temporal profiles of transcripts during the YMC.

The mRNA turnover rate normally cannot be measured directly. To answer this question, it is necessary to measure the activities of transcription or the levels of nascent RNA during the YMC by sequencing. Currently, there are three major strategies to study nascent RNA. They are the nucleotide labelling methods, the RNAPII co-IP methods and the direct isolation methods. Their major differences lie in the purification of the nascent RNA. The nucleotide labelling methods, such as 4SU-seq, PRO-seq and GRO-seq, need to incubate the cells or the nuclei with the modified nucleotides (Core et al. 2008; H. Kwak et al. 2013; Rabani et al. 2011). Once the nucleotides are incorporated into the nascent RNA by RNA polymerases, the nascent RNA can be purified in the later procedures by pulling down the modified nucleotides. In terms of RNAPII co-IP, NET-seq is the most typical method (Nojima et al. 2015; Churchman & Weissman 2011). It uses an antibody to pull down RNAPII, which can co-precipitate the nascent RNA. The direct isolation methods, such as nascent-seq, use various methods to isolate the chromatin and extract the chromatin associated RNA, which is believed to be mostly nascent RNA (Khodor et al. 2011).

Between these three strategies, the RNAPII co-IP method, which is NET-seq, is more applicable to the YMC. The direct isolation methods require the purification of the chromatin. As budding yeast cells have the cell wall, which has to be digested by enzymes first, it takes a relatively long time (at least 15 mins, based on the protocol in Mellor lab) to isolate the chromatin. Considering the rapid changes in transcript levels during the YMC, the nascent RNA isolated

after 15 mins is unlikely to reflect the original transcription levels at the time point of sampling. The nucleotide labelling methods also cannot be applied, because most of these methods require the cellular nucleotide synthesis enzymes to convert the labelled precursors (such as 4-thiouracil and 4-thiouridine) into the labelled nucleotides. It is already known that the expression of the genes coding those enzymes are highly temporal specific during the YMC (Kuang et al. 2014). Hence, the levels of nascent RNA measured by the nucleotide labelling methods are very likely to be biased at certain time points during the YMC. Although PRO-seq uses the biotin labelled nucleotides directly, it requires cell permeabilization or nuclear isolation steps, which has the same problem as the direct isolation methods (Mahat et al. 2016). The sampling procedure of yeast NET-seq is quick (within 3mins). Furthermore, by pulling down RNAPII with epitope tags, the transcription levels can be measured in a relatively unbiased manner. Hence, NET-seq is more suitable than other methods for studying transcription during the YMC.

However, before studying the relationship between transcription and transcripts during the YMC, there are two things that still need to be improved. First, the original protocol of yeast NET-seq has a relatively high non-specific background. It can complicate the analysis, because the background is mainly mature RNA and can change periodically during the YMC. Actually, a preliminary experiment has been performed on the YMC by using the original protocol of NET-seq (Please refer to appendix section S1-3). Due to the presence of the high background, it was hard to obtain conclusive results for this research topic. Hence, the current protocol of yeast NET-seq needs to be improved, in order to reduce the background. Second, there are still many unannotated transcription units in the yeast genome. Without knowledge of their transcription start sites (TSS), it is difficult to quantify their expression levels and perform other analysis. Although it is possible to ignore these unannotated transcription units and focus only on annotated ones, including unannotated transcription units in the analysis may bring a more comprehensive view of the transcription changes during the YMC.

1.10 The two aims of this thesis

This study has two major aims. The first one is to use the YMC system to study the relationship between acetyl-CoA metabolism, histone acetylation and gene expression. Currently, it is known that histones can be acetylated in response to an increase in the cellular levels of acetyl-CoA. Histone acetylation is associated with gene expression and many chromatin remodeling factors recruited by histone acetylation can regulate gene expression. There are three remaining issues. First, it is still not entirely clear how histone acetylation is linked to gene expression. Second, it is not known whether chromatin remodeling factors are involved in the interplay between acetyl-CoA metabolism, histone acetylation and gene expression. Third, it is not clear whether the gene expression associated with histone acetylation can regulate cell metabolism. In other words, is there a reciprocal regulation between cell metabolism and gene expression, which is bridged by acetyl-CoA and histone acetylation? The background knowledge of these questions will be further introduced in the context of the YMC system in section 3.1.2. The experiments for answering these questions are mainly covered by the third chapter of this thesis.

The second aim of this study is to pave the way for using the YMC system to study the relationship between transcription and transcript levels by developing two techniques. The first technique is called ultra-low background NET-seq. It can solve the high background issue of the NET-seq. The second one is called size-fractionated NET-seq. It uses a completely different strategy to map the positions of TSS in a sensitive, reliable and relatively unbiased manner. It has determined the positions of more than 9000 TSS. More than one-third of them have been missed by other TSS mapping techniques. The two techniques will be covered by the fourth chapter of this thesis.

Chapter 2.

Materials and methods

2 Materials and methods

2.1 Strain construction

All the yeast strains used in this study are listed in Table 2.1. The strain construction was based on homologous recombination and involves three major steps, making DNA constructs, yeast transformation and strain verification.

Name in this study	Origin	Genotype	Nutrient requirement
wild-type	From Euroscarf	CENPK113-7D MATa, URA3, HIS3, LEU2, TRP1, <i>mal2-8c</i> , SUC2	Prototroph
<i>RPB3-3xFLAG</i>	wild-type	<i>RPB3-3xFLAG::KanMX6</i>	Prototroph
<i>sth1-BDΔ</i>	wild-type	<i>sth1Δ1273-1359::KanMX6</i>	Prototroph
<i>rsc4-Y225AY226A</i>	wild-type	<i>rsc4-Y225AY226A::KanMX6</i>	Prototroph
<i>rsc2-BD1Δ</i>	wild-type	<i>rsc2Δ17-131::KanMX6</i>	Prototroph
<i>rsc2-BD2Δ</i>	wild-type	<i>rsc2Δ282-378::KanMX6</i>	Prototroph
<i>rsc2-BAHΔ</i>	wild-type	<i>rsc2Δ431-536::KanMX6</i>	Prototroph
<i>rsc2-CT1Δ</i>	wild-type	<i>rsc2Δ541-646::KanMX6</i>	Prototroph
<i>rsc1Δ</i>	<i>RPB3-3FLAG</i>	<i>rsc1::hphMX6</i>	Prototroph
<i>rsc2Δ</i>	<i>RPB3-3FLAG</i>	<i>rsc2::hphMX6</i>	Prototroph
<i>aco2Δ</i>	wild-type	<i>aco2::KanMX6</i>	Prototroph

Table 2.1 A list of yeast strains used in this study

The Longtine method (one-step PCR) was followed to make the DNA construct for full length deletion mutant and epitope tagged strains (Longtine et al. 1998). Both forward and reverse primers are comprised of two regions. The 5' side has a 50nts gene specific sequence. For gene deletion, it maps to the sequence immediately upstream of the start codon or downstream of the stop codon. For epitope tagging, it maps to both sides of the stop codon. To make the DNA construct for point mutations or domain deletions, the original Longtine method was modified by using two-step PCR instead of one-step. The workflow is illustrated in Figure 2.1. In the 1st PCR, the point mutation or domain deletion was introduced to the target gene by primer F1. Primer F2 has one end mapping to the end of the intergenic region and the other end mapping to the promoter of the marker. After the 1st PCR, the product was separated on 1% agarose gel and recovered by Monarch DNA gel extraction kit (T1020L, NEB). The purified 1st PCR product was then combined with the normal Longtine plasmid carrying

the appropriate selection marker and primer pairs (F1 and F3) to start the 2nd PCR. Primer F3 has its one end mapping to the end of the marker gene, while the other end of F3 maps to the intergenic region right after the target gene. In this study, it is essential to have F3 and F2 mapping to the upstream and downstream positions of the intergenic region respectively. The reason is that the PAS of RSC2 is close to the TSS of its downstream gene and the current arrangement of F2 and F3 can avoid truncating either the terminator of RSC2 or the promoter of its downstream gene by duplicating the intergenic region after homologous recombination. After the 2nd PCR reaction, the product was purified in the same way as the 1st PCR purification. Then the DNA sequence of the purified product was then checked by Sanger sequencing provided by Source Bioscience. After confirming there was no mutation except the designed mutation site, the PCR product was used in the yeast transformation. For a comprehensive list of primers used in this study, please refer to table 2.2.

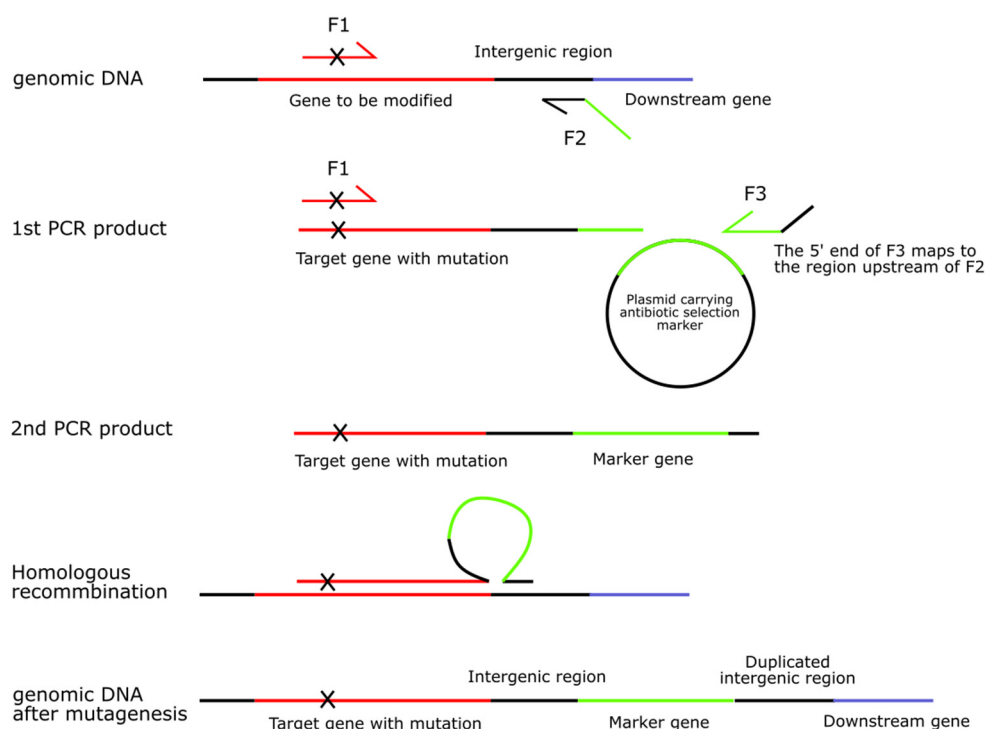


Figure 2.1 The modified workflow of the Longtine method for introducing point mutations. The intergenic region was duplicated after transformation, which could avoid disturbing the promoter of the downstream gene. Red, green and blue colors represent the target gene, the downstream gene and the selection marker respectively. F1, F2 and F3 are primers for PCR reactions. The same color indicates the sequence homology. The cross represents the point mutation.

Table 2.2 A list of primers used in this study

Name	Sequence (from 5' to 3')
Rpb3-CTD(tag)-F	AATGGGTAATACTGGATCAGGAGGGTATGATAATGCTTGGCGGATCCCCGGGTTAATTAA
Rpb3-CTD(tag)-R	TTTCGGTTCGTTCACTTGTTTTTTCTCTATTACGCCGAATTCGAGCTCGTTTAAAC
Rpb3-(Chk)1F	GTGCAAAGGCTTCGTACCAG
Rpb3-(Chk)1R	TCGGTTCGTTCACTTGTTTTT
STH1 bromo deletion F	AATTGGTGGAAGAAATGCGCGAACAGTTAGATGAGGTAGACTCCCATCCATAACGTTTTTCATCACCACATGTTT
STH1 bromo deletion R	GAGCTGCGCACGTCAAGACTGTCAAGGAGGGTATTCTGGGCCTCCATGTCTTGCGGCACTAATAAGCCAGAG A
STH1 HpH tagging R	GAATCGCAAACGGTCGAAGCTAACTGGTCCATCGCCGCCGGGAATTATTGAATTGAGCTCGTTTAAAC
STH1 checking F	CTAAACTAAATCCAAGTCGCTTGGGA
STH1 checking R	AGCCAAAGGATATTAGTTACTCAAAGA
RSC4 K25A F	GAAAGGCCTAAATACTTGCCGGGAGCACATCTAAAAACCAAGAGAAAACT
RSC4 K92A93A F	CCAAGTAGGAAGTTTCATCCACAGGCTGCTTACAAGATCCAGCAGCCAATGTCT
RSC4 K225A226A F	GTCGATAAAGATGAACCTCCTGAAGCTGCTGAAATAGTGCATAGCCCCATGGCT
RSC4 bromodomain mutation R	GAGCTGCGCACGTCAAGACTGTCAAGGAGGGTATTCTGGGCCTCCATGTCAGTACTTTTTCTTTACGGACGCA ACT
RSC4 HpH tagging R	GCCATATCCTTACATACATCTTTCCCTACATAACTAACCTACCCATTTTATGAATTGAGCTCGTTTAAAC
RSC4 chk F	TAGAGAGCAAACCAAGAATAATAGAT
RSC4 chk R	GCTCAATATCATTATCAAGAGCAGATA
RSC4 del3 F	GAAACGGAATTCATGAATTTCTGGATAAATTGATGACTATCTCTCTTCATAGTCTTCC
RSC4 del4 F	CATGAAACGGAATTCATGAATTTCTGGATATGATGACTATCTCTCTTCATAGTCTTCC
RSC2 BD2 deletion	TAAAGCGTGGGAGACCCCAATCATCGACTTACCGTATATTCAAAGAATGGCAAGATTGAACTATCCAAACCT
RSC2 BD1 deletion	GCATGATGCCTGATGACAATTCAAACTCGTCCACTCAAACTCAAGCGCAGATTTGGGTCCTCTACCCGATGAA
RSC2_BD1_purer	GCATGATGCCTGATGACAATTCAAACTCGTCCACTCAAACTCAAGCGCAGATTTGGGTCCTCTACCCGATGAA
RSC2_BD1_chk2	GAAATTAAGAGAAAAAGCTGAGGAAG
RSC2_BD1_chk_R	TAGTGGACCTGTTGTCTCTTAATG
RSC2_C_termi_chk	AGGTCATCTATGTCTTTAGAAAACC
RSC2 BAH deletion	ATTGGGCACTTCTTCGAAATCAAATGATCCTCAAAAACCTATAGTGGGCCGTGATTGGACGAAGCTACCATT
RSC2 CT deletion	AGATTAGAACCTGGAAGGCTTGCTTACCAGAGGAAATTCGTGATTTGGACTACTCAAGTAGTAAACTATCCAG TTTAAGA
RSC2_marker_tagging	GAGCTGCGCACGTCAAGACTGTCAAGGAGGGTATTCTGGGCCTCCATGTCGTCCTTGGGGAAGAAATTTT
RSC2 deletion Reverse	GGAAGATATTATGCTGCCATTGCTTTTACAATAAAGGTGAGAATTCGAGCTCGTTTAAAC
RSC2_chk_F2	GGATGCTGATTTTCCTGATTGAT
RSC2_chk_R2	TTGTAACCTTTGCAATTGGTGTT
CLN3 deletion F	TCTCCCTTTACTCTGTTCAAGACACTGATTTGATACGCTTTCTGTACGCGACATGGAGGCCAGAATA
CLN3 deletion R	TTTGCAAATTTAATTTATTTGTTTAAATGCATTTTTTTTGTGCTTGAATTCGAGCTCGTTTAAAC
CLN3_chk_F	TGCTCTCACTGTAATGATCA
CLN3_chk_R	TTAGATAGAGAGGGGCAGAT
RSC2 FLAG tagging F	AAGTTCACGGCGCACAGACTCTCTATGCTGCGGCCTCCTTCGTCGCTTCagactacaaggacgacgatgac
RSC2 FLAG tagging R	aaggattacaagatgacgacgataaggactataaggacgatgatgataaaTGATCACCTTTATTGTAAAAGCAATG
RSC2 full length chk R	GTAGCAACTTACTGGAAAAAG
RSC2 FLAG tagging chk	TAAGACCCTCGGCTAAGTTC
ACO2_deletion_F	TTCGACTTTCTTTAAACATAAACTCTAAGAGCAATAAGCAAGAGATTCGGGATCCCCGGGTTAATTAA
ACO2_deletion_R	TATGACTATTAAATCGGACTCTTAATTTTTAGCTAATATCTCTCAGAATTGAATTCGAGCTCGTTTAAAC
ACO2 F chk	TTATTCGATTTGCGCACTTG
ACO2 R chk	TTTGCACCGTCTATATGCAG
RSC1 del new F	CTAGCTACAAAACAGAGATAAAAAAATTATATTTCAGCCGACATGGAGGCCAGAATA
RSC1 deletion R	ATTCTCGTAATCTGAATGCAAGGTATTTCAGCGGAA GAATTGAATTCGAGCTCGTTTAAAC
RSC1 checking F	CTTACCCTCTTCCGTCAAT
RSC1 checking R	AAAACGATTGAGGTCTATCG
RSC2 del new F	AGAACCAGACGAAGCGGAGAATATTCTACATTGAC AGTGCCGACATGGAGGCCAGAATA
RSC2 deletion R	GGAAGATATTATGCTGCCATTGCTTTTACAATAAAG GTGAGAATTCGAGCTCGTTTAAAC
RSC2 checking F	AGTTCAATACGTGATCAAAAT
RSC2 checking R	GTCCTTGGGGAAGAAATTTT

To carry out transformation, the yeast strain to be transformed was inoculated in 5ml of YPD and cultured overnight at 30°C. On the next day, 250ul of the overnight culture was diluted in 25ml of YPD and grown for 6 hours at 30°C. Cells were harvested by spinning down at 3000rpm for 5mins at room temperature. After a brief wash with PBS, cells were suspended in 450ul of 0.1M sterile LiAc/TE and left at 4°C overnight. On the following day, 100ul of the cell suspension was mixed with 10ul of calf thymus DNA, 10ul of the purified PCR products and 700ul of 0.1M LiAc/TE with 40% PEG. The mixture was incubated at 30°C for 30mins. After adding 50ul of DMSO, the mixture was heat shocked at 42°C for 20mins. Cells were collected by spinning at 7000rpm for 2mins at room temperature. After washing cells once with autoclaved miliQ water, cells were suspended in 1ml of YPD and incubated at 30°C for 3 hours. After that, cells were collected by spinning at 7000rpm for 1min. The cell pellet was suspended in 100ul of fresh YPD and plated onto either Kanamycin or hygromycin B selection plates. Depending on the growth rates of the strain and the mutant, colonies normally started to appear after 3 days, if the transformation was successful.

To screen for positive colonies, the colonies on the selection plates were restreaked onto a new selection plate. Some cells from each colony were also grown in 1ml of YPD in an Eppendorf tube overnight. Genomic DNA was extracted from each of the 1ml culture by non-phenol method (please refer to section 2.5.2). PCR was performed on each genomic DNA preparation with primers flanking the deleted, inserted or mutated region. For full length deletion, domain deletion or tagging, a certain size shift of PCR products could be observed, if the colony was correct. For point mutation, the PCR product for each colony was first purified with Monarch PCR&DNA cleanup kit (T1030S, NEB). If there was a restriction site at the mutated position, the PCR product could be digested with the appropriate restriction enzyme and examined by running a 1% agarose gel. A differential fragmentation pattern could be observed for correct colonies. If there was no suitable restriction site, the PCR product had to be sequenced with Sanger sequencing to find out the correct colonies.

For gene tagging, western blot with the corresponding primary anti-epitope antibody was performed to further validate the positive colonies. For domain deletion and point mutation, Sanger sequencing was carried out to examine the DNA sequence of the whole gene, in order to make sure there was no unexpected mutation. After all the processes above, there were still multiple correct colonies. The growth rate of each colony was measured by growing cells in 25ml of YPD and monitoring OD₆₀₀ over time. The colonies with growth rates close to wild-type could be suspended in glycerol/YPD solution and stored at -80°C.

2.2 Operating the fermenter

Before the start of a new run, all major parts of the fermenter vessel need to be disassembled, washed with MiliQ water thoroughly and autoclaved. After that, 900ml of YMC medium was prepared in a 1L bottle without adding antifoam 204. For the full recipe, please refer to table 2.3. Then the 900ml of YMC medium was poured into the fermenter vessel and 0.5ml of antifoam 204 was added directly to the vessel as well via a 1ml syringe. Once the pH probe was calibrated with pH standard and the dissolved oxygen probe had its electrolyte replaced, they were installed onto the fermenter vessel. Then the vessel was autoclaved again. After autoclave, the condenser should be installed onto the vessel in a flow hood. Then the vessel was left in room temperature to cool down. Once its temperature was below 60°C, all the probes and pipes were connected to the controller. The DO probe should be polarized for at least 7 hours by connecting to the controller before use. On the same day, the 10ml pre-culture was inoculated by transferring roughly 200million cells from the YPDA plate to 10ml of YPD media. The FALCON tube containing the 10ml medium should be closed tightly and wrapped with parafilm.

On the next day, the water cooler and compressor were first switched on. The temperature of the medium in the fermenter was adjusted to 30°C and the pH was raised to above 3.50 by switching on the base pump which adds NaOH stock solution (10g/L) to the fermenter.

Aeration and agitation were maintained at 1L/min and 1000rpm respectively. The pre-culture was checked under the microscope to make sure there was no bacteria contamination. The OD₆₀₀ of the pre-culture was also measured and recorded. After that, the 10ml pre-culture was injected into the fermenter by a syringe with a 19G needle through the inoculation membrane. The base pump was switched to the mode of 'base cascade' to correct the pH of the medium, while the feeding pump was switched off. The collection pump was set to 'level 2', in order to maintain a constant culture volume.

After inoculation, the cells in the fermenter kept growing until it used up all the glucose in the medium. Then cells entered the starvation phase in which the level of dissolved oxygen in the medium was close to saturation. In principle, the cells should be starved for at least 7 hours to scavenge any residual carbon sources. In this study, the starvation time was roughly 14 hours. Although longer starvation time is also allowed, it is not recommended to starve the cells for more than 48 hours. After starvation, the feeding pump was switched on and set at a constant rate of 1.5ml/min. At this point, the OD₆₀₀ of the cell culture was about 10units per ml. After 10 hours, the OD₆₀₀ should reach 16units per ml. The metabolic cycle should normally appear 6 hours after turning on the feeding pump.

There are two types of sampling, the small scale (1ml) and the large scale (100ml). The two types of sampling require different preparation procedures and are used for different purposes. The 1ml small scale sampling does not disturb the metabolic cycle, so multiple samples can be taken constitutively within the same metabolic cycle. It can be used for western blot, RNA-seq, FACS, metabolome, proteome and other biochemical assays. To carry out the small scale sampling, 5ml of cell culture was taken from the fermenter by a 10ml lure-lock syringe via the sampling valve. The 5ml of cell culture was first transferred to a 15ml FALCON tube. It was then further transferred to several Eppendorf tubes with a pipette. Each Eppendorf tube should contain 1ml of cell culture. Then the tubes were centrifuged at room

temperature by pressing the quick spin button for exactly 10 secs. The supernatant was quickly removed by a pipette. The cell pellet was snap frozen in liquid nitrogen. The time of snap freezing should be recorded as the sampling time. The whole procedure is expected to be completed within 3mins.

The large scale (100ml) sampling normally disturbs the metabolic cycle, so only one sample should be taken in a metabolic cycle. It is advised to leave a time interval of at least 8 hours between two large scale sampling events for the cells to recover. The large scale sampling is intended for NET-seq, which requires a large amount of starting material. To carry out this type of sampling, the glass filter unit should be assembled before starting. For details of using the glass filter, please refer to the NET-seq protocol for yeast published by Stirling Churchman in 2011. A 50ml lure-lock syringe was used to take 100ml of cell culture from the fermenter. The cell culture was immediately filtered through the nitrocellulose membrane. A spatula was used to scrape cells off the membrane. After that, the cell pellet was snap frozen by liquid nitrogen. The time of snap freezing should be recorded as the sampling time. 100ml of cell culture from the fermenter normally contains 1.4g of cells, which is enough for one NET-seq experiment. In order to capture the original transcriptional status of the cells within the fermenter, the whole process should be completed within 2.5mins.

2.3 Next generation sequencing

2.3.1 Nascent elongating transcripts sequencing (conventional protocol)

In the first step of NET-seq, frozen *S.cerevisiae* cell pellets were mixed with *S.pombe* cell pellets and grounded into fine powder together in the mixer mill. Before starting the experiment, the grinding chambers, the spatulas and any other equipment that may directly contact the sample need to be pre-chilled in a clean polystyrene box containing liquid nitrogen. Then both *S.cerevisiae* cell pellets and *S.pombe* cell pellets were transferred from -80°C freezer to the same polystyrene box. Next, the weight of the empty grinding chamber was

measured by a balance. A piece of polystyrene was placed in between the balance and the empty grinding chamber to avoid heating up the grinding chamber. In the following steps, the chamber should be kept on the balance. *S.cerevisiae* cell pellets were first added into the chamber and the changes in weight should be taken down. The amount of *S.pombe* to be added was calculated from the weight of *S.cerevisiae* divided by a fixed ratio. In this study, the ratio of *cerevisiae* to *pombe* was 2.5 to 1. After adding *S.pombe*, the changes in weight should be taken down again. Adding slightly more or less *S.pombe* was not a problem for normalization, as long as the weights for both *S.cerevisiae* and *S.pombe* were taken down accurately. Once the weighing step was completed, the two grinding chambers were closed and shaken in the mixer mill. The program for shaking was 15Hz for 3mins. Then the grinding chamber was placed in liquid nitrogen for 2mins to maintain the low temperature. The above process was repeated for another 5 times. Fine cell powder was transferred to two pre-chilled clean 50ml falcon tubes, A and B. Tube A contained 1.40g of cell powder, while tube B contained the rest of cell powder. Tube A can either be stored at -80°C or used for cell lysis directly. Tube B can be saved for other applications. All the weighing steps need to be done as fast as possible, as water vapour may condense on the surface of the grinding chamber and slightly change the weight.

In the second step of NET-seq, cell power was clarified by lysis buffer. RNAPII and its associated RNA was isolated from the cell lysate by immunoprecipitation. Normally two samples were processed in parallel. For most of the steps, the original NET-seq protocol was followed (Churchman and weissman, 2011). Briefly, the NET-seq lysis buffer was prepared first and incubated on ice. The anti-FLAG M2 agarose beads (A2220-10ML, Sigma) were washed twice with the ice cold NET-seq lysis buffer. Next, 12.5ul of SUPERase.In (AM2696, ThermoFisher Scientific) and 660ul of RQ1 RNase-free DNaseI (M6101, Promega) was added to 5ml of ice cold NET-seq lysis buffer. 1.40g of cell powder taken directly from the -80°C freezer or liquid nitrogen should be mixed with the lysis buffer above immediately by vigorous shaking and/or

pipetting up and down. Once there were no obvious clumps seen, the lysate of each sample was transferred to 7 Eppendorf tubes. The whole procedure should be mostly carried out on ice. The total incubation time was 20mins, counting from the beginning of mixing the cell powder and the lysis buffer. After 20mins, the 7 Eppendorf tubes were centrifuged at 16,000g for 10mins at 4°C. The supernatant was combined and incubated with the washed agarose affinity beads in a 15ml Falcon tube at 4°C with rotation for 2.5 hours. After that, the suspension of beads was centrifuged at 1000g, 4°C for 2mins (this setting was used for all the following centrifugation steps). The supernatant was decanted carefully. The beads were washed 4 times, each with 10ml of NET-seq wash buffer followed by centrifugation. After washing, RNAPII was eluted from the beads by incubating the beads with NET-seq elution buffer at 4°C for 30mins with rotation. Next, the elution buffer was collected and the elution step was repeated for one more time. After that, the elution buffer was combined. RNA was purified from the eluate with miRNeasy mini kit (217004, Qiagen). In the last step of the kit, RNA was eluted from the column by 30ul of MiliQ water. Nanodrop 2000 was used to quantify the amount of RNA purified. If the sample comes from the YMC, it is expected to yield 1.5~4.5 ug of RNA. If the sample is from batch cultures, normally 3-9ug of RNA is expected.

The third step is library construction, in which the RNA undergoes linker-1 ligation, alkaline fragmentation, reverse transcription, circularization and PCR amplification sequentially. The typical amount of starting material was 3ug of RNA purified from the last step. However, it can also go down to 0.3ug. Most procedures were exactly the same as Churchman 2011 paper. Here, most library construction procedures are briefly described. Some key or modified procedures are described in more detail. To perform linker-1 ligation, three 1.5ml non-stick RNase-free tubes were used. Each tube contained 10ul of heat denatured RNA (100ng/ul, dissolved in 10mM Tris pH 7.0), 4.8ul of 50% PEG (mol. Wt. 8000), 1.2ul of MiliQ water, 2ul of RNA ligase buffer (B0216L, NEB), 1ul of linker-1 (S1315S, NEB) and 1ul T4 RNA ligase 2 truncated (M0242L, NEB). If the total amount of starting material is lower than 3ug, at least

make sure the volume of RNA solution is 10ul in each tube. As the presence of PEG makes the solution become viscous, a vigorous vortex is needed to fully mix all the reagents. After incubating the mixture at 37°C for 3 hours, 0.7ul of 0.5M EDTA was added to each tube to quench the reaction. To fragment the RNA into the desired size range, 20ul of alkaline fragmentation buffer was added to each tube. Then the mixture was incubated at 95°C for 40mins (the exact time needs to be calibrated for each batch of alkaline fragmentation buffer). After fragmentation, RNA was purified by isopropanol precipitation, followed by size selection on 10% TBE-urea polyacrylamide gel. The excised gel piece was mashed with a 1ml pipette tip in a 1.5 ml Eppendorf tube. 600ul of MiliQ water was added to the mashed gel piece, followed by incubation at 70°C for 10mins with vigorous shaking. The supernatant was collected by Costar Spin-X column. RNA was recovered by isopropanol precipitation for 2 hours in -20°C. RNA pellet was dissolved in 10ul of 10mM Tris.Cl (pH 7.0). After reverse transcription, the cDNA was size selected again with 10% TBE-urea polyacrylamide gel and recovered in the same way as just described. Then cDNA was circularized by CirLigase (CL4111K, epicentre). To obtain the final product for sequencing, PCR amplification was performed on the circularized cDNA. A range of PCR cycle numbers should be tested (normally 3-7 cycles). The PCR product was size selected on an 8% TBE gel. Under the optimal PCR cycle number, the PCR product should give a bright smear without any high-molecular weight products on the 8% TBE gel. The PCR product with the optimal cycle number was recovered by soaking the mashed gel pieces in DNA soaking buffer overnight. PCR products can then be precipitated by isopropanol precipitation from the supernatant. Finally, the PCR product was dissolved in 10ul of 10mM Tris.Cl (pH 8.0) and was ready for illumina sequencing.

The sequencing was performed on NextSeq 500 by the Biopolymer Facility of Harvard University. The setting for sequencing was high throughput flow cell, 50 cycles and single-read. For the data processing of NET-seq, please refer to section 2.6.1.

2.3.2 Nascent elongating transcripts sequencing (ultra-low background protocol)

This is a new protocol developed in this study. It can greatly reduce the background of NET-seq. However, because of the lowered background, the yield of RNA after immunoprecipitation is also reduced. Extra care must be given when performing the downstream steps. In this new protocol, the agarose beads were washed for eight more times with the stringent wash buffer before elution. Specifically, after the four normal washes, agarose affinity beads were transferred to a COSTAR spin-x column and centrifuged at 4°C, 1000g for 1min to get rid of residual normal wash buffer. To make the stringent wash buffer, RNase-free urea was supplemented to the normal wash buffer to make a final concentration of 1M. 300ul of the stringent wash buffer was added to the column containing beads. It was then incubated at 4°C for 10 mins with gentle rotation. Next, the column was spun at 4°C, 1000g for 1 min to remove the stringent wash buffer. The two steps above were repeated for 7 additional times. After that, RNAPII can be eluted from the beads as described in the conventional protocol. miRNeasy Micro kit (217084, Qiagen) was used to extract RNA from eluate. In the last step, RNA was eluted from the column with 10ul of MiliQ water. The expected RNA yield of the 1.40g batch culture cell powder is around 0.6ug. Since the yield was lower than that of the conventional protocol, one tube was used instead of three in the linker ligation step. All the following steps were the same as the conventional protocol except that additional two PCR cycles may be required in the final PCR amplification step.

Note that sodium chloride cannot be used to increase the stringency of the normal wash buffer, because dissolved sodium chloride can dissociate RNAPII and result in the failure of immunoprecipitation.

2.3.3 Size fractionated NET-seq

Size fractionated NET-seq, developed in this study, is a new technique of mapping TSS on the genome. Its procedures are described in this section. For its mechanism, please refer to section 4.2.2.1. All the steps before linker-1 ligation were performed in the same way as normal NET-seq (the conventional protocol).

For size fractionated NET-seq, it is recommended to use 6ug of RNA as the starting material for linker-1 ligation, although 3ug of RNA is also possible. If 6ug of RNA is used, six non-stick RNase-free tubes should be used instead of three to carry out the ligation reaction. The recipe for ligation reaction is exactly the same as normal NET-seq (the conventional protocol). After 3 hours incubation at 37°C, the reaction was quenched by adding 0.7ul of 0.5M EDTA to each tube.

After that, the ligation product was purified and concentrated by isopropanol precipitation, instead of undergoing alkaline fragmentation directly. Specifically, 0.56ml of RNA precipitation solution and 0.75ml of isopropanol were added to each tube. After vortex, the mixture was incubated at -20°C for 2 hours, followed by centrifugation at 4°C, 20,000g for 30mins. Next, the supernatant was removed without disturbing the RNA pellet. The pellet was then washed once with 0.75ml of 80% ice-cold ethanol. After brief centrifugation, residual ethanol was carefully removed by pipetting. The pellet was air-dried at room temperature for 10mins. 15ul of 10mM Tris.Cl (pH 7.0) was used to dissolve all the RNA pellets in the six tubes.

The next step is to perform the size selection. The 15ul of RNA solution was mixed with 15ul of 2x TBU denaturing loading buffer and heat denatured at 80°C for 3 mins. Heat denatured 10bp, 50bp, 100bp and 1kb DNA ladder were loaded onto the 3.5% TBU polyacrylamide gel first (for the procedures of casting the gel, please refer to Note 1 at the end the current section). The RNA sample was evenly split into 3 wells of the gel (10ul in each well). It is advised to leave two empty wells between the ladders and the sample, in order to avoid

contamination. The 3.5% TBU polyacrylamide gel was assembled into the BioRad electrophoresis device. The tank of electrophoresis device was filled with ice-cold 1xTBE running buffer. The whole device was then incubated in ice. After setting up the device, the electrophoresis was run at 80V for 80mins. Once the electrophoresis was completed, the gel can be stained for 5mins in 50ml of 1xTBE buffer containing 5ul of SYBR Gold nucleic acid gel stain (S11494, ThermoFisher Scientific), followed by destaining in 1xTBE buffer for another 5mins. After that, the gel was frozen at -20°C for 15mins. As 3.5% polyacrylamide gel is floppy and fragile, freezing the gel can make it easier to cut. In this study, five size ranges were selected. They were 40-100nts, 100-400nts, 400-1Knts, 1K-3Knts and above 3Knts, according to the DNA ladder. Other size ranges can also be used, however it is advised to keep at least two small fractions, e.g. 40-100nts and 100-400nts.

To recover the RNA from the gel, electroelution was used. Gel pieces were first mashed in the same way as described in ordinary NET-seq protocol (please refer to section 2.3.1). The mashed gel from each size fraction was loaded into separate electroelution devices (For how to make this device, please refer to note 3 at the end of this section). All the electroelution devices were placed in the electrophoresis tank. After filling the tank with 1xTBE buffer, electrophoresis was performed at 80V. To avoid heating up the buffer, the whole electrophoresis device was placed on wet ice (need more details). In the meantime, the position of fluorescence should be monitored every 10mins with the transilluminator. Once the fluorescence moved from the gel to one end of the electroelution device, the electroelution device was taken out immediately. All the buffer in the electroelution device should be collected. The time required to electroelute RNA depends heavily on the size of RNA. For the smallest size fraction, 10mins is normally sufficient, while the largest RNA fragment may need as long as 2 hours. After electroelution, RNA from each size fraction was normally dissolved in 600ul of 10mM Tris.Cl (pH 7.0). To purify and concentrate RNA, 2ul of Glycoblue, 50ul of 3M sodium acetate (pH 5.5) and 0.75ml of isopropanol were added. The following

procedures were the same as the isopropanol precipitation described before. In the end, RNA pellet of the smallest fraction was dissolved in 10ul of 10mM Tris.Cl (pH 7.0). All the other size fractions were dissolved in 20ul of the same buffer.

RNA from each fraction is treated as separated samples and is ready to be alkaline fragmented. As it is not necessary to fragment the smallest fraction, it can proceed directly to reverse transcription. For the rest of fractions, the same alkaline fragmentation time can be used. All the following library construction procedures are the same as the ordinary NET-seq protocol. In the final PCR amplification step, two more PCR cycles might be used to get the products for sequencing. For the data processing of size fractionated NET-seq library, please refer to section 2.6.4 and section 4.2.2.1.

Note 1: Casting 3.5% TBU polyacrylamide gel

To minimize the contamination of RNase, all the glass plates for gel casting were first wiped with 1% SDS and washed thoroughly with MiliQ water. The cleaned glass plates should be kept for RNA purpose only. To make the gel solution, the following recipe was followed.

For 10ml of gel solution (sufficient for casting one gel):

Urea	4.8g
30% acrylamide solution	1.17ml
10% APS	50uL
10X UltraPure TBE buffer	1ml
MiliQ water	added to make the total volume become 10ml
TEMED	10uL (added before casting)

The gel should be left in room temperature for at least 1 hour to solidify. After that, the gel is ready to be used or can be stored at 4°C overnight. Ideally, the gel should be used within 2 days. The 3.5% gel is extremely fragile and must be handled carefully. Due to the presence of 8M urea, 3.5% is the lowest percentage that can be achieved with the protocol above.

Note 2: Pre-run polyacrylamide gel

The TBU gel should be pre-run for 15mins without loading samples. Before loading, a pipette with a clean tip should be used to flush the wells with 1XTBE buffer several times.

Note 3: Making the electroelution device

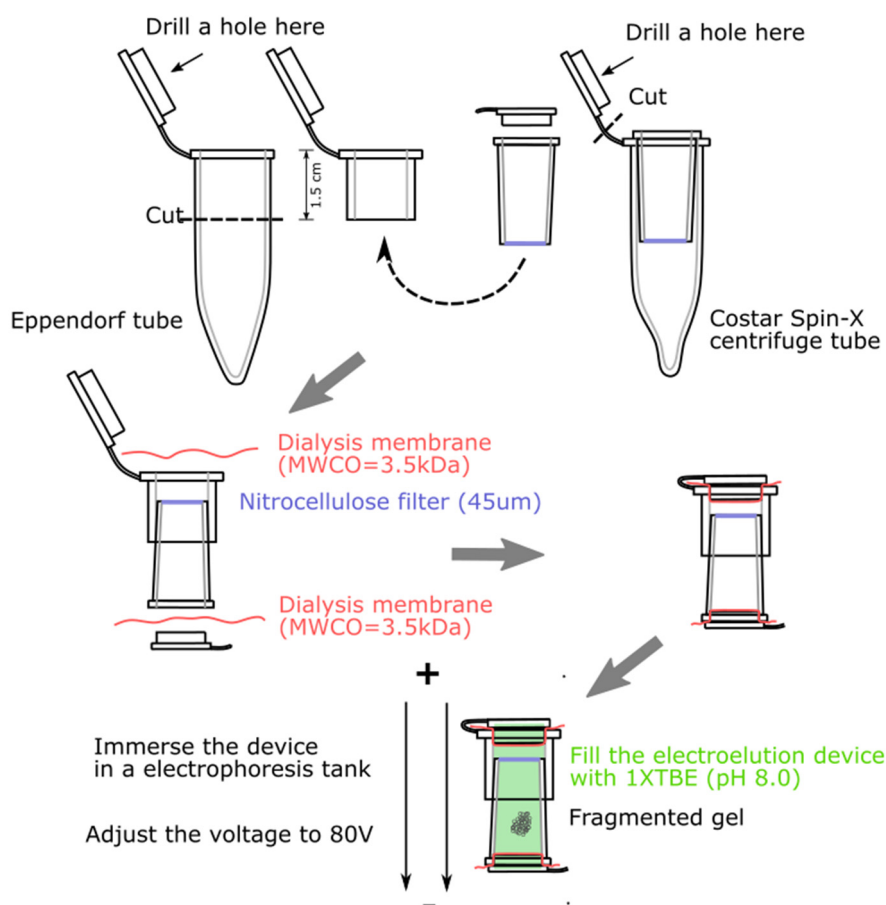


Figure 2.2 The steps of making an electroelution device from an Eppendorf tube and a Costar Spin-X tube for the size fractionated NET-seq. The electroelution allows the efficient recovery of long RNA molecules from polyacrylamide gel. The dialysis membrane and the nitrocellulose filter membrane are colored red and blue respectively. The TBE buffer is shown in green.

The diagram for making the electroelution device is shown in Figure 2.2. A hole with a diameter of 7mm was first drilled on the cap of both Eppendorf tube and Costar Spin-X centrifuge tube respectively. The Eppendorf tube was cut into two at the position shown on the diagram. Its top half along with the cap was retained. For the Costar Spin-X column, its inner tube as well as the cap were kept. Then the inner column was inverted and inserted into the top half of the Eppendorf tube as shown in the middle left of Figure 2.2. After that, two pieces of dialysis membrane with a MWCO (molecular weight cut off) of 3.5 kDa were excised

and soaked in 1xTBE. Before assembling the device, the chamber of the device was filled with 1xTBE and the fragmented polyacrylamide gel was loading into the inner tube from the Costar Spin-X column. After that, both ends of the device were covered with the two pieces of dialysis membrane and the two caps were attached to both ends to fix the dialysis membranes. Finally, the electroelution was placed in an electrophoresis tank with 1xTBE. The buffer should be added until the dialysis membrane is completely immersed.

2.3.4 mRNA 3'end sequencing

This is a method for quantifying the level of mRNA in the sample. tRNA, rRNA and mitochondrial RNA cannot be studied by this method. Briefly, mRNA is reverse transcribed by poly-T primer and the complementary strand of cDNA is synthesised by random priming.

First of all, 1ml of culture from the fermenter was centrifuged at room temperature by pressing the quick spin button for 10 secs. After removing the supernatant, the cell pellet was snap-frozen by liquid nitrogen. The cell pellet can either be stored at -80°C or proceeded to the next step straightway. In the following procedures, the sample tubes were kept in dry ice. To add RNA spikein, an aliquot (Note 4) of frozen *S.pombe* cell pellet (containing 180 OD₆₀₀ units of *S.pombe* cells harvested from a chemostat) was quickly suspended in 12.78ml of ice cold TE buffer (10mM Tris.HCl pH 7.5, 10mM EDTA). 200ul of the suspension was quickly transferred to each of the sample tubes. If there is a large number of samples to be spiked in, the *S.pombe* cell suspension should be resuspended after adding every 5 samples, in order to make sure the suspension is homogenous. After adding spikein to all the samples, 200ul of TES buffer (10mM Tris.HCl pH 7.5, 10mM EDTA, 1%SDS) and 400ul of Acid-Phenol-Chloroform (AM9722, ThermoFisher) were added to each sample sequentially. The sample mixture was transferred to a thermomixer and incubated for 20mins at 65°C with vortexing at 1400RMP for 30 secs every 5mins. After that, it was stored in -80°C freezer for 30mins followed by centrifugation at 13,000 RMP for 15mins at room temperature. To perform ethanol

precipitation, the supernatant was mixed with 600ul of freshly made NaOAc/EtOH solution (2.5ml of 3M NaOAc pH5.2 mixed with 47.5ml of 100% EtOH) and stored at -80°C for 30mins. After centrifugation at 13,000rpm for 20mins at 4°C, the supernatant was removed and the RNA pellet was washed with 70% of ethanol. After air dry, the pellet was dissolved in 400ul of water. The ethanol precipitation procedure was repeated to further improve purity. Finally, the air dried pellet was dissolved in 100ul of MiliQ water. The expected total RNA yield was around 200ug with 260/280 ratio higher than 2.4 and 260/230 ratio higher than 2.2.

The purified RNA can be used for library prep. To construct the library, QuantSeq 3' mRNA-Seq library prep kit for Ion Torrent (012.24A, LEXOGEN) was used and its associated protocol was strictly followed. Once the library is made, its quality and concentration can be checked with bioanalyzer. The sequencing was performed on Ion Proton. With *S.pombe* spikein, it is advised to pool 12 mRNA 3' end sequencing samples on the same sequencing chip, in order to balance the sequencing depth and cost. For more detailed instruction, please refer to the manual of Ion Proton sequencer. The data processing of mRNA 3' end sequencing is described in detail in section 2.6.1.

Note: One such aliquot of *S.pombe* is sufficient for spiking in 48 samples. The final OD₆₀₀ ratio of *S.cerevisiae* to *S.pombe* is roughly 6.5 to 1. Normally, the final ratio of reads between the sample and *S.pombe* spikein should be around 2.5 to 1. It is possible to use a higher ratio.

2.4 Proteome

It is advised to perform all the procedures on ice. 70 million cells were first resuspended in 1ml of ice cold proteome lysis buffer. For buffer recipe, please refer to table 2.3. The suspension was transferred to a 2ml tube with a screw cap. 500ul of acid washed glass beads was added to the 2ml tube. Once all the screw caps were tightened, all the 2ml tubes were transferred to a Magna Lyser bead miller and shaken for 1min at 6000 rpm, followed by incubation in ice for 3mins. It was repeated for 7 more times. After that, the tubes were

centrifuged at 4°C for 10mins at 16,000g to remove cell debris. The supernatant was transferred to a new tube and stored at -80°C. The protein concentration was measured by the BCA assay. For more details of the assay, please refer to the manual of Pierce BCA protein assay kit (23225, ThermoFisher). The mass spectrometry analysis of the samples was performed by Benedikt Kessler Lab at the Target Discovery Institute (TDI) in University of Oxford. Briefly, 100ul from each of the 10 samples were taken and proteins were precipitated by methanol-chloroform method. The proteins were digested overnight by trypsin. After digestion, the peptides were further labelled with TMT labels. After pooling all 10 samples together, the pooled sample was run on the mass spectrometry instrument separately.

2.5 Biochemical assays

2.5.1 Western blot

A yeast cell pellet containing roughly 290 million cells was quickly mixed with 300ul of 8M urea, 240ul of western blot loading buffer and 100ul of acid-washed glass beads. If a large number of samples need to be processed, it is advised to perform the procedure above on ice. Then the mixture was vortexed for 3mins, followed by incubation at 95°C for 5mins. After centrifugation at 16,000g for 10mins at room temperature, the supernatant was transferred to a new Eppendorf tube. 10ul of the supernatant was saved to carry out the BCA assay. The rest was mixed with 60ul of 1M DTT and 10ul of 3% bromophenol blue stock solution, followed by 5mins incubation at 95°C. The loading volume of each sample should be calculated according to the measured protein concentration of each sample, in order to make sure the same amount of protein is loaded in each well. For the recipe of hand-made polyacrylamide gel, please refer to table 2.3. After loading the gel, the electrophoresis was first performed at 80V for 1 hour, followed by 160V until the marker reached the bottom of the gel. Then the separated protein was transferred to nitrocellulose membrane by the semi-dry method. The electrophoresis of semi-dry transfer was run at 15V for 45mins. On the next day, the

nitrocellulose membrane was blocked with 5% BSA in TBST for 1 hour. Then the membrane was incubated in 2.5% BSA/TBST with primary antibody for 1 hour. After 3 quick washes (1mins each) and 3 long washes (10 mins each), the membrane was incubated with secondary antibody in 2.5% BSA/TBST for 45mins. The washing procedures were repeated. After incubating the membrane in Pico ECL mix at room temperature for 3mins, the membrane can be exposed to a film in a dark room. Normally, the exposure time of 10secs, 1min, 5mins and overnight can be used. For checking RNAPII pull down in NET-seq experiment, 1sec exposure time was used, as the signal for RNAPII was normally extremely strong.

2.5.2 Non-phenol genomic DNA extraction

The cells were collected from the 1ml overnight culture by spinning down at 13,000rpm for 1min. Then the cell pellet was suspended in 100ul of 200mM LiOAc/1%SDS and incubated at 65°C for 5 mins. After mixing with 300ul of 100% EtOH, the mixture was centrifuged at 13,000rpm for 3mins. The supernatant was discarded and the pellet was washed with 100ul of 70% EtOH, followed by centrifugation at 13,000rpm for 1min. After removing the supernatant, the pellet was suspended in 100ul of MiliQ water and incubated at 65°C for 5 mins. After centrifugation at 13,000rpm for 15 secs, the supernatant was collected and was ready for downstream applications.

2.6 Bioinformatics

2.6.1 The processing of the raw sequencing data

The sequencing of NET-seq libraries was performed on the illumina NEXTseq 500 sequencer with the high throughput flow cell, single end reads and 50 cycles. With the original protocol, NET-seq normally requires 100 million reads to obtain a good coverage on each *S.pombe* spiked *S.cerevisiae* sample. With the new protocol, 50 million reads are enough. The adapter sequence 'ATCTCGTATGCCGTCTTCTGCTTG' was first trimmed from the raw reads by

running cutadapt. Then bowtie was utilized to align the reads to the hybrid genome of *S.pombe* and *S.cerevisiae* with the following settings: '-q -p 16 -S -m 1 -n 1 -e 70 -l 28 -k 1 --best --phred33-quals'. The uniquely aligned sam files were eventually converted into the bedgraph files by samtools, bamtobed and genomecov.

The sequencing of RNA-seq libraries was performed on the Ion Proton sequencer with Ion PI Chip v3. The 3' end sequencing of *S.pombe* spiked *S.cerevisiae* sample requires 6 million reads. The raw reads were aligned by STAR without using cutadapt. The settings for STAR were: '--runThreadN 2', '--alignIntronMax 2500' and '--outSJfilterReads Unique'. All the downstream data format conversions for RNA-seq were performed the same as those for NET-seq.

2.6.2 Pearson correlation, K-means clustering, statistical tests and plotting

metagene

The Pearson correlation coefficient was calculated by using the MATLAB function 'corr'. The K-means clustering was performed by using the MATLAB function 'kmeans' with 16 clusters, 1000 maximal iterations and 500 replicates. The Kolmogorov-Smirnov test and the single sided Fisher's test were carried out by using the MATLAB functions 'kstest2' and 'fishertest' respectively. Metagenes were plotted by following the procedures described by Churchman and Weissman 2011.

2.6.3 Mapping time points to the wild-type time axis

The Smoothing spline fitting was first performed on the 22 wild-type RNA-seq data points for each gene by the MATLAB function 'fit'. Specifically, the 'fittype' parameter was specified as 'smoothingspline' and the smoothing parameter was set at 0.166. The function returned a fit object. Using time as the input, the fit object returned the estimated expression data at the time point during the YMC. Once all the genes were fit to smoothing spline curves. The transcriptome was estimated at any time points during the YMC. The fit objects for all the

genes in the wild-type were sampled every 1mins during the YMC, which produced 180 estimated wild-type transcriptomes. The Pearson correlation coefficients between the 180 estimated wild-type transcriptomes and RNA-seq data to be mapped were calculated, which gave 180 Pearson correlation coefficients on the time axis of the wild-type. The same smoothing spline fitting was applied to the 180 coefficients. By using MATLAB function 'findpeaks', the position of the peak in the curve was determined. The peak position was the estimated time point on the wild-type time axis. All the RNA-seq samples including the mutants were mapped in the same way. The NET-seq and ChIP-seq samples have their companion RNA-seq samples, which allows NET-seq and ChIP-seq data to be mapped as well.

2.6.4 The dry lab part of size-fractionated NET-seq

The normal NET-seq alignment method should be used. For fraction I, which was the smallest size fraction, the reads trimming by 'genomecov' was performed on either the 3' ends and the 5' ends of reads, which produced two sets of bedgraph files for fraction I. The bedgraph files for track A, B, C and D were imported into MATLAB and smoothed by zero-phase digital filtering algorithm 'filtfilt'. Specifically, the digital filter was first set up by using function 'designfilt' with the following parameters: 'lowpassiir', 'FilterOrder', 12, 'HalfPowerFrequency', 0.05, 'DesignMethod', 'butter'. It returned a digital filter object, which was used as one of the inputs of the 'filtfilt' function together with the data on track A, B, C and D. The output data were the smoothed data for track A, B, C and D. The TIF-seq annotation was then utilized to generate a metagene for each of the smoothed track A, B, C and D. The data of the first 400bps downstream of TSS of the four metagenes were combined, which was the reference pattern. At each nucleotide position on the genome, the data of the 400bps downstream of that position for all four data track were combined as well, which was the pattern to be tested. The Pearson correlation coefficient was calculated between the reference pattern and the pattern to be tested. As a result, all positions on the genome had a Pearson correlation coefficient,

which was plotted as a curve. The x-axis of the curve was the position on the genome. The position of the peaks on the curve was called by using 'findpeaks' with the minimal peak height of 0.34 and the minimal peak distance of 50bps. The positions of the peaks were the positions of the TSS. A TSS was filtered if the total read number in its downstream 100bp region of all four data tracks was lower than 10. The mapped TSS were compared to several existing annotations. A TSS was assigned to the existing annotation, if their distance was smaller than 70bps.

2.6.5 Motif enrichment analysis

The motif enrichment analysis was performed on the MEME server version 5.01. Specifically, the differential enrichment mode was used. The size distribution was set as 'zero or one occurrence per sequence'. The number of motif to find was set at 5. '0-order model of sequences' was chosen as the background model. The width of motifs was set between 6 and 50bps. Tick the box for searching given strand only but do not tick the box for searching palindromes only.

2.6.6 Gene ontology (GO) analysis

The GO analysis was performed on the GORILLA server (<http://cbl-gorilla.cs.technion.ac.il/>). The organism was *Saccharomyces cerevisiae*. The running mode was 'Two unranked lists of genes'. The gene names in each cluster were pasted into the box of 'Target set'. The gene names of all 5579 genes were pasted into the box of 'Background set'. All the ontology terms, including 'Process', 'Function' and 'Component' were chosen. All the other settings were set as default.

Chapter 3.

Investigating the role of the RSC complex during the YMC

3 Investigating the role of the RSC complex during the YMC

3.1 Introduction

3.1.1 An introduction to the YMC system

3.1.1.1 The basics of the fermenter

In this study, the fermenter was the device that provided the chemostatic growth conditions required for maintaining the YMC. Specifically, prototrophic *S.cerevisiae* cells were grown at a high cell density in the culture vessel of the fermenter. Under the control of the fermenter, the fresh medium containing 1% of glucose was added to the vessel dropwise at a constant speed. Within the vessel, the rapid rotation of the stirrer ensured quick and efficient mixing. At the same time, the fermenter also maintained the constant pH, temperature, aeration and volume of the cell culture. These growth conditions for maintaining the YMC are referred to as the YMC conditions in the later discussions.

The general components of the fermenter are depicted in figure 3.1. It is composed of two major parts, which are the culture vessel and the control unit. The culture vessel has several sensors to monitor cellular growth conditions. The readings from these sensors are sent to the control unit. The control unit has pre-set values for pH, temperature and the culture volume. That is, when the sensors detect deviations from the pre-set values, the control unit will activate the corresponding effectors to correct the deviations. The detailed mechanisms for maintaining the constant pH, temperature, aeration and culture volume are explained below.

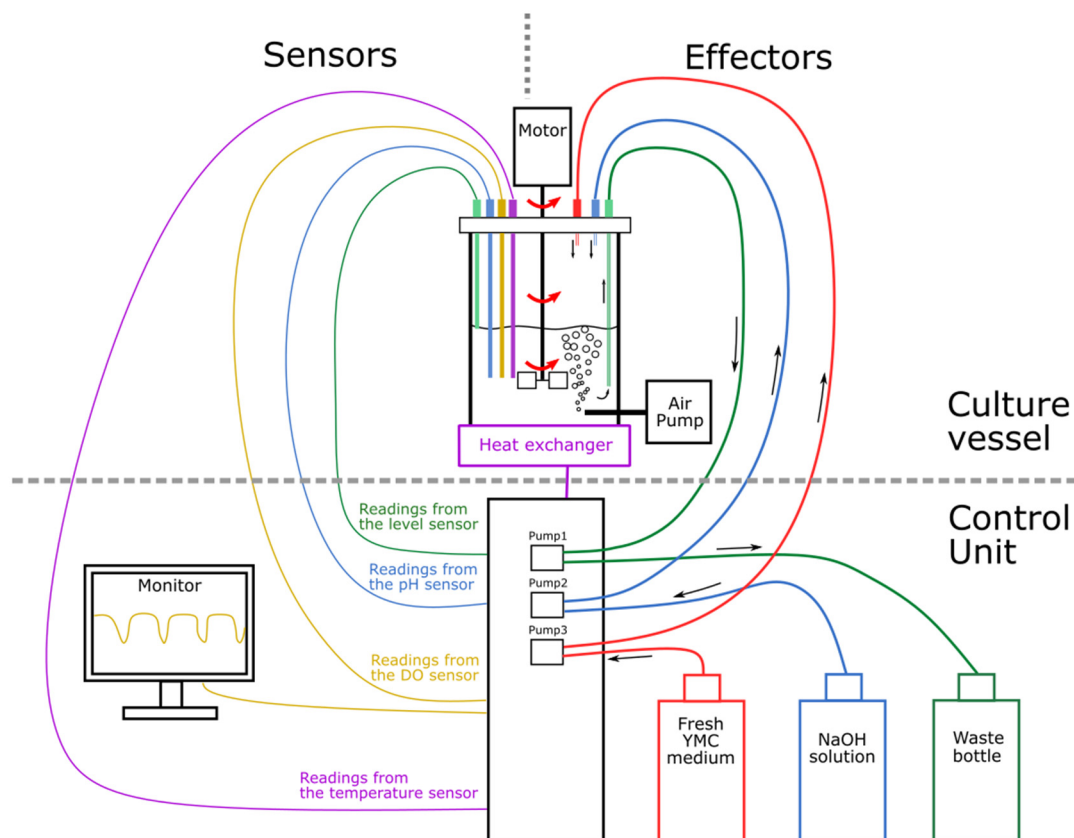


Figure 3.1 A schematic diagram of the fermenter. The fermenter is composed of two parts, the culture vessel and the control unit. The culture vessel is installed both the sensors (green for the level sensor, blue for the pH sensor and purple for the temperature sensor) and the effectors. The sensors detect changes in the medium and send the readings to the control unit, which instructs the effectors to counteract those changes. In this study, the pH, temperature and culture volume are maintained constant through the feedback loop above.

The cells were grown at a constant pH of 3.5. However, the pH of the medium in the vessel always had a tendency to drop, mainly because of two reasons. First, the fresh medium being constantly added to the vessel had a pH of around 2.7 (The low pH can inhibit bacteria growth in the feeding bottle). Second, the cells produced metabolites that could acidify the medium during the course of the YMC. To maintain a constant pH of 3.5, the pH of the medium was monitored by the pH sensor. If the pH fell below 3.5, the corresponding pump was activated to add the NaOH solution to the vessel. As the effect of acidification caused by the addition of the fresh medium was constant, the changes in the activity of the base pump were mostly caused by the production rate of the acidic metabolites by the cells. The temperature and the

cell culture volume were maintained in a similar way. Any decrease or increase in the temperature was monitored by the temperature probe, which instructed the heat exchanger to heat up or cool down the cell culture. For the culture volume, as the fresh medium and the NaOH solution were added to the vessel, the liquid volume in the vessel tended to increase. Once it reached the tip of the liquid level sensor, the corresponding pump was turned on to remove the cell culture from the vessel, which suppressed any further increase in the cell culture volume.

The dissolved oxygen (DO) sensor monitored the DO levels in the medium throughout the time. The DO levels could be converted to the cellular oxygen consumption rates by performing calibrations (For details, please refer to section 3.1.1.3). In this study, the oxygen consumption rate was the major indicator for different metabolic states during the YMC.

3.1.1.2 The changes in the DO levels of the medium during a typical fermenter run

During the typical course of a fermenter run, the cell growth will experience three stages, which can be indicated by the changes in the DO readings (Figure 3.2). The first stage starts from inoculating the cells from the overnight pre-culture. In this stage, the cell population in the vessel experiences rapid growth and the cell density keeps increasing, while the DO values decrease. Typically, the cells will use up the majority of the glucose for aerobic metabolism after 18 hours, which correlates with the sudden increases in the DO values of the culture medium. At this point, the cell density normally has 10 units of OD₆₀₀. Then the cell population enters the second phase, in which the cell population is starved and only has a low level of aerobic metabolism to use up the residual carbon sources. Theoretically, as long as the fresh YMC medium is not added to the cell culture, the cell population will stay in this stage. In practice, the cells are left in this stage for at least 14 hours but not longer than 48 hours. Once the feeding pump is turned on, the cells enter the third stage, which is the chemostatic stage. The cell density will keep increasing at the beginning and eventually reaches the saturation

level, which is about 16 units of OD₆₀₀. Some mutants may have a slightly lower saturation level. Normally, the YMC starts to appear after the OD₆₀₀ unit reaches 13. Once the YMC appears, its frequency and amplitude will gradually stabilize.

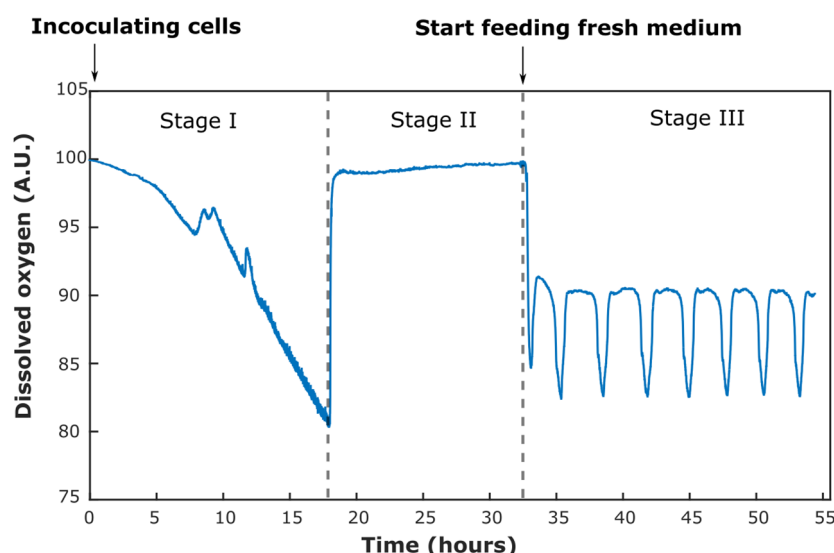


Figure 3.2 The changes in the dissolved oxygen (DO) of the medium during a typical fermenter run of the wild-type. This figure shows the raw readings from the DO sensor. It is also directly displayed on the monitor of the fermenter during the experiment. Lower readings of DO indicate higher oxygen consumption rates of the cells. The data will be converted to the level of oxygen consumption rate in the later analysis. The growth of cells can be divided into three stages. Stage I starts from inoculating cells and ends when the majority of glucose is used up. Once the cells are fed with the fresh glucose containing medium, they enter the chemostatic stage (stage III) and exhibit the yeast metabolic cycle.

3.1.1.3 The DO level measured by the DO sensor can be converted to the cellular oxygen consumption rate by performing calibration

The DO (dissolved oxygen) sensor measures the amount of oxygen dissolved in the medium and uses the DO value as the direct output. The DO value is the percentage of saturation of oxygen in the culturing media. The cellular metabolic state should be indicated by the cellular oxygen consumption rate. It can be converted from the DO value by performing calibration of the DO probe, which will be explained below.

First of all, there exists the following equation at any time points during the YMC:

The change rate of the DO level = oxygen dissolution rate – oxygen consumption rate of the cell population ----- Equation 1

Equation 1 can be rearranged as below:

Oxygen consumption rate of the cell population = oxygen dissolution rate - the change rate of the DO level ----- Equation 2

The oxygen dissolution rate is mainly affected by three factors, the air flow rate, the rotation speed of the stirrer and the current DO level. The former two factors are both set at fixed values throughout the course of the YMC. Hence, the oxygen dissolution rate becomes a function of the current DO level. Governed by the basic physical rules, the higher the current DO level is, the lower the oxygen dissolution rate will be. The relationship between the current DO level and the oxygen dissolution rate can be directly calibrated by examining the DO curve of the culture medium in the absence of cells. In this case, Equation 2 will become as below:

Oxygen dissolution rate = the change rate of the DO level ----- Equation 3

To perform the calibration, the fermenter was first aerated with pure nitrogen gas to remove all the dissolved oxygen, which made the DO value become zero. Then the gas supply was switched to air and all parameters were set as they were during the course of the YMC. It produced a curve of the DO levels (Figure 3.3a). At each point of the curve, the slope represents the oxygen dissolution rate at that particular DO level. Then the relationship between the oxygen dissolution rate and the DO level can be plotted as a new curve in figure 3.3b. Hence, the cellular oxygen consumption rate at any time points during the YMC can be deduced by combining this relationship and equation 2. In the following sections, all the DO readings have already been converted to the oxygen consumption rates.

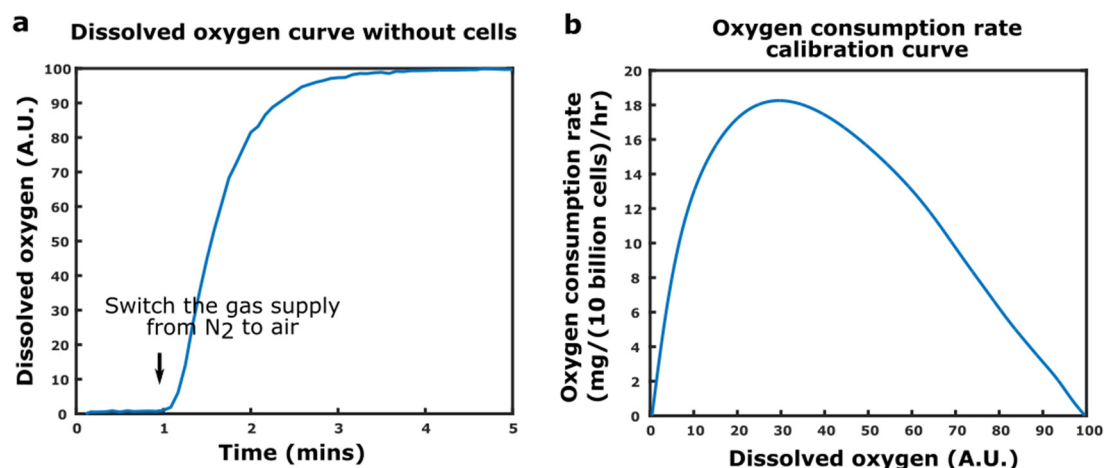


Figure 3.3 The raw dissolved oxygen values can be converted into the levels of oxygen consumption rate by performing the calibration. **a.** The gas supply was switched from nitrogen to air, in order to obtain the relationship between the oxygen dissolution rate and the dissolved oxygen values. **b.** The curve shows the functional relationship between the y-value and the slope at every point of figure a.

3.1.1.4 The typical oxygen consumption profile of the YMC

The typical periodic changes in the DO of the medium during the YMC is shown in Figure 3.4a. After calibration, the DO values of the medium were converted into the cellular oxygen consumption rates (Figure 3.4b, note the change in the unit of the Y-axis). The characteristic changes in the cellular oxygen consumption rate were revealed by overlaying multiple cycles of the YMC of the wild-type strain (Figure 3.4c). The YMC can be first divided into two phases, the high oxygen consumption (HOC) phase (red) and the low oxygen consumption (LOC) phase (blue). This study mainly focuses on the HOC phase. Upon entering the HOC phase, the oxygen consumption rate keeps increasing at the beginning. When a specific level is reached, the oxygen consumption rate is held at that level for 3 mins before it further increases. This feature appears in all the YMC cycles in different studies (Tu et al. 2005; Kuang et al. 2014; Blank et al. 2009; Henry et al. 2010). A number of changes in metabolism, epigenetics and gene expression can happen around this particular time point. Its biological significance is further demonstrated in this chapter. For simplicity, it is referred to as “point B” in all the following discussions. After passing point B, the oxygen consumption rate further increases,

until it reaches the maximal point (point A hereafter). Then it starts to decrease and hits the lowest point (point C hereafter) before it re-enters the LOC phase. To simplify the discussion in the following sections, the HOC phase is further divided into three phases (HOC-1, HOC-2 and HOC-3). Point B is the boundary between HOC-1 and HOC-2. Point D separates HOC-2 and HOC-3. It has the same oxygen consumption rate as point B. Although these phases are separated based on the oxygen consumption rates, the RNA-seq data shown in section 3.2.9 and 3.2.10 agree with this phase separation.

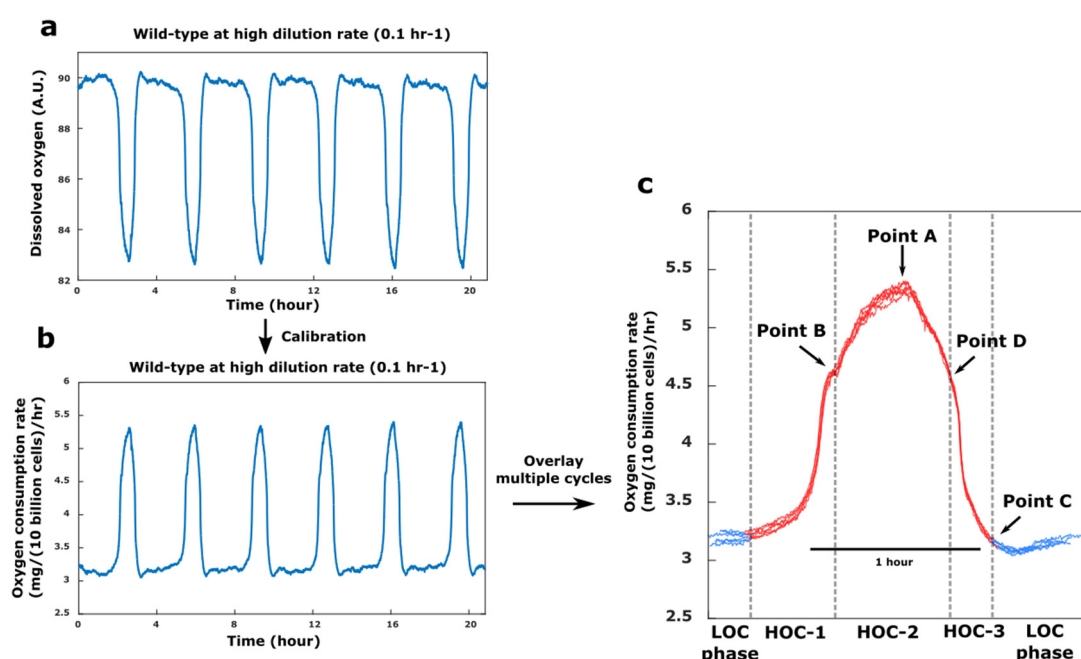


Figure 3.4 The workflow of obtaining the overlaid curve of the yeast metabolic cycle (YMC). **a.** The raw readings from the dissolved oxygen (DO) sensor. **b.** The raw DO values were converted into the oxygen consumption rate by the calibration. (Note the difference in the y-axis). **c.** After overlaying multiple cycles, the YMC can be divided into four phases, LOC (low oxygen consumption, blue color), HOC-1, HOC-2 and HOC-3 (high oxygen consumption, red color). Some critical points on the curve are named with A, B, C and D.

3.1.1.5 Cells in the LOC phase can be induced to enter the HOC phase

When the cells are in the LOC phase of the YMC, they can be induced to enter the HOC phase prematurely by many methods. The primary method is injecting extra carbon sources such as ethanol, acetaldehyde, acetate, lactose and glucose to the cell culture (Robertson et al. 2008; Murray et al. 2003). Acetate has been used on both the wild-type and the *gcn5Δ* strain (Cai et al. 2011). Although the *gcn5Δ* strain is unable to show the YMC and its curve of oxygen consumption rates is almost a flat line, injecting acetate still induced a HOC-like response. Apart from extra carbon sources, some chemicals seem to have similar effects. For example, either injecting hydrogen peroxide or briefly aerating the cells with carbon monoxide can induce the cells to enter the HOC phase in the wild-type strain (Chen et al. 2007; Tu & McKnight 2009).

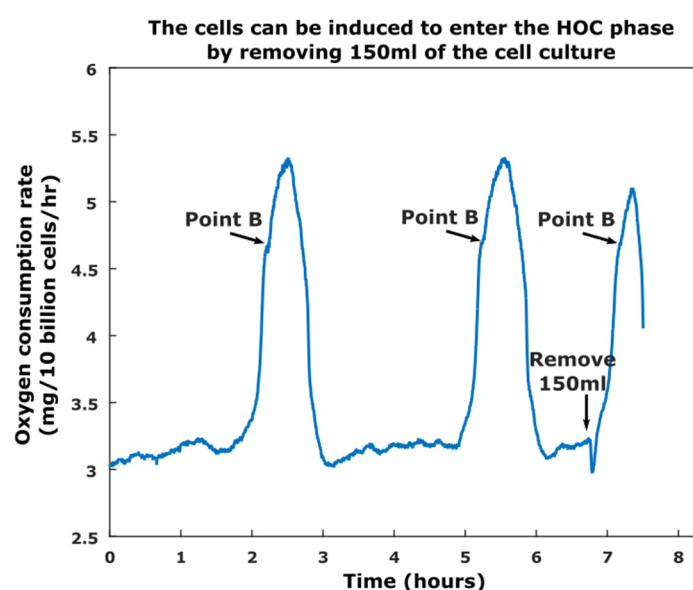


Figure 3.5 The cells can be induced to enter the HOC phase by removing 150ml of the culture

In this work, it was found that the cells can be induced to enter the HOC phase by suddenly removing 150ml (1/7 of the total volume) of the cell culture from the vessel. As is shown in Figure 3.5, after a transient decrease in the oxygen consumption rate, the cells enter the HOC phase of the YMC. The shape of the induced HOC phase is highly similar to the previous cycles.

Specifically, the point B of the induced HOC phase has at a similar level of oxygen consumption rate, while the point A of the induced HOC phase is lower.

Currently, it is not known why the remaining cells will enter the HOC phase by removing 150ml of the cell culture. One explanation is that this procedure suddenly reduces the culture volume but the rate of adding fresh glucose-containing YMC medium remains the same. It effectively increases the availability of carbon sources for the remaining cells, which might mimic the effect of injecting extra carbon sources. In this chapter, this method of inducing the HOC phase is utilized to study the mutants that do not exhibit the YMC.

3.1.2 An introduction to the research question

3.1.2.1 The current understanding of the relationship between acetyl-CoA, histone acetylation and gene expression during the HOC phase of the YMC

It is already known that the transcript levels of more than 60% of genes exhibit periodic oscillations during the YMC (Tu et al. 2005). The levels of various metabolites and histone modifications also show clear periodicities (Cai et al. 2011; Tu et al. 2007; Kuang et al. 2014). This feature of the YMC makes it become a good system to study the relationship between acetyl-CoA, histone acetylation and gene expression. This section will first summarize the current understanding of this topic in the context of the YMC and then introduce the unsolved issue as well as the hypothesis of this study.

First of all, the intracellular level of acetyl-CoA changes periodically during the YMC. Specifically, it starts to increase from the beginning of the HOC-1 phase, where the oxygen consumption rate of the cells also starts to rise. Then it reaches the maximum at point B. This phenomenon was initially found in the metabolome datasets across the YMC (Tu et al. 2007). Later, the isotope tracing experiment suggested that this peak of acetyl-CoA is produced by the increased intracellular levels of carbon sources at the beginning of the HOC-1 phase (Cai et al. 2011). The relevant metabolic pathways in this process are shown in Figure 3.6.

Interestingly, disrupting the production of acetyl-CoA by mutating *ACS1* and *ACS2* can cause the loss of the YMC (Cai et al. 2011).

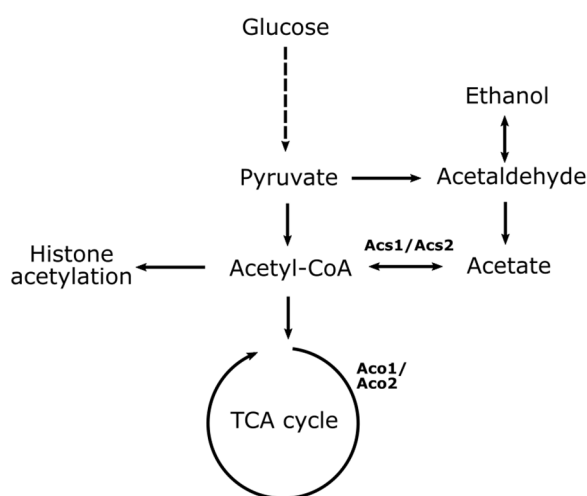


Figure 3.6 Acetate can be converted into acetyl-CoA, which is the substrate for histone acetylation.

The increased intracellular level of acetyl-CoA is associated with the acetylation of both histones and non-histone proteins. The overall acetylation levels on different histone H3 and H4 lysine residues were first shown by the western blot (Cai et al. 2011). Particularly, the levels of H3K14,18,23,27Ac are elevated from point B and remain high throughout the HOC-2 phase. H3K9 seems to be acetylated from point B and the level is further increased in the HOC-2 phase. The increases in the acetylation levels of histone H3 was found to be dependent on lysine acetyltransferase Gcn5, which is normally recruited to the promoter as a part of the SAGA complex (Cai et al. 2011). Apart from histones, several other subunits of the SAGA complex, including Ada3, Sgf73 and Spt7 were also shown to be acetylated (Cai et al. 2011). Notably, both the deletion of *GCN5* and the point mutation that inactivates the catalytic activities of Gcn5 can lead to the abortion of the YMC (Cai et al. 2011). Besides, the penta-lysine-to-arginine mutations (K9,14,18,23,27R) on histone H3 also causes the loss of the YMC (Kuang et al. 2014).

Histone H3 acetylation was found to be highly specific in terms of both its genomic localizations and timing. Furthermore, these acetylation marks have strong associations with gene expression. Initially, by performing ChIP-qPCR at *RPS11B*, *RPL33B* and *SHM2* loci, it was found that their promoter regions have much higher levels of histone H3 acetylation in the HOC phase of the YMC. The transcript levels of these genes also have a sharp peak during the HOC phase and the deletion of *GCN5* can disrupt that peak (Cai et al. 2011). Similar observations were also made on the G1/S cyclin gene, *CLN3* (Shi & Tu 2013). The histone H3 acetylation levels at the promoter region of *CLN3* was examined by ChIP-qPCR with a relatively high temporal resolution. It revealed that H3K9, 14, 18 and 23 all have a sharp peak at point B (Shi & Tu 2013). Later, the genome-wide changes in histone acetylation levels and gene expression were both investigated at a high temporal resolution during the YMC, which also allowed the measurement of the relative timing between histone H3 acetylation and gene expression (Kuang et al. 2014). Particularly, on growth related genes, including ribosomal protein genes, rRNA processing genes and amino acid synthesis genes, there are peaks of H3K9Ac and H3K14Ac appearing at point B. The levels of their corresponding transcripts peak few minutes after histone H3 acetylation. All these studies show that the peak of histone H3 acetylation has a strong association with gene expression.

However, it is still not known how the expression of genes can happen after the acetylation of histone H3 during the YMC. It was suggested that acetylation can neutralize the positive charges of the lysine residues on histone H3, which could loosen the interaction between DNA and histones and further induce gene expression. This study will examine a different possibility. That is, the acetylation marks on histone H3 can recruit a certain chromatin remodeler, which further activates the expression of genes. Luckily, in a preliminary experiment, deleting *RSC2*, which encodes a bromodomain containing subunit of the RSC chromatin remodelling complex, was found to cause the loss of the YMC, which suggests that the RSC complex may be a candidate of the chromatin remodeler mentioned above. Hence, this study will focus on the

RSC complex. In budding yeast, another two chromatin remodelling complexes, the SWR1 complex and the SWI/SNF complex, also have subunits containing bromodomains. Due to the limitation of time, they were not investigated in this study. Some future experiments about them are suggested in section 3.3.5.1 and 3.3.5.2.

3.1.2.2 The RSC complex

The RSC complex is a member of the SWI/SNF family of the chromatin remodelling complex. It harbours more than half of all the bromodomains in budding yeast and bromodomains can bind acetylated histones. Budding yeast is found to have 15 bromodomains among all its proteins, while the RSC complex possesses 8 of them (Zhang et al. 2010). The RSC complex has 17 subunits (Tang et al. 2010). Among them, 5 of them have bromodomains. Specifically, Rsc1, Rsc2 and Rsc4 each contain two bromodomains, while Rsc58 and Sth1 have a single bromodomain. The SWI/SNF chromatin remodeling complex, which is homologous to the RSC complex, also contains a single bromodomain in its Snf2 subunit (the catalytic subunit of the SWI/SNF complex). However, unlike the RSC complex, the SWI/SNF complex is not essential for budding yeast (Cairns 1998). Furthermore, the abundance of the RSC complex is ten times higher than that of the SWI/SNF complex (Cairns et al. 1996; Côté et al. 1994; Cairns et al. 1994).

The RSC complex can interact with acetylated histone H3. In an early study, the affinity of the RSC complex for nucleosomes was found to be greatly increased by tetra-acetylated histone H3 (at lysine 9, 14, 18 and 23 residues) but not by tetra-acetylated histone H4 (Chatterjee et al. 2011). Later, it was shown that the two bromodomains of yeast expressed Rsc4 can bind to H3K14Ac (Kasten et al. 2004; VanDemark, Kasten, Ferris, Heroux, Hill, Cairns, et al. 2007). Specifically, the second bromodomain have a higher affinity. A dot blot assay of bacterially expressed yeast bromodomains suggested that the bromodomains of different RSC subunits can bind acetylated histones with various affinities and specificities. The bromodomains of

Rsc1 and Rsc2 have weak affinities for histone H3 acetylation, while the bromodomains of Rsc4 can bind to acetylated histone H3, particularly H3K14Ac. The bromodomain of Sth1 can bind H3K115Ac, in addition to H3K14Ac (Zhang et al. 2010; Josling et al. 2012). The affinities of bromodomains in this study are generally lower than the previous studies. It might be related to the differences between protein expression methods.

The RSC complex can be recruited to the promoters of more than 700 genes and has been shown to be required for the expression of many genes (Ng et al. 2002). For example, the Rsc1 isoform of the RSC complex can activate *ATG8*, which is associated with autophagy during nitrogen starvation (Yu et al. 2015). The deletion of either *RSC1* or *RSC2* can impair the expression of certain sporulation specific genes (Bungard et al. 2004). Sth1 is the enzymatic subunit of the RSC complex. It consumes ATP and can activate gene expression by moving or evicting nucleosomes. In an early microarray experiment, the transcriptomes of the temperature sensitive mutant of Sth1 and the wild-type were compared. It was found that the RSC complex is required for the normal expression levels of mitochondrial related genes and carbohydrate metabolism genes (Tsuchiya et al. 2015). The deletome data from Holstege lab also shows that more than 1000 genes can be down-regulated by deleting *RSC1* or *RSC2* (Kemmeren et al. 2014).

To sum up, the RSC complex may have the potential to bind to the acetylated histone H3 at point B and further activate genes required for the HOC-2 phase of the YMC.

3.1.2.3 The structures and functions of Rsc1, Rsc2 and Rsc4

Among all the subunits of the RSC complex, only the Rsc1 and Rsc2 subunits are mutually exclusive. The presence of Rsc1 or Rsc2 in the RSC complex defines two different isoforms of the complex (Cairns et al. 1999). The two isoforms of the RSC complex have overlapping but non-identical functions. For example, the growth rates of the *rsc1Δ* and *rsc2Δ* strains can be similar or quite different, depending on the types of stress conditions (Cairns et al. 1999). The

two subunits cannot complement each other by an extra copy of the same gene in certain stress conditions (Chambers et al. 2012). Double deletion of *RSC1* and *RSC2* can cause lethality (Chambers et al. 2012).

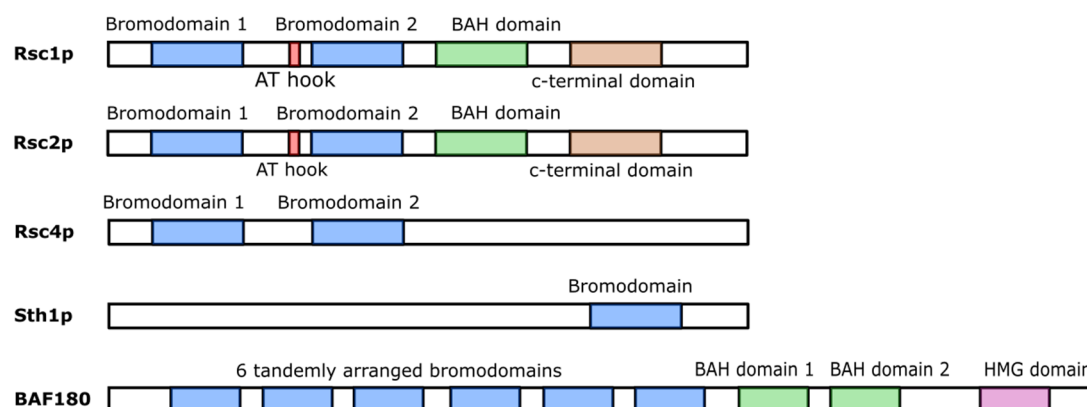


Figure 3.7 The schematic structures of several bromodomain containing proteins.

Rsc1 and Rsc2 have some similarities in terms of their amino acid sequences and domain arrangements (Cairns et al. 1999) (Figure 3.7). Both Rsc1 and Rsc2 have two bromodomains (BD1 and BD2), a BAH domain and two C-terminal domains (CT1 and CT2). An AT hook motif is found in between the bromodomains for both proteins. It preferably binds to the minor groove of DNA in the AT-rich regions. Some domains are essential for the functions of Rsc1 or Rsc2, while others are only required for certain functions (Cairns et al. 1999). For both Rsc1 and Rsc2, their BD2, BAH and the C-terminal domains are essential for all their functions tested (Cairns et al. 1999). BD1 is not required for Rsc1, but it is required for the function of Rsc2 in strains growing on the plate containing 15mM of caffeine, which impairs the cAMP signaling (Cairns et al. 1999). The AT hook is only required for Rsc1 but not Rsc2 (Cairns et al. 1999). Furthermore, the minimal construct required for cell viability is the presence of BD2+AThook+BAH+CT1+CT2 in either Rsc1 or Rsc2. For the AT hook, having one in either Rsc1 or Rsc2 is enough for cell viability (Cairns et al. 1999). Among all the domains, only the c-terminal domains are required for assembling Rsc1 or Rsc2 into the corresponding isoforms of the RSC complex (Cairns et al. 1999). The BAH domains of both Rsc1 and Rsc2 have affinities

for histone H3 (Chambers et al. 2013). This binding is independent of acetylation and only contribute partially to the overall binding affinity of Rsc1 and Rsc2 to the nucleosomes (Chambers et al. 2013). The swapping of the BAH domains between Rsc1 and Rsc2 seems to alter the functions of Rsc1 and Rsc2 (Chambers et al. 2012).

The Rsc4 subunit of the RSC complex has two bromodomains (BD1 and BD2) (Figure 3.7). Deleting either BD1 or BD2 is lethal. Both bromodomains can bind to H3K14Ac and BD2 has a stronger binding affinity. Such binding can be abolished by introducing the point mutation Y92F to BD1 or Y225F to BD2. The double point mutation is also lethal (Kasten et al. 2004).

Interestingly, Rsc1, Rsc2 and Rsc4 are homologous to the mammalian BAF180 (BRG1-associated factor 180)/PB1 (Protein polybromo-1) subunit of the BRG1/PBAF complex (Brownlee et al. 2012). BAF180 contains 6 bromodomains and 2 BAH domains (Figure 3.7).

3.1.2.4 The hypothesized acetylation-recruitment pathway and plans for testing the hypothesis

Based on the previous discussions, the following hypothesis is proposed. The RSC complex may play an important role in the YMC. At point B, the acetylation of the lysine residues on histone H3 may recruit the RSC complex to the promoters of certain genes via the bromodomains of the RSC complex. The recruited RSC complex could then activate the expression of these genes. The activated genes may further upregulate the cellular aerobic respiration in the HOC-2 phase. For simplicity, this pathway will be referred to as the acetylation-recruitment pathway (Figure 3.8). Although this pathway is not directly reported for the RSC complex, an early study suggested that the acetylation of histones is both a prerequisite and a determinant of the following nucleosome remodeling by the SWI/SNF complex at the gene *PHO8* (Reinke et al. 2001).

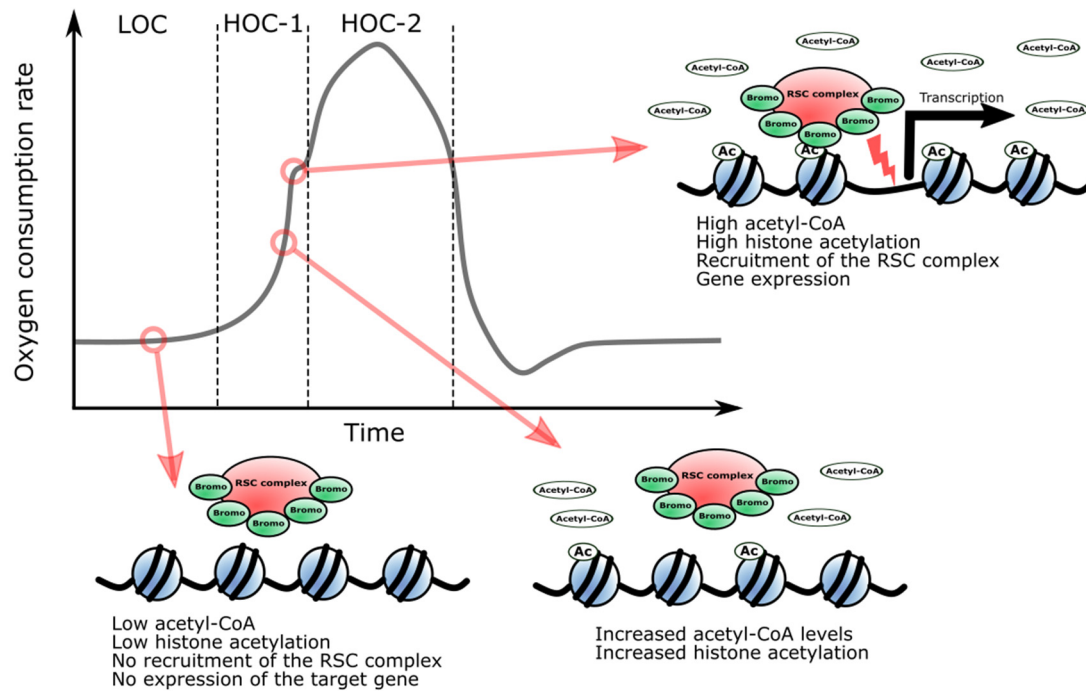


Figure 3.8 The hypothesized acetylation-recruitment pathway. Upon entering the HOC-1 phase, there is an increase in the level of acetyl-CoA, which leads to the acetylation of histones. The acetylated histones recruit the RSC complex via the interaction with bromodomains at point B, which leads to gene expression and the entry of the HOC-2 phase. The red, green and white bubbles represent the RSC complex, the bromodomains and acetyl-CoA respectively.

This chapter will attempt to examine this hypothesis by answering three key questions. First, does the RSC complex play a role during the YMC? To answer this question, the yeast strains carrying deletions or mutations of various bromodomain containing subunits of the RSC complex were tested in the YMC conditions individually. The RSC complex may play a role during the YMC, if some of these mutants show YMC defects. The results are shown in section 3.2.1 - 3.2.6.

If the answer to the first question is yes, the second question will be whether there exist genes that are activated in the HOC-2 phase by the RSC complex via the hypothesized acetylation-recruitment pathway. To answer this question, the temporal changes in transcripts during the YMC were compared between the wild-type and the RSC mutant, which could abolish certain function of the RSC complex. After combining the data of histone H3 acetylation measured by ChIP-seq, several statistical tests were utilized to test if there was a significant positive

association between histone H3 acetylation and the gene expression changes in response to the loss of the function of the RSC complex. Individual genes were checked in detail to find the candidate to test in the third question below. The results are shown from section 3.2.8 to 3.2.15.

Third, are the genes identified from the second question important for the entry of the HOC-2 phase? To answer this question, the strain carrying the deletion of the gene was constructed and tested in the YMC conditions. If it was required for the entry of the HOC-2 phase, the strain might exhibit defects in the HOC-2 phase during the YMC. The results are shown from section 3.2.16 to 3.2.19.

3.2 Results

3.2.1 The *rsc2Δ* strain has a fluctuating oxygen consumption profile instead of the normal YMC, while no obvious defects can be seen in the *rsc1Δ* strain

As was mentioned before, the RSC complex has two forms, depending on whether it is associated with its Rsc1 or Rsc2 subunit (Cairns et al. 1999). To examine if either Rsc1 or Rsc2 could be required for the YMC, both the *rsc1Δ* strain and the *rsc2Δ* strain were tested in the same YMC conditions at the high dilution rate (0.1 hr⁻¹). The results are shown in Figure 3.9 and 3.10.

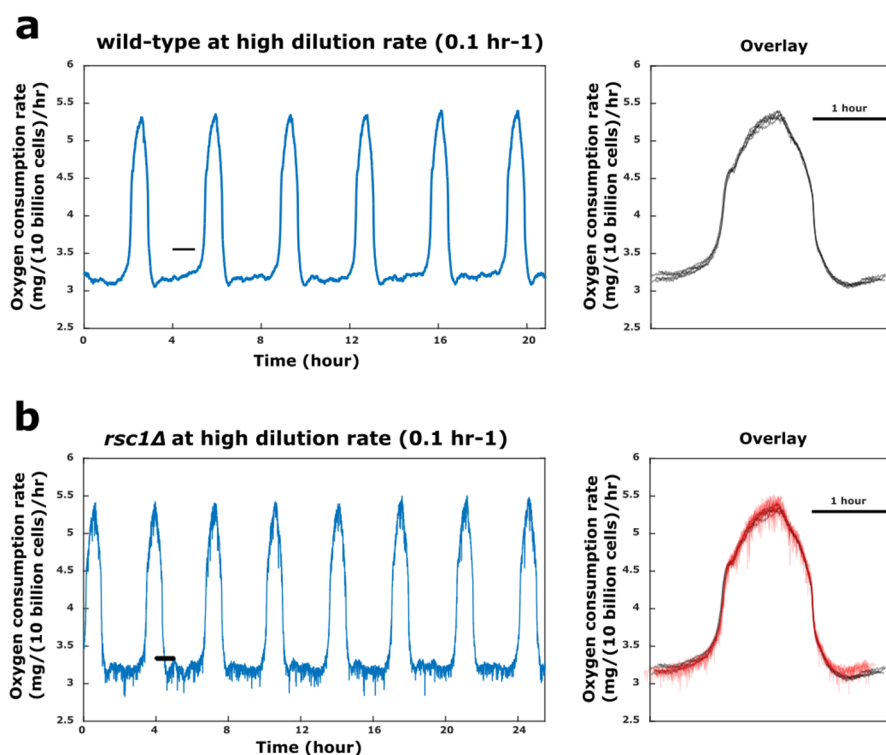


Figure 3.9 The oxygen consumption profiles for the wild-type strain (Figure a) and the *rsc1Δ* strain (Figure b) at the high dilution rate. The profiles of the *rsc1Δ* strain (the red curves in the overlay) can almost overlap with the wild-type (the black curves in the overlay). The black bars represent the length of one hour for the wild-type. The spikes seen in the trace of the *rsc1Δ* strain are not biological relevant. They are the electrical noise of the DO sensor.

The loss of Rsc1 does not seem to disturb the cellular oxygen consumption profile. This is evidenced by the fact that the oxygen consumption curves of the *rsc1Δ* strain almost completely overlap with that of the wild-type. This result strongly suggests that Rsc1 is not

required for the YMC, unless the *rsc1Δ* strain has certain YMC defects which cannot be manifested by the curve of cellular oxygen consumption rate.

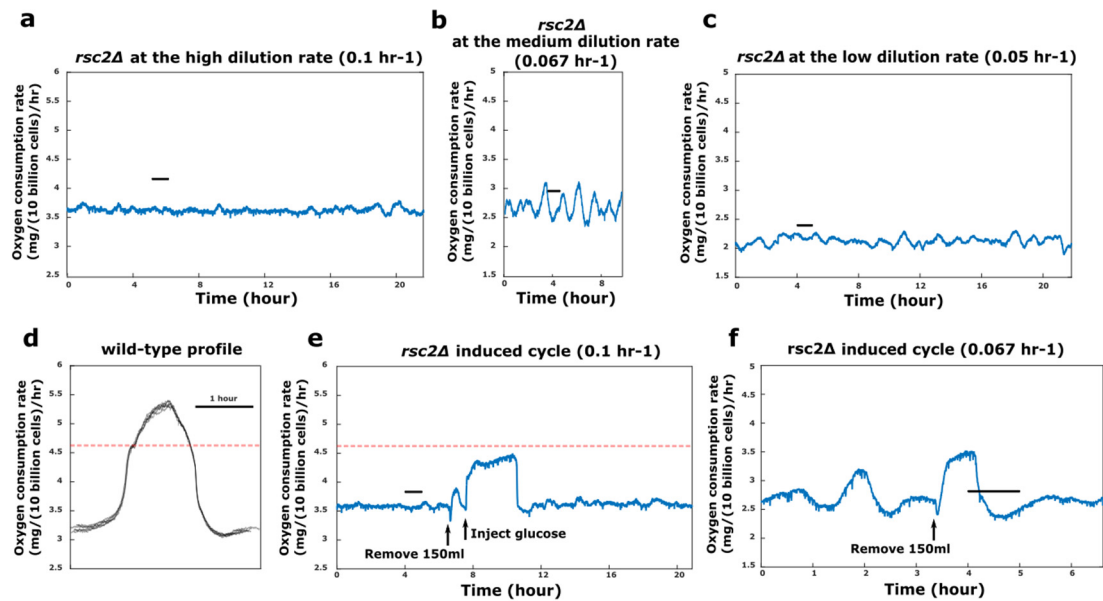


Figure 3.10 The oxygen consumption profiles for the *rsc2Δ* strain at different dilution rates. **a.** the high dilution rate. **b.** The medium dilution rate. **c.** The low dilution rate. **d.** The wild-type profile at the high dilution rate. **e.** The cells of the *rsc2Δ* strain can be induced to enter the HOC phase, however, the maximum oxygen consumption rate is lower than the wild-type, which is indicated by the red dashed line. **f.** The cells of the *rsc2Δ* strain can also be induced to enter the HOC phase at the medium dilution rate.

Strikingly, the loss of Rsc2 does cause strong defects in the YMC. In the same conditions as the *rsc1Δ* strain, instead of the normal YMC, the *rsc2Δ* strain can only produce small random fluctuations in the oxygen consumption rate (Figure 3.10a), which is reminiscent of the phenotype of the *gcn5Δ* strain as well as the strain bearing penta-lysine-to-arginine mutations on histone H3 (Cai et al. 2011; Kuang et al. 2014). Considering the wild-type can also lose the YMC at a much higher dilution rate (0.2 hr⁻¹), there is a possibility that the current dilution rate (0.1 hr⁻¹) is too high for the *rsc2Δ* strain. To test this possibility, the strain was grown at the medium and low feeding rates (0.067 hr⁻¹ and 0.05 hr⁻¹) as well. The *rsc2Δ* strain still produces fluctuations instead of the normal YMC at both feeding rates (Figure 3.10b,c), albeit

the medium feeding rate seems to have a slightly higher amplitude of fluctuations than the high and low feeding rates.

It is already found that a YMC-like response can be induced in the *gcn5Δ* strain (which does not exhibit the normal YMC) (Cai et al. 2011). To see whether the YMC-like response can also be induced in the *rsc2Δ* strain, 150ml of cell culture was removed from the fermenter, which has a similar effect to the injection of the carbon sources (For details, please refer to section 3.1.1.5). The treatment does induce a YMC-like response, however the maximal oxygen consumption rate is much lower than that of the normal YMC (Figure 3.10e). To further test the maximal oxygen consumption rate of the *rsc2Δ* strain, a large dose of carbon sources (5ml of 20% glucose solution) was injected into the cell culture (Figure 3.10f). In the beginning, the oxygen consumption rate rises sharply. After that, it reaches a plateau. The plateau is higher than the previous YMC-like response, but still lower than the level of point B in the normal YMC. After 4 hours, the oxygen consumption rate eventually returns to the ground level. The presence of the plateau indicates that the maximal oxygen consumption rate of the cells is unlikely to be further elevated by injecting a larger dose of glucose. It also suggests that there may exist certain defects in the *rsc2Δ* strain that can cause it to be unable to further increase the oxygen consumption rate.

To conclude, based on the oxygen consumption profiles, Rsc2 but not Rsc1 is required for the normal YMC. However, it is not known whether Rsc2 regulates the YMC on its own or via the RSC complex during the YMC.

3.2.2 The Rsc2 isoform of the RSC complex is required for the YMC

The c-terminal domain of Rsc2 has been shown to be responsible for its assembly into the RSC complex and the c-terminal domain deletion did not influence the protein level of Rsc2 (Cairns et al. 1999). If the deletion of the c-terminal domain can cause a phenotype that mimics *rsc2Δ*, Rsc2 may function as a subunit of the RSC complex during the YMC rather than on its own.

The *rsc2-CT1Δ* strain was tested at both high (0.1 hr⁻¹) and medium (0.067 hr⁻¹) dilution rates. The curve of oxygen consumption rates shows almost a flat line with small fluctuations for both conditions, which is similar to the phenotype of the *rsc2Δ* strain (Figure 3.11). This result supports the view that the interaction between Rsc2 and the rest of the RSC complex is required for the YMC.

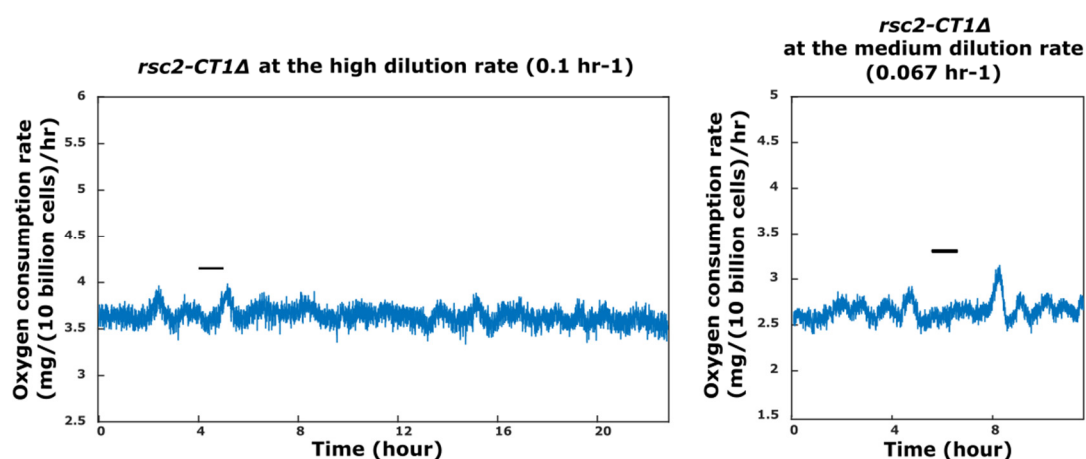


Figure 3.11 The oxygen consumption profiles for the *rsc2-CT1Δ* strain at the high and medium dilution rates. The black bar represent the length of one hour.

This result also indicates that the remodelling function of the RSC complex may be required for the YMC. To test this, an experiment can be designed to disturb the function of Sth1 (the catalytic subunit of the RSC complex, required for chromatin remodelling) and observe the phenotype in the YMC conditions. Since the direct deletion of *STH1* can cause lethality, certain knockdown systems, such as the auxin-degron system or the anchor-away technique, may need to be used instead (Nishimura et al. 2009; Haruki et al. 2008).

So far, the first question in section 3.1.2.4 has been answered. The Rsc2 isoform of the RSC complex is required for the YMC. Apart from the c-terminal domain, Rsc2 also contains two bromodomains (BD1 and BD2) and a BAH domain. In order to obtain more mechanistic insights, the strains carrying deletions of other domains of Rsc2 were tested under the YMC conditions. The results are shown in the following sections.

3.2.3 The loss of Rsc2 bromodomain 1 results in defects in the HOC-2 phase

Based on the growth tests in various stress conditions, the BD1 domain of Rsc2 is known to be required for the growth on caffeine containing plate but not required for other the functions of Rsc2 (Cairns et al. 1999). In order to know whether it is required for the YMC, the *rsc2-BD1Δ* strain was grown in the YMC conditions at the high dilution rate (0.1 hr⁻¹).

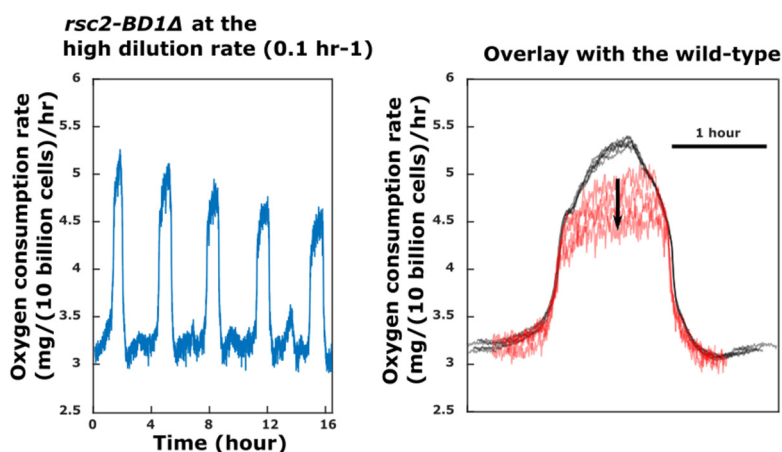


Figure 3.12 The oxygen consumption profiles for the *rsc2-BD1Δ* strain at the high dilution rate. In the right panel, the red curves are the profiles of the *rsc2-BD1Δ* strain, while the black curves are profiles of the wild-type at the same dilution rate. The black arrow indicates how the curves of the *rsc2-BD1Δ* strain change with time. The black bar represents the length of one hour.

As is shown in Figure 3.12, although the mutant is able to exhibit the periodicity of the YMC, defects in the HOC-2 phase of the YMC develop progressively after the YMC appears. Specifically, compared to the wild-type, the HOC-2 phase of the mutant is relatively flat. The maximal oxygen consumption rate of the mutant also fails to reach the same level as the wild-type. Due to the limited testing time, it is not known whether these defects will further develop or stabilize at certain states. This experiment needs to be repeated in future in order to get the answer.

To conclude, unlike the full length deletion of Rsc2, the bromodomain 1 of Rsc2 is required for the HOC-2 phase of the YMC, but not the periodicity of the YMC.

3.2.4 The *rsc2-BD2Δ* strain cannot exhibit the YMC at the high dilution rate

The study in the literature suggests the BD2 domain is required for most of the functions of Rsc2 (Cairns et al. 1999). To find out whether it is required for the YMC, the *rsc2-BD2Δ* strain was first tested at the high dilution rate (0.1 hr⁻¹).

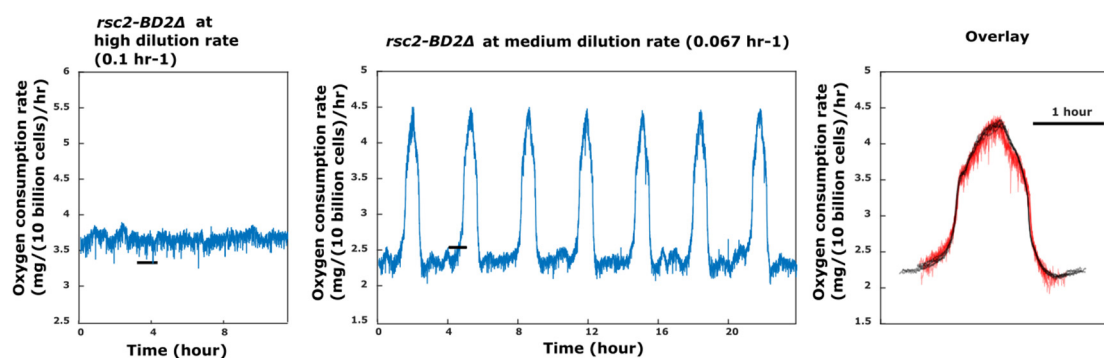


Figure 3.13 The oxygen consumption profiles for the *rsc2-BD2Δ* strain at the high (left panel) and medium (middle panel) dilution rate. In the right panel, the curves of the *rsc2-BD2Δ* strain (red) at the medium dilution rate is overlaid with that of the wild-type (black) at the same dilution rate. The black bar represents the length of one hour.

As is shown in Figure 3.13, the *rsc2-BD2Δ* strain has an almost flat profile of the oxygen consumption rate, which is highly similar to the observations on the strain *rsc2Δ* under the same condition. To test whether this strain is also similar to the *rsc2Δ* strain at a lower dilution rate, it was further grown at the medium dilution rate (0.067 hr⁻¹). Interestingly, it exhibits an almost normal YMC profile, which suggests that the maximal permissive dilution rate of the *rsc2-BD2Δ* strain is lower than that of the wild-type.

Although the mechanism behind such a difference is not known, one explanation can be the underexpression of a certain metabolic enzyme in the central metabolic pathway. At the high dilution rate, the substrate of this enzyme may exceed its maximal capacity. The accumulation of the substrate may further disturb the metabolic flux and the progression of the YMC, which leads to the loss of the YMC. It can be supported by two facts. First, the wild-type can also lose

the YMC if the dilution rate is too high (data not shown). Second, constantly adding a small amount of ethanol to the culture medium is enough to suppress the YMC (data not shown).

To summarize, the deletion of the BD2 domain of Rsc2 can cause the loss of the YMC at the high but not the medium dilution rate.

3.2.5 The *rsc2-BAHΔ* strain has defects in the HOC-2 phase of the YMC as well

Like the BD2 domain, the BAH domain was shown to be responsible for the majority of the functions of Rsc2 (Cairns et al. 1999). To test whether the YMC also requires the BAH domain of Rsc2, the *rsc2-BAHΔ* strain was grown in the YMC conditions at the high dilution rate (0.1 hr⁻¹).

Unlike the *rsc2Δ*, *rsc2-CT1Δ* and *rsc2-BD2Δ* strain, the *rsc2-BAHΔ* strain does exhibit periodic changes in its oxygen consumption rate. However, the shape of its oxygen consumption curve is different from that of the wild-type (Figure 3.14). During the LOC phase, the oxygen consumption rate of the wild-type is mostly maintained at a stable level, while the *rsc2-BAHΔ* strain seems to have an extra peak of oxygen consumption. In the HOC-2 phase, the oxygen consumption curve of the *rsc2-BAHΔ* strain is more flattened than that of the wild-type. Furthermore, on the oxygen consumption curve, the highest point of this strain is also lower than that of the wild-type. Both features are reminiscent of the phenotype of the *rsc2-BD1Δ* strain.

To sum up, the *rsc2-BAHΔ* strain is similar to the *rsc2-BD1Δ* strain. It shows periodic changes in the oxygen consumption rate, but has defects in the HOC-2 phase.

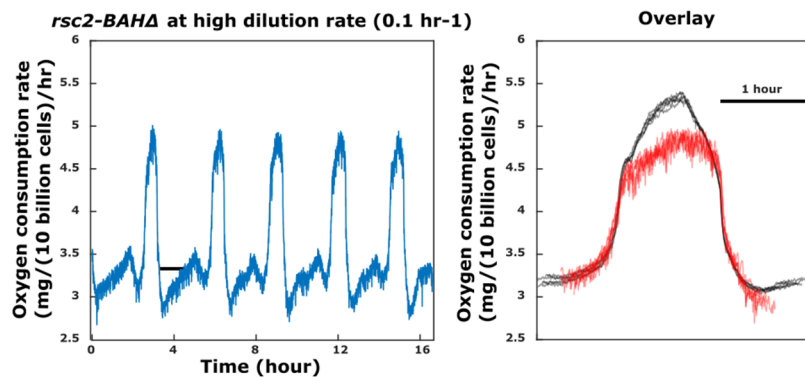


Figure 3.14 The oxygen consumption profiles for the *rsc2-BAHA* strain at the high dilution rate. In the right panel, the curves of the *rsc2-BD2Δ* strain (red) at the high dilution rate is overlaid with that of the wild-type (black) at the same dilution rate. The black bar represents the length of one hour.

3.2.6 No obvious YMC defects can be seen in either *sth1-BDΔ* or *rsc4-Y225AY226A*

This study also tested the bromodomains for the Sth1 and Rsc4 subunits of the RSC complex. For Sth1, the only bromodomain of it was deleted directly. There is a paper suggesting that this mutant is temperature sensitive and has abnormal cell shapes (Du et al. 1998). However, such phenotype was not observed in the prototrophic *sth1-BDΔ* strain (based on CEN.PK113-7D) used in this study. In the other strain, *rsc4-Y225AY226A*, the two key residues of the second bromodomain of Rsc4 were mutated, which can weaken the interaction between the bromodomain and H3K14Ac (Kasten et al. 2004).

Both strains were tested under the YMC conditions at the high dilution rate (0.1 hr⁻¹). The results are shown in Figure 3.15. Both strains have an almost wild-type temporal profile of the oxygen consumption rate. It suggests that the bromodomain of Sth1 or the second bromodomain of Rsc4 is dispensable for the YMC, however, it is not known if a combination of Sth1 bromodomain deletion and Rsc4 bromodomain mutation can cause defects.

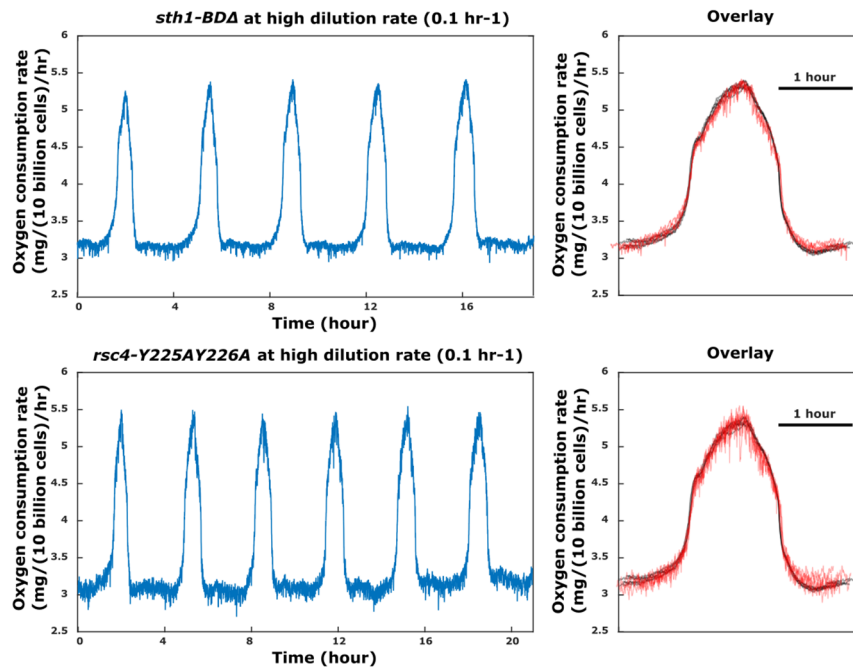


Figure 3.15 The oxygen consumption profiles for the *sth1-BDA* (upper panels) and *rsc4-Y225AY226A* (lower panel) strain at the high dilution rate. In the right panels, the mutant strains (red) are overlaid with the wild-type (black). The black bar represents the length of one hour.

3.2.7 A summary for the answer to the first question

The Rsc2 isoform of the RSC complex is required for the YMC. All domains of Rsc2 are required for the normal YMC. Specifically, the BD1 and BAH domains of Rsc2 are required for the HOC-2 phase of the YMC. The bromodomains of Rsc1 and Sth1 as well as the second bromodomain of Rsc4 seem to be dispensable for the YMC.

The only two bromodomains of the RSC complex, which were not tested, are the bromodomain of Rsc58 subunit and the first bromodomain of Rsc4 subunit. They should also be tested in future experiments. The first bromodomain of Rsc4 can be tested by mutating its tyrosine 92 and 93 residues to alanine, which disrupts its interaction with H3K14Ac (Kasten et al. 2004). For Rsc58, the full length deletion is known to cause lethality (Giaever et al. 2002). Currently, it is not known whether its bromodomain deletion is also lethal. If yes, conditional knock-down methods, such as the degron system (Natsume & Kanemaki 2017), may be used.

Currently, histone H3 acetylation is hypothesized to recruit the Rsc2 isoform of the RSC complex. However, it is worth to note that the current evidence is still not enough to prove this hypothesis. Other possibilities also exist. For example, acetylated non-histone proteins might be able to recruit the RSC complex. To validate this hypothesis, a possible experimental plan is proposed in section 3.3.1. The following sections focuses on checking the genes downstream of and dependent on the Rsc2 isoform of the RSC complex during the YMC.

3.2.8 RNA-seq samples are taken from the induced YMC-like response of the *rsc2Δ* strain and the normal YMC of the wild-type

More than 60% of genes have highly dynamic changes in their transcripts levels during the YMC (Tu et al. 2005). To test the hypothesized acetylation-recruitment pathway, it is important to answer two questions first. Question one: what genes may require the Rsc2 isoform of the RSC complex for their expression? Question two: when does the Rsc2 isoform of the RSC complex induce their expression during the YMC? If the acetylation-recruitment hypothesis is correct, the genes requiring the RSC complex for their expression at a certain time during the YMC should have a preceding peak of histone acetylation just before that time. It can be examined by statistical tests. Here, to answer the two questions, the temporal profiles of gene expression in the *rsc2Δ* strain were compared to those in the wild-type by performing RNA-seq on samples taken from both strains.

As shown in section 3.2.1, the *rsc2Δ* strain does not exhibit the YMC. However, a YMC-like response could still be induced in this strain by quickly removing 150ml of the cell culture (out of the total volume of 1050ml) from the fermenter (Figure 3.16a,b). 10 samples were taken from the *rsc2Δ* strain, which covered the whole induced YMC-like response. For the wild-type, the whole YMC cycle was sampled at 22 different time points. As most of the dynamic changes in gene expression happen in the HOC phase, the HOC phase was sampled at a higher frequency (every 5mins). To enable a quantitative comparison between gene expression

profiles, all the samples were spiked in with the same batch of *S.pombe* cells. Finally, RNA-seq libraries were constructed from all the samples.

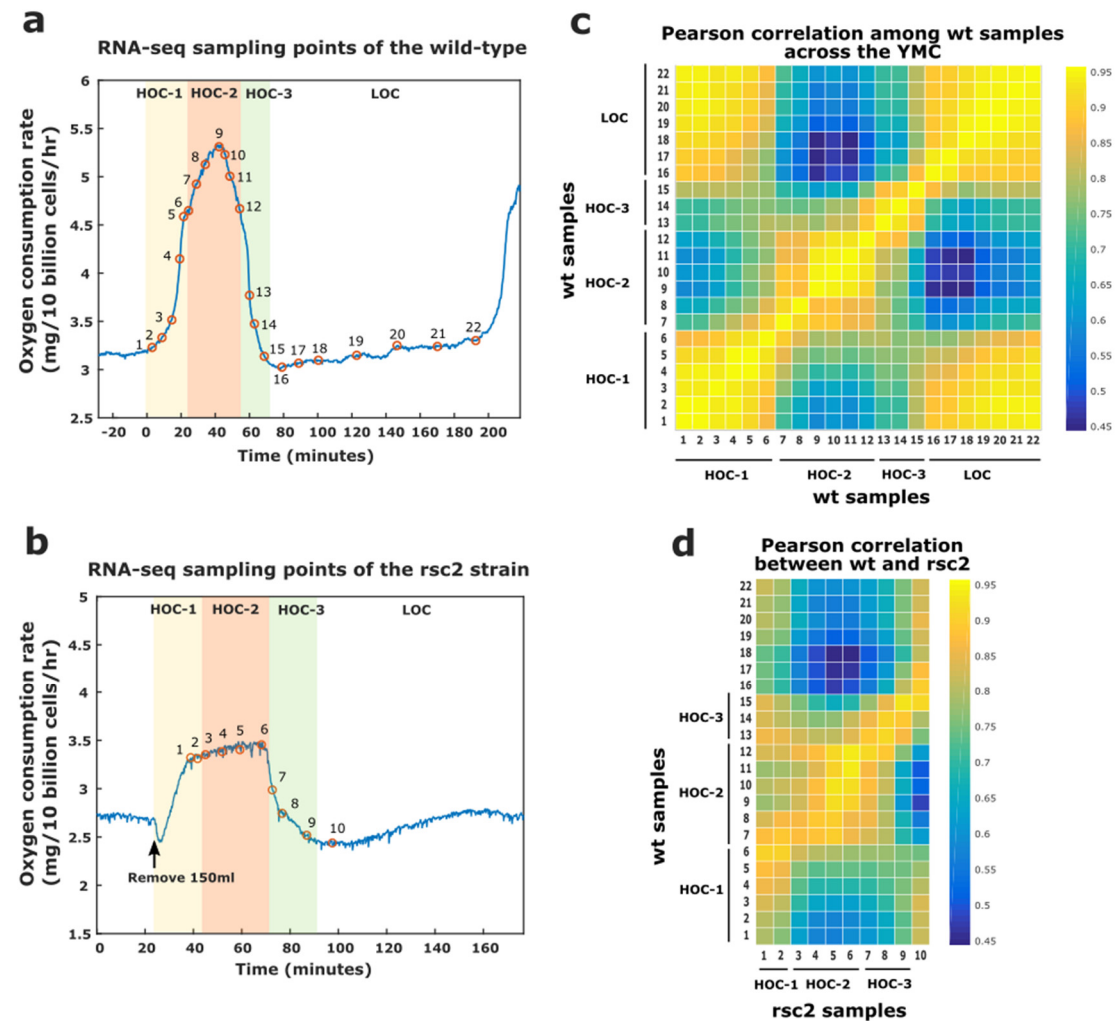


Figure 3.16 The sampling time points and the Pearson correlation matrix of the RNA-seq samples from the wild-type and the *rsc2Δ* strain. **a.** 22 samples were taken from a normal YMC of the wild-type. Different HOC phases are marked with different colors. **b.** 10 samples were taken from the induced YMC-like response of the *rsc2Δ* strain. **c.** The heat map visualizes the self-correlation matrix of the 22 wild-type RNA-seq samples. Warm colors represent high correlations and similarities, while cold colors indicate low correlations and low similarities. **d.** The heat map visualizes the correlation matrix between the 22 RNA-seq samples of the wild-type and the 10 RNA-seq samples of the *rsc2Δ* strain.

It is worth to bear in mind that the induced YMC-like response can be different from the normal YMC in many aspects. Although many results in the later sections suggest they are similar in general, certain specific differences may still exist. The following sections attempt to find the changes in mRNA levels that may be associated with the deletion of *RSC2*. However,

there always exists two alternative possibilities. First, the observed differential gene expression is associated with the difference between the induced YMC-like response and the normal YMC, rather than the deletion of *RSC2*. Second, the changes in mRNA level can always be the indirect effects of *RSC2* deletion, such as the stress response or certain compensatory effects. For the sake of brevity, the two possibilities are mentioned repeatedly in the following sections.

3.2.9 The matrix of Pearson correlation coefficient maps the sampling points of the *rsc2Δ* strain to the time axis of the wild-type and also largely agrees with the phases of the YMC

Before comparing the gene expression profiles between the wild-type and the *rsc2Δ* strain, it is necessary to map the sampling time points of the *rsc2Δ* strain to the time axis of the wild-type. One way is to find the most similar time point among all the wild-type sampling points for each of the *rsc2* samples. Pearson correlation coefficient is a good metric for assessing the degree of similarity between the genome-wide gene expression levels of two RNA-seq samples. It may be possible to map each *rsc2* sample to the time axis of the wild-type by finding the wild-type time point where the maximal Pearson correlation coefficient is achieved.

This mapping method was first tested by calculating the Pearson correlation coefficients between every two wild-type RNA-seq samples. If the mapping method works, for each wild-type sample, the highest Pearson correlation coefficient should be achieved at its temporally closest sample. Figure 3.16c visualizes the 22-by-22 matrix of Pearson correlation coefficients with a heat map. Obviously, the values on the diagonal line are always 1, because the samples were compared to themselves. Importantly, for each sample, the second highest value of the coefficient is always achieved by its most adjacent sample, which justifies this method.

Interestingly, the degree of similarity of genome-wide gene expression profiles largely agrees with the YMC phases defined by the curve of oxygen consumption rates (Figure 3.16c). Wild-

type samples from 7 to 12 have relatively high levels Pearson correlation coefficient for their gene expression. They cover the whole HOC-2 phase. Likewise, wild-type samples from 13 to 15, which covers the whole HOC-3 phase, are also similar in terms of their gene expression. Strikingly, the gene expression profiles between the LOC phase and the HOC-1 phase are highly similar, although the two phases are significantly different in terms of their cellular oxygen consumption profiles. It suggests that the increase in the oxygen consumption rates upon entering the HOC-1 phase is not accompanied by genome-wide remodeling of gene expression. The entry of the HOC-1 phase from the LOC phase is more likely to be driven by changes in the transcription of few genes or purely by post-transcriptional events.

To find the corresponding YMC phases of the *rsc2* samples, the matrix of Pearson correlation coefficient between the samples of the *rsc2Δ* strain and the wild-type was generated and visualized (Figure 3.16d). The pairs of samples with high correlation coefficients also form a diagonal line on the heat map. It suggests that the progressive changes in the genome-wide gene expression profiles of the induced YMC-like response of the *rsc2Δ* strain generally follow the same temporal sequence as the normal YMC in the wild-type. By matching the best correlated wild-type counterpart for each *rsc2* sample, it is found that *rsc2* samples 1-2, 3-6 and 7-9 belong to the HOC-1 phase, the HOC-2 phase and the HOC-3 phase respectively. Interestingly, the oxygen consumption curve of the *rsc2Δ* strain also largely fits such a phase separation. The major difference between the wild-type and the *rsc2Δ* strain lines in the HOC-2 phase. The oxygen consumption rate is further increased upon entering the HOC-2 phase in the wild-type but not in the *rsc2Δ* strain. It implies that the Rsc2 isoform of the RSC complex may be required for the entry of the HOC-2 phase.

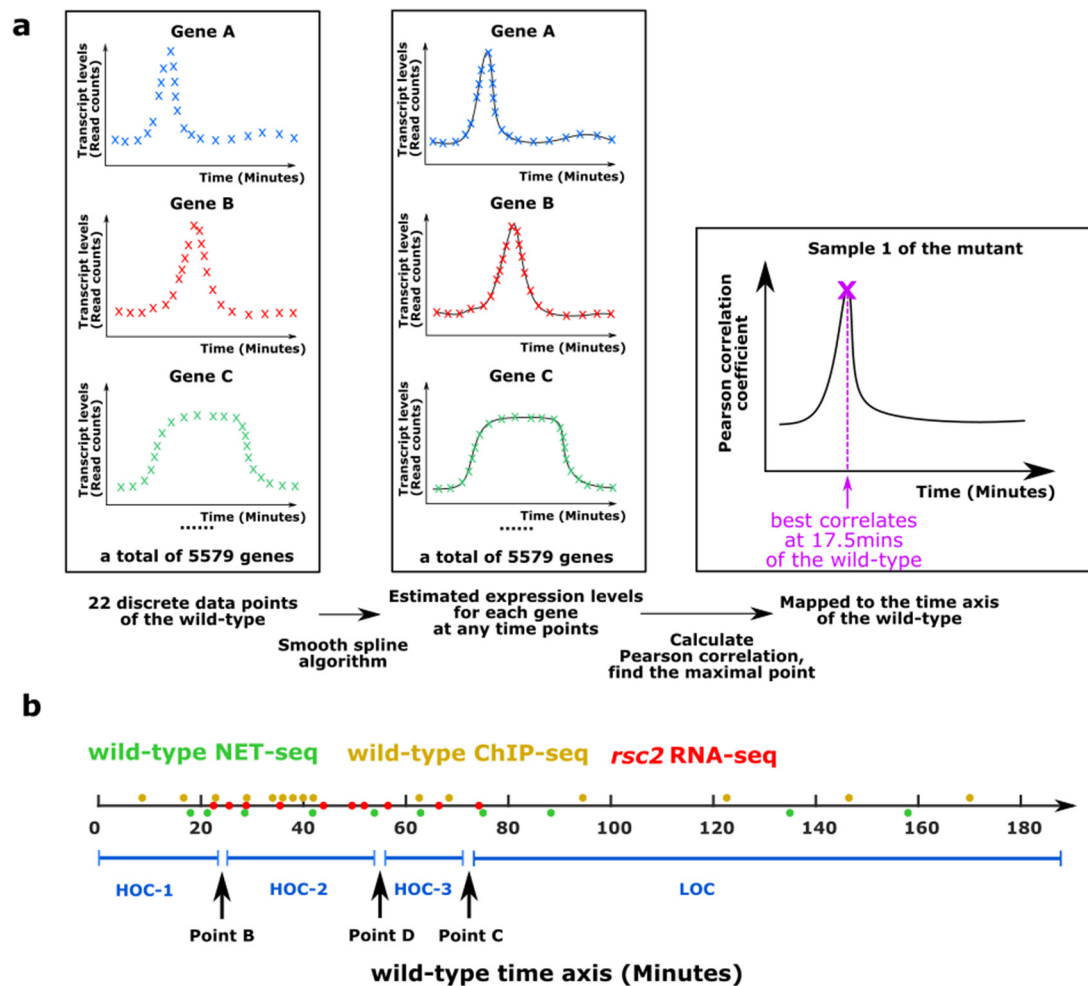


Figure 3.17 Different datasets were mapped to a uniform time axis. a. The workflow of mapping the time points of multiple datasets to the time axis of the wild-type. **b.** The estimated time points of NET-seq (green dots), ChIP-seq (yellow dots) and *rsc2* RNA-seq (red dots) on the time axis of the wild-type strain.

Before progressing to the analysis of gene expression profiles, it is critical to obtain the best estimates for the timings of each *rsc2Δ* sample on the time axis of the normal wild-type YMC. The major difficulty was that all the data points were discrete, so there was no direct information for the gene expression profiles between two adjacent wild-type samples. To circumvent it, a smooth curve was fitted to the 22 discrete wild-type data points for each gene by smoothing spline algorithm (Figure 3.17a). After that, the wild-type genome-wide gene expression profiles at any points on the time axis were estimated. Although the estimation has the risk of underrepresenting the potential rapid gene expression changes between two adjacent samples, this risk was minimized by the high sampling density (roughly every 5 mins)

of the wild-type. For each *rsc2Δ* sample, the exact point where the maximal Pearson correlation coefficient was achieved could be found on the time axis of the wild-type normal YMC. After that, all the *rsc2Δ* sample and the wild-type samples could be mapped to a uniform time axis of the normal YMC. NET-seq (For details, please refer to appendix section S1) and ChIP-seq (data from Kuang et al. 2014) datasets all have their companion RNA-seq datasets prepared from the same samples, which allows the wild-type YMC NET-seq and the wild-type YMC ChIP-seq datasets to be mapped to the same uniform time axis by the same method (Figure 3.17b). Now, it is possible to examine the differences in their temporal profiles of gene expression. It is worth noting here that these yeast NET-seq libraries were prepared before the ultra-low background yeast NET-seq (chapter 4.1) was developed, so the NET-seq libraries may still contain 10-30% non-specific background from mature RNA.

3.2.10 There are two particular time windows in the HOC phase where the peaks of the transcripts of certain genes are not seen in the *rsc2Δ* strain

To provide an overview of the impact of the *RSC2* deletion on the transcriptome during the YMC, genes were categorized into different clusters, according to the shapes of their temporal expression profiles in both the wild-type and the *rsc2Δ* strain. This section attempts to provide a general answer to two important questions. First, what kinds of genes are affected by the deletion of *RSC2*? Second, when are these genes affected during the YMC?

Specifically, the YMC time-course RNA-seq datasets of both the wild-type and the *rsc2Δ* strain were utilized to carry out the analysis. 4499 genes were retained after filtering lowly expressed genes. Before performing the clustering algorithm, the time course data for each gene were subtracted by their median and then divided by their standard deviation. It allows all the genes to have a median equal to zero for both wild-type and the *rsc2Δ* strain. The expression changes were transformed into the number of standard deviation from the median. These procedures are required for the K-means clustering and allow the algorithm to focus on

the shape differences between the temporal profiles of the wild-type and the *rsc2Δ* strain. Then a total of 16 clusters were created by the K-means clustering algorithm. Each cluster contains around 280 genes and has a characteristic temporal profile of gene expression for the wild-type, which is indicated by the light blue curves. The corresponding characteristic gene expression profile of the *rsc2Δ* strain for each cluster is plotted with pink curves (Figure 3.18).

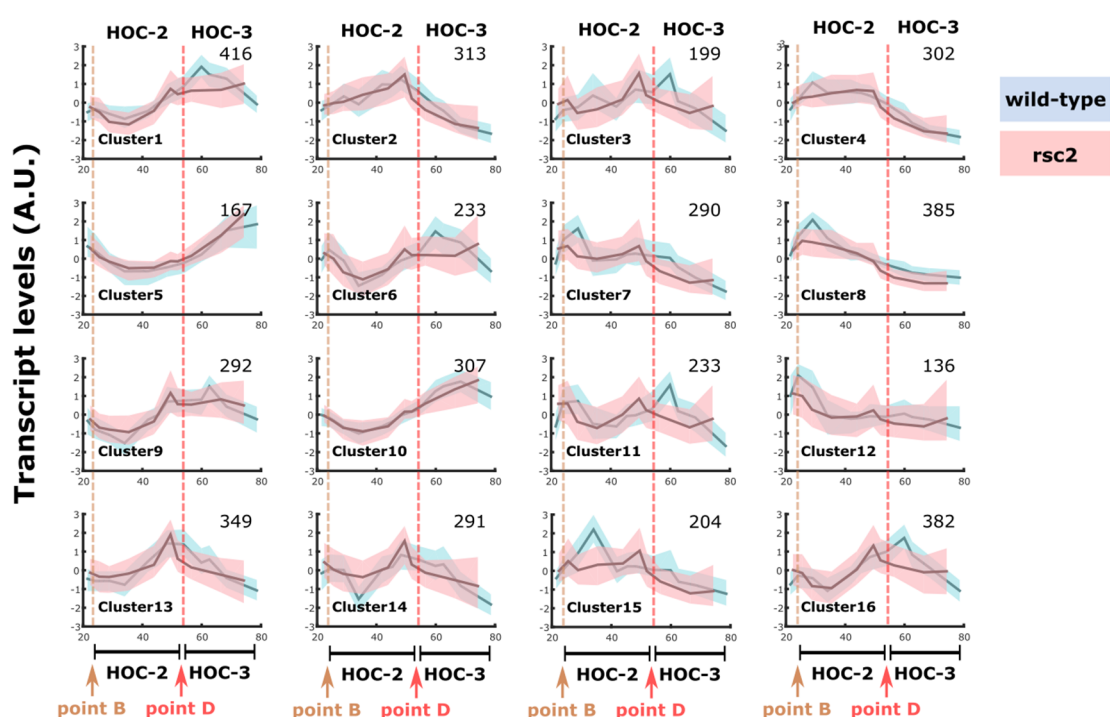


Figure 3.18 Genes were categorized into 16 clusters based on their temporal expression profiles in both the wild-type and the *rsc2Δ* strain. The characteristic profiles of the wild-type and the *rsc2Δ* strain in each cluster are colored blue and red respectively. The upper and lower bounds of a characteristic profile are the top and bottom 10 percent of the gene expression levels in a cluster. The median level in each cluster is represented with a thick black curve. The name of the cluster and the number of genes in the cluster are labelled at the bottom-left and top-right corners respectively.

In each cluster, the characteristic gene expression profile of the wild-type was compared to that of the *rsc2Δ* strain. It is found that the general shapes of the profiles of both strains are similar for most of the clusters. It indicates that the induced YMC-like response in the *rsc2Δ*

strain is largely comparable to the normal YMC of the wild-type and the deletion of *RSC2* does not lead to a complete change in the temporal profiles of gene expression.

However, in many clusters, the peak of transcripts appearing right after point B or point D in the wild-type seems to be lost in the *rsc2Δ* strain. In Figure 3.19, the clusters are rearranged into four groups (A-D). Group A contains cluster 4, 7, 8, 12 and 15. They have a peak of transcripts in the wild-type strain after point B, while this peak is greatly reduced or lost in the *rsc2Δ* strain (Figure 3.19 Group A). The peak normally spans the 18mins after point B. In the following discussion, this time window is referred to as TW1 (time window 1). In group A, cluster 4 and 8 are significantly enriched for ribosome biogenesis genes and translation related genes. Interestingly, these are the genes that are known to have a peak of histone acetylation at point B and their expression can be induced by injecting acetate, which can cause histone acetylation (Cai et al. 2011; Kuang et al. 2014). Similarly, Cluster 1, 3, 6, 11 and 16 have a peak of transcripts after point D (Figure 3.19 Group B). This peak is also not seen in the *rsc2Δ* strain. It normally spans the 10mins after point B. This time window will be referred to as TW2 (time window 2). In group B, cluster 6 is enriched for genes associated with proteasomal degradation, while cluster 1 and 16 have significantly more mitochondrial function related genes than the background. The rest of the clusters which do not show obvious differences in neither TW1 nor TW2 are grouped into C and D, according to the similarities in their general shapes (Figure 3.19 Group C and D). Clusters in group D have a 'U' shape. They have a strong association with aerobic respiration.

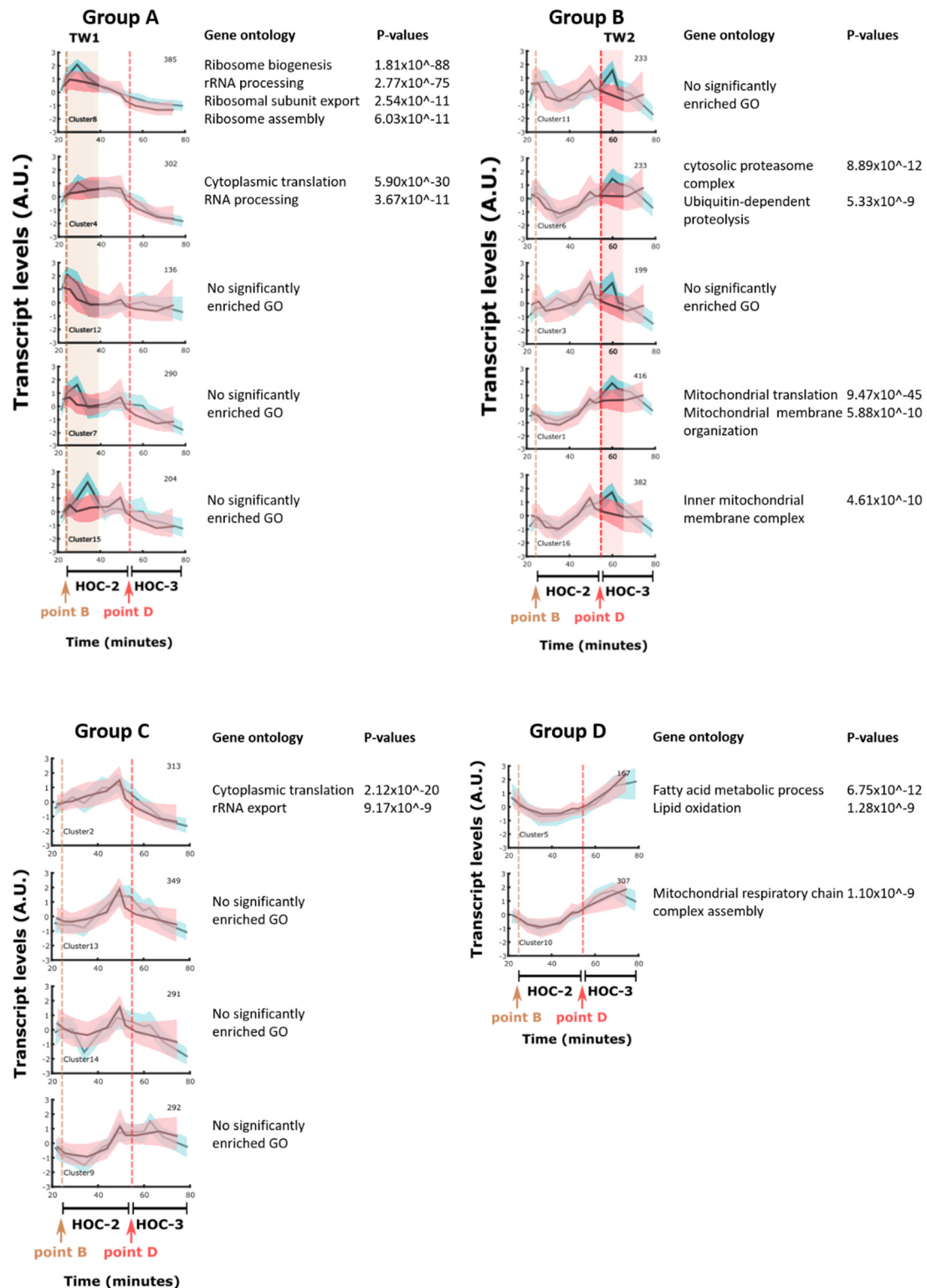


Figure 3.19 The 16 clusters were categorized into four groups (A, B, C and D). The gene ontology terms and corresponding p-values are listed on the right side of each cluster. Group A has the loss of transcript peaks in TW1, while group B has the loss of transcript peaks in TW2. The rest of clusters are grouped into C and D which differ in their general shapes of profiles.

Besides the two time windows, all clusters seem to gain an additional peak of transcript levels just before point D in the *rsc2Δ* strain. It could be a point where the RSC complex exerts its transcriptional suppressive functions. However, it is more likely to be an artefact, because this peak is only represented by a single RNA-seq sample and it almost appears in all clusters. It can be explained by the error of adding less *S.pombe* cells to that sample of the *rsc2Δ* strain, which can lead to the global overestimation of gene expression at that point.

To summarize, there are two time windows (TW1 and TW2) where the peaks of transcripts seen in the wild-type are lost in the *rsc2Δ* strain. The two time windows are located right at the beginning of the HOC-2 and HOC-3 phase respectively. Some clusters having peaks lost in TW1 are known to have peaks of histone acetylation at point B (just before TW1).

3.2.11 Genes losing their peaks of transcripts in TW1 in the *rsc2Δ* strain have a strong positive association with a peak of H3K9Ac, H3K14Ac, H3K18Ac or H4K5Ac at point B

If the genes losing their peaks of transcripts in either/both time windows after deleting *RSC2* are regulated by the acetylation recruitment mechanism, these genes, in the wild-type should be significantly associated with a peak of histone acetylation (particularly histone H3 acetylation, because the RSC complex is shown to bind to acetylated H3 rather than H4) before their peaks of transcripts.

To test the hypothesis, the single sided Fisher's exact test was performed to check the potential association between histone H3 acetylation near point B/D and the responsiveness to the deletion of *RSC2* in TW1/TW2. Genes were selected, if they had a peak of mRNA in TW1 in the wild-type. These genes were divided into 4 groups, based on two dimensions. The first dimension was whether they had a peak of histone H3 acetylation near point B. The second dimension was whether their peaks of mRNA levels in TW1 were lost in the *rsc2Δ* strain (the responsiveness to the deletion of *RSC2*). The contingency table for the statistical test was

created by counting the number of genes in each group. Finally, Fisher's exact test was performed on the contingency table and the p-values were calculated. The same procedures were repeated for point D and TW2. The level of histone H4 acetylation, histone H3 methylation and the total level of histone H3 were also tested as controls in this analysis.

All the p-values are summarized in Table 3.1. This analysis includes three H3 acetylation marks (H3K9Ac, H3K14Ac and H3K18Ac) which are known to have physical interactions with the RSC complex in the literature. For TW1, all the three histone H3 acetylation marks have a highly significant positive correlation with the responsiveness of *rsc2Δ* in TW1. Interestingly, H4K5Ac also has a low p-value for TW1. Except for H4K5Ac, none of the other controls have significant positive associations for TW1. Hence, the results for TW1 support the acetylation-recruitment pathway. For TW2, none of the tested histone modifications are significant. In other words, no significant positive association can be observed between the tested histone modifications near point D and the loss of mRNA peaks in TW2 in the *rsc2Δ* strain. It indicates that the peaks of mRNA in TW2 in the wild-type might be activated by the RSC complex through a different pathway or the loss of the peaks is an indirect effect of deleting *RSC2*. It could even just be the inherent difference between the induced YMC-like response and the normal YMC.

The preceding peaks of histone modifications at either point B or D	p-values	
	Loss transcript peaks in TW1	Loss transcript peaks in TW2
H3	0.99	0.95
H3K9Ac	1.74×10^{-20}	0.25
H3K14Ac	2.94×10^{-26}	0.73
H3K18Ac	3.05×10^{-24}	1.79×10^{-3}
H3K56Ac	0.15	0.74
H3K4Me3	1.99×10^{-4}	0.11
H3K36Me3	0.85	0.49
H4K5Ac	2.94×10^{-21}	0.92
H4K16Ac	0.98	0.44

Table 3.1 Testing the positive correlation between the responsiveness to *RSC2* deletion and the presence of a preceding peak of various histone modifications by single sided Fisher's tests. The p-values for all the combinations are listed. The combinations that show a significant ($p < 10^{-7}$) positive correlation are colored red.

To further test the acetylation-recruitment hypothesis, the single sided Fisher's exact test was performed on the 16 individual clusters (in section 3.2.10) to check whether they were enriched for genes that had a peak of H3 acetylation near point B or D. The results are shown in Table 3.2a and 3.2b. Clusters in group A are significantly more likely to have a peak of H3K9Ac, H3K14Ac, H3K18Ac and H4K5Ac at point B. At the same time, they are the clusters showing the loss of the peaks of transcripts in TW1 after deleting *RSC2*. Hence, this result also supports the acetylation-recruitment pathway. On the contrary, most clusters do not have significant enrichment for the peaks of any histone modification marks near point D. There are only two clusters (1 and 10) showing the significant enrichment for H3K18Ac near point D, however, they do not have significant enrichment for other H3 acetylation marks.

a

Histone modifications peaks at point B

		H3	H3K9Ac	H3K14Ac	H3K18Ac	H3K56Ac	H3K4Me3	H3K36Me3	H4K5Ac	H4K16Ac
Group A (peaks lost in TW1)	Cluster8	1.00	2.10x10 ⁻⁸⁰	2.24x10 ⁻⁹¹	2.71x10 ⁻¹⁰²	0.20	9.04x10 ⁻¹⁸	0.10	2.24x10 ⁻¹¹³	1.00
	Cluster4	1.00	5.49x10 ⁻⁵⁷	8.52x10 ⁻⁵⁶	6.91x10 ⁻⁶⁵	1.04x10 ⁻⁰⁷	3.44x10 ⁻⁰³	0.82	1.98x10 ⁻⁶⁷	1.00
	Cluster12	0.99	1.15x10 ⁻²⁸	2.51x10 ⁻²¹	1.08x10 ⁻²⁷	2.42x10 ⁻⁰⁹	0.94	0.06	3.60x10 ⁻²⁰	0.92
	Cluster7	0.60	2.23x10 ⁻¹⁹	1.42x10 ⁻¹³	1.59x10 ⁻⁰⁸	0.74	0.30	0.80	2.11x10 ⁻¹³	0.36
	Cluster15	1.00	4.57x10 ⁻¹⁰	1.38x10 ⁻⁰⁹	2.09x10 ⁻¹²	1.55x10 ⁻⁰³	0.75	0.39	2.05x10 ⁻¹¹	0.37
Group B (peaks lost in TW2)	Cluster11	0.72	0.03	0.93	0.56	0.99	0.50	0.22	1.00	0.69
	Cluster6	0.19	1.00	1.00	1.00	1.00	0.01	0.01	1.00	0.32
	Cluster3	0.38	0.95	0.89	0.94	0.99	0.99	0.25	1.00	0.47
	Cluster1	0.01	1.00	1.00	1.00	1.00	0.95	0.96	1.00	0.04
	Cluster16	0.01	1.00	1.00	1.00	1.00	1.00	0.73	1.00	0.53
Group C	Cluster2	1.00	0.01	0.01	0.01	1.23x10 ⁻⁰⁶	0.02	0.05	0.44	0.13
	Cluster13	0.08	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.05
	Cluster14	0.30	0.04	0.90	0.48	1.00	0.73	0.33	0.78	1.00
	Cluster9	0.26	1.00	1.00	1.00	0.99	0.89	0.39	1.00	0.58
Group D	Cluster5	0.02	1.00	1.00	1.00	7.38x10 ⁻⁰⁸	0.95	0.93	1.00	4.09x10 ⁻⁰⁴
	Cluster10	1.44x10 ⁻⁰³	1.00	1.00	1.00	0.53	0.17	0.57	1.00	2.47x10 ⁻⁰³

b

Histone modifications peaks at point D

		H3	H3K9Ac	H3K14Ac	H3K18Ac	H3K56Ac	H3K4Me3	H3K36Me3	H4K5Ac	H4K16Ac
Group A (peaks lost in TW1)	Cluster8	0.89	1.00	1.00	1.00	0.08	1.00	1.00	1.00	0.88
	Cluster4	0.01	1.00	1.00	1.00	0.40	0.99	0.34	1.00	0.01
	Cluster12	0.61	0.75	0.46	1.00	1.00	0.99	0.82	0.26	0.27
	Cluster7	0.22	0.92	1.00	1.00	0.14	0.01	0.32	0.94	0.31
	Cluster15	0.01	1.00	0.97	0.96	0.58	0.32	0.05	0.57	0.21
Group B (peaks lost in TW2)	Cluster11	0.37	0.87	0.59	0.95	0.10	0.23	0.58	0.78	0.42
	Cluster6	0.22	0.09	0.33	0.02	0.04	1.00	0.49	0.52	0.67
	Cluster3	0.29	0.87	0.69	0.54	0.97	0.02	0.03	0.33	0.07
	Cluster1	0.53	4.99x10 ⁻⁰⁴	0.07	1.78x10 ⁻¹⁰	0.04	0.70	0.78	0.01	0.69
	Cluster16	0.74	0.01	0.19	0.16	0.43	0.00	0.04	3.99x10 ⁻⁰⁵	0.66
Group C	Cluster2	0.05	1.00	0.99	1.00	1.00	0.02	0.07	1.00	0.06
	Cluster13	0.71	0.09	0.06	0.55	0.23	2.05x10 ⁻⁰⁸	3.40x10 ⁻⁰³	0.02	1.00
	Cluster14	0.91	0.85	0.83	0.88	0.02	0.31	0.92	0.98	0.15
	Cluster9	0.75	0.07	0.02	7.17x10 ⁻⁰⁴	0.65	0.99	0.65	7.06x10 ⁻⁰⁴	0.77
Group D	Cluster5	0.92	0.59	0.03	7.65x10 ⁻⁰⁴	1.00	0.99	1.00	0.25	0.94
	Cluster10	1.00	4.00x10 ⁻⁰³	1.76x10 ⁻⁰⁴	2.91x10 ⁻¹⁷	0.94	1.00	0.97	0.17	0.86

Table 3.2 a and b. Testing whether the 16 clusters were enriched for genes having a peak of certain histone modification at point B (Table a) or point D (Table b) by single sided Fisher's test. The p-values for all the combinations are listed. The combinations showing a significant ($p < 10^{-7}$) positive correlation are colored red.

To summarize, if a gene has a peak of mRNA in TW1 and a peak of H3 acetylation near point B, its expression is significantly more likely to be affected by *RSC2* deletion. Hence, the results for TW1 support the acetylation-recruitment hypothesis introduced in section 3.1.2.4. However, the strong positive correlation between histone H3 acetylation and *rsc2Δ* responsiveness seen in this study does not essentially mean that they have a causative

relationship. More experiments are required to examine whether they have a causative relationship. This is discussed in more detail in section 3.3.1.

The following sections try to check if the genes that are controlled by the hypothesized acetylation-recruitment pathway can have some functional implications on the cellular metabolism during the YMC. There were 906 genes whose peak of transcripts in TW1 is lost after deleting *RSC2* and also has a preceding peak of H3 acetylation near point B. Ideally, to see their functional implications on the YMC, all the 906 genes should be deleted and tested in the fermenter one by one. Based on the current throughput of the fermenter, this is impossible to achieve (The fermenter can test up to 15 mutants per year in its ideal condition). To solve this problem, I propose a high throughput screening experiment as shown in section 3.3.5.2. However, due to time limitations, this screening experiment cannot be accomplished in this DPhil study. Therefore, in order to narrow down the targets and find at least one or more genes to test in the fermenter, this study focused on the G1/S cyclin gene *CLN3* (section 3.2.12) and genes involved in the central metabolic pathway (section 3.2.13-19). The rationale for choosing them is provided in their corresponding sections.

3.2.12 The expression of *CLN3* is *Rsc2* dependent during the HOC-2 phase

CLN3 is the yeast G1/S cyclin. It has been shown that after injecting acetate into the fermenter, the cellular acetyl-CoA level increases dramatically. At the same time, histone H3 is found to be acetylated at the promoter of *CLN3*. A corresponding peak of the *CLN3* transcript is also seen after point B (Shi & Tu 2013). Apart from regulating the entry of a cell cycle, Cln3 is also found to upregulate catabolism via activating cyclin kinase Cdk1, which can promote the utilization of the stored trehalose by phosphorylating Nth1 (neutral trehalase 1) (Zhao et al. 2016). These findings make *CLN3* an attractive target to test, because of two reasons. First, the data suggest it is likely to be the downstream player of the hypothesis of the acetylation-

recruitment pathway. Second, it may have some regulatory role in the cell cycle and cellular metabolism during the YMC.

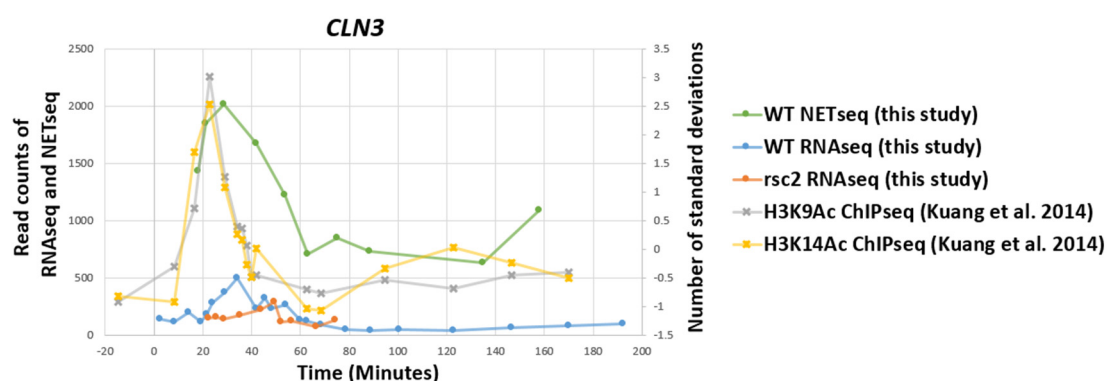


Figure 3.20 The temporal changes in the levels of transcription, transcripts and histone acetylation of *CLN3* during the YMC. The levels of transcription for the wild-type (blue curve) were assessed by NET-seq. The amounts of transcripts for the wild-type (blue curve) and the *rsc2Δ* mutant (red curve) were measured by RNA-seq. The transcript levels for the wild-type and the *rsc2Δ* mutant are comparable, because of the normalization by RNA spike-in. The signals of NET-seq and RNA-seq are expressed as the number of reads (the y-axis on the left hand side). The ChIP-seq signals for histone acetylation (Kuang et al. 2014) are expressed as the number of standard deviations from the median (the y-axis on the right hand side).

To examine *CLN3*, its temporal profiles of histone H3 acetylation, transcription and mRNA levels are plotted together in Figure 3.20. There is a clear peak of histone H3 acetylation at point B, which is associated with a rapid increase in the levels of transcription, measured by NET-seq. For the wild-type, the mRNA level of *CLN3* starts to increase after point B, while it largely remains as a flat line in the *rsc2Δ* strain. It means that the expression of *CLN3* requires Rsc2, which suggests that the Rsc2 isoform of the RSC complex may read the histone H3 acetylation marks and activate the expression of *CLN3*.

To conclude, the observations on *CLN3* support the hypothesis of the acetylation-recruitment pathway. To see the functional implications of the acetylation-recruitment pathway to cell metabolism during the YMC, *CLN3* could be a potential candidate to test in the fermenter.

3.2.13 A few genes have their expression patterns changed in the *rsc2Δ* strain; *ACO2*

is most likely to be the downstream of the hypothesized acetylation-recruitment pathway

As is shown in section 3.2.1, 3.2.8 and 3.2.9, the *rsc2Δ* strain has defects in further increasing oxygen consumption rates after passing point B, so the oxygen consumption curve of the HOC-2 phase of the *rsc2Δ* strain looks like a flat platform rather than the peak seen in the wild-type. One possibility is that the *rsc2Δ* strain may have defects in expressing genes in the central metabolic pathway, including glycolysis and TCA cycle. These genes may be upregulated by the hypothesized acetylation-recruitment mechanism after point B, which allows the aerobic respiration to increase. After deleting *RSC2*, these genes may fail to be expressed after point B, which causes the defects seen in the oxygen consumption curve of the HOC-2 phase of the *rsc2Δ* strain. To check it, the temporal profiles of genes in the central metabolic pathway were compared between the wild-type and the *rsc2Δ* strain.

The heat map (Figure 3.21 b) provides an overview of the log₂ fold gene expression changes in each of the sampling points of the *rsc2Δ* strain, with respect to the estimated gene expression levels of the wild-type at the same time point (The estimation method is the same as described in the section 3.2.9). Among all genes, *PDC6* and *ACO2* have the most prominent reduction in their transcripts levels just after passing point B, which is similar to *CLN3* shown in the previous section. Interestingly, except for *ACO2*, most TCA cycle genes have slightly elevated expression in their first four sampling points. In terms of glycolysis related genes, some of them seem to be moderately decreased in the first few sampling points. Some genes, such as *HXK1*, *PDC5*, *ALD4* and *CIT1* appear to have large fold increases at certain time points. Calculating log₂ fold changes allows the overall patterns to be assessed and highlights the most significant changes. However, it does not show the actual levels of gene expression. Artefacts may appear, if there exists some inaccuracies in the mapping of time points or some

denominators (the WT levels) are small values. Next, the shapes of the temporal profiles of multiple sequencing datasets, including ChIP-seq, RNA-seq and NET-seq are assessed for each gene.

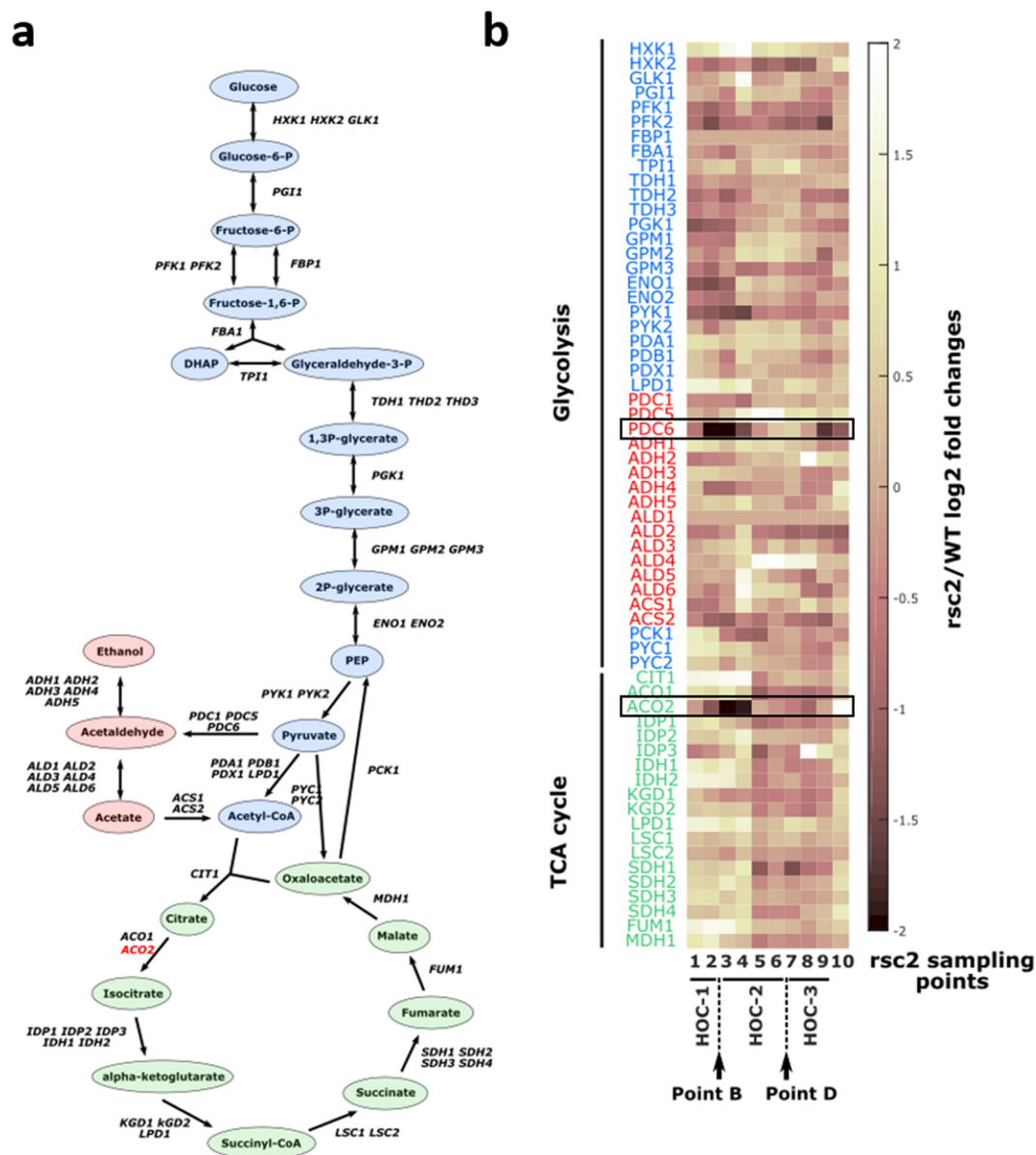


Figure 3.21 Changes in the expression of genes involved in the central metabolic pathways during the YMC after deleting *RSC2*. **a.** A diagram of the central metabolic pathways. Metabolites in the glycolysis, ethanol metabolism and TCA cycle are represented with blue, red and green bubbles. The genes coding the enzymes are written in bold italics. **b.** The heat map visualizes the $\log_2$ fold changes in the transcript levels of the *rsc2Δ* samples relative to the wild-type samples. Glycolysis, ethanol metabolism and TCA cycle related genes are written in blue, red and green respectively. The bright and dark colours of the heatmap represent overexpression and underexpression respectively. ACO2 and PDC6 show obvious reduction in gene expression at the boundary between HOC-1 and HOC-2 phases, so they are highlighted by the black boxes.

For each gene, the temporal profiles of histone H3 acetylation (yellow and grey curves), transcription (green curves) and transcript levels (blue for the wild-type; red for the *rsc2Δ* strain) are plotted together in Figure S6 (appendix). Some genes, such as *GLK1*, *TDH1*, *TDH3* and *IDP2*, do not have obvious changes in their temporal profiles of transcripts in the *rsc2Δ* strain (Figure 3.22). This suggests that the *rsc2Δ* samples are mapped to the time axis of the wild-type correctly and the induced YMC-like response is largely comparable to the normal YMC. After checking all 61 genes of the central metabolic pathway, three of them (*PDC6*, *ACO2* and *ADH4*) were found to lose (or at least reduce) their peak of transcripts in the *rsc2Δ* strain after point B (Figure 3.23). *ACO2* and *PDC6* have a dramatic peak of transcripts in the wild-type and the peak seems to be completely lost in the *rsc2Δ* strain. *ADH4* only has a small peak of transcripts in the wild-type after passing point B. Although this peak is also suppressed, the change in transcript levels is less than 2 fold. By checking the wild-type temporal profiles of NET-seq across the YMC, a peak of transcription activities could be found at point B for all three genes. This suggests that their transcript peaks are produced by transcription. By checking the ChIP-seq data for histone acetylation, both *ACO2* and *ADH4* were found to have a corresponding peak of histone H3 acetylation at point B. However, *PDC6* does not seem to have a peak of H3 acetylation, which suggests that it is not the downstream target of the hypothesized acetylation-recruitment pathway. To conclude, among the 61 genes involved in the central metabolism, *ACO2* is most likely to be the downstream target of the RSC complex in the acetylation-recruitment pathway.

Apart from the loss of the transcript peak after passing point B, there are three major differences in the transcripts temporal profiles of the *rsc2Δ* strain. First, many genes seem to have a lowered peak of transcripts after passing point D. Moreover, these peaks do not seem to be associated with a preceding peak of histone H3 acetylation. This is consistent with the statistical tests in section 3.2.11. The typical examples are *ALD6*, *ACS2*, *PYC1*, *ACO1*, *KGD1* and *KGD2* (Figure 3.24). It implies that the Rsc2 isoform of the RSC complex might participate in

gene expression after point D via a mechanism which is different from the putative acetylation-recruitment pathway. Second, some glycolysis related genes, such as *PYK1* and *PFK2*, show general reduction in mRNA levels (Figure 3.25). In other words, the reduction is not temporally specific. The third difference is that some genes appeared to overexpress after deleting *RSC2*. The most typical example is *PDC5* (Figure 3.26). As mentioned before, some TCA cycles genes also show slightly higher levels of mRNA in the first sample of the *rsc2Δ* strain. They could be the secondary effects of deleting *RSC2*. One explanation is that some genes may be upregulated to compensate for the loss of function of certain genes. Alternatively, the RSC complex may also have inhibitory effects on transcription. All these observations suggest that the Rsc2 isoform of the RSC complex may influence gene expression via multiple mechanisms. For a more detailed discussion, please refer to section 3.3.2.

To conclude, the deletion of *RSC2* can cause changes in the temporal mRNA profiles of many genes in the central metabolic pathway. Among them, *ACO2* is most likely to be the downstream target in the acetylation-recruitment pathway.

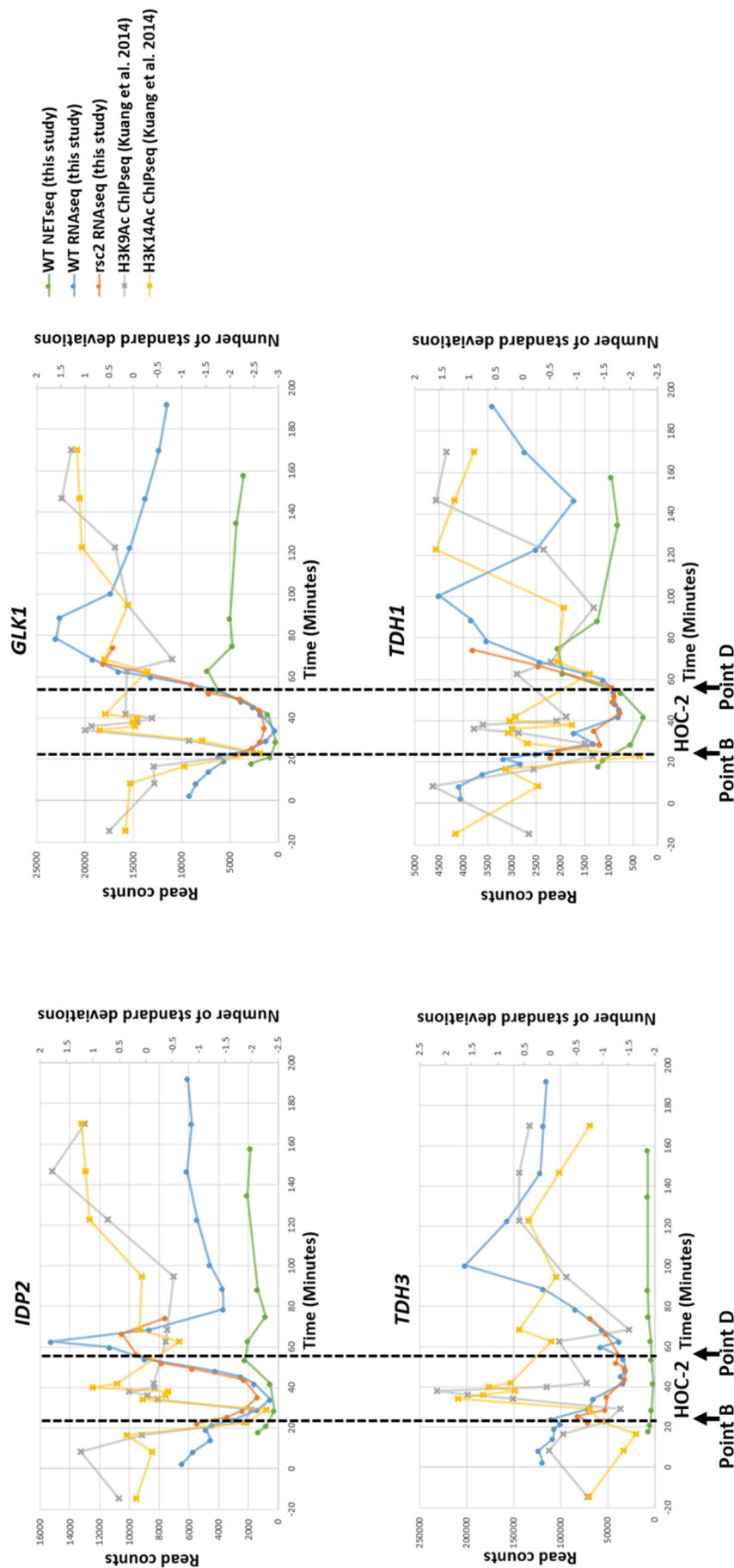


Figure 3.22 The temporal changes in the levels of transcription, transcripts and histone acetylation of *GLK1*, *IDP2*, *TDH1* and *TDH3* during the YMC. The levels of transcription for the wild-type (blue curve) were assessed by NET-seq. The amounts of transcripts for the wild-type (blue curve) and the *rsc2Δ* mutant (red curve) were measured by RNA-seq. The transcript levels for the wild-type and the *rsc2Δ* mutant are comparable, because of the normalization by RNA spike-in. The signals of NET-seq and RNA-seq are expressed as the number of reads (the y-axis on the left hand side). The ChIP-seq signals for histone acetylation (Kuang et al. 2014) are expressed as the number of standard deviations from the median (the y-axis on the right hand side).

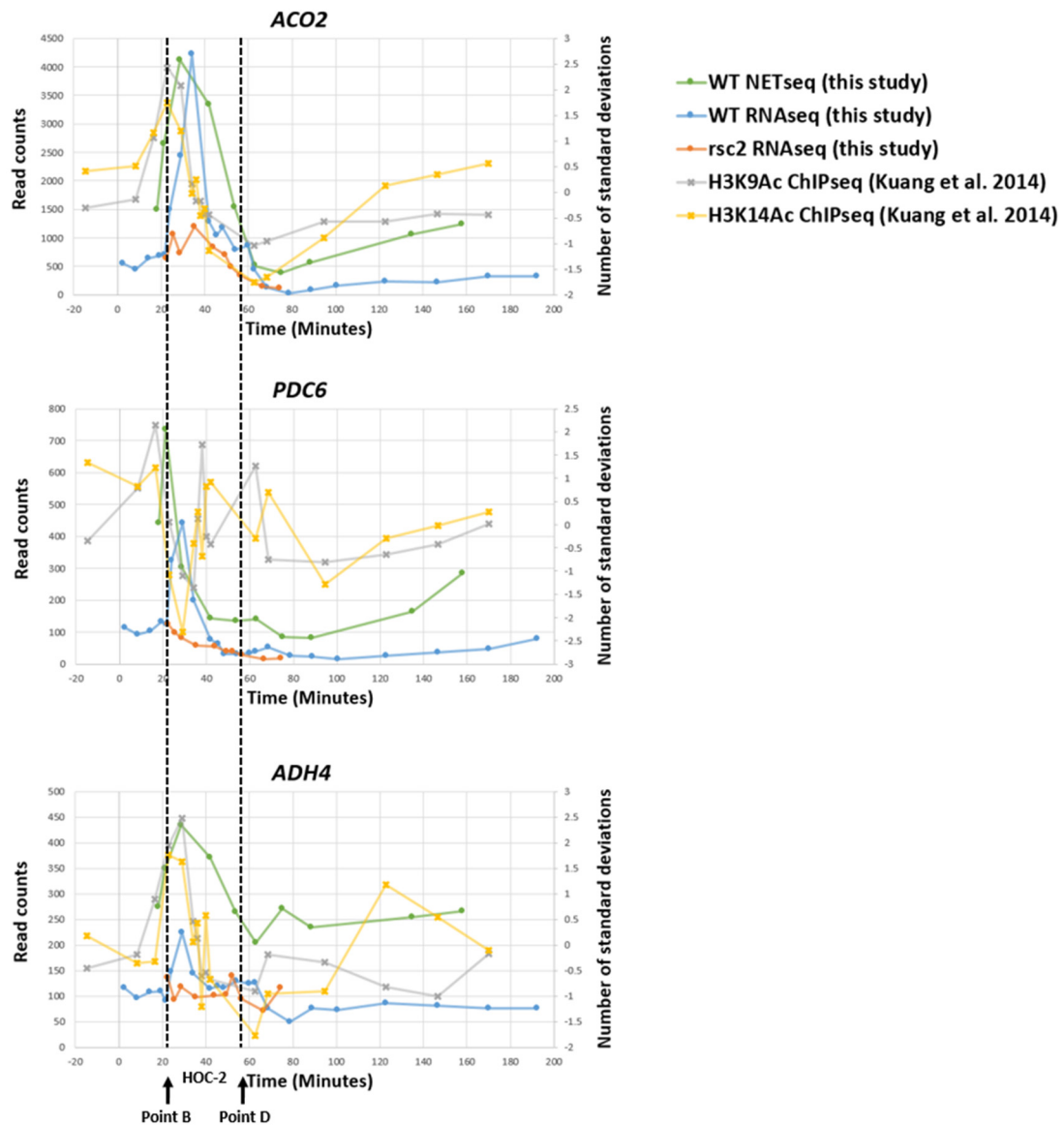


Figure 3.23 The temporal changes in the levels of transcription, transcripts and histone acetylation of *ACO2*, *PDC6* and *ADH4* during the YMC. The levels of transcription for the wild-type (blue curve) were assessed by NET-seq. The amounts of transcripts for the wild-type (blue curve) and the *rsc2Δ* mutant (red curve) were measured by RNA-seq. The transcript levels for the wild-type and the *rsc2Δ* mutant are comparable, because of the normalization by RNA spike-in. The signals of NET-seq and RNA-seq are expressed as the number of reads (the y-axis on the left hand side). The ChIP-seq signals for histone acetylation (Kuang et al. 2014) are expressed as the number of standard deviations from the median (the y-axis on the right hand side).

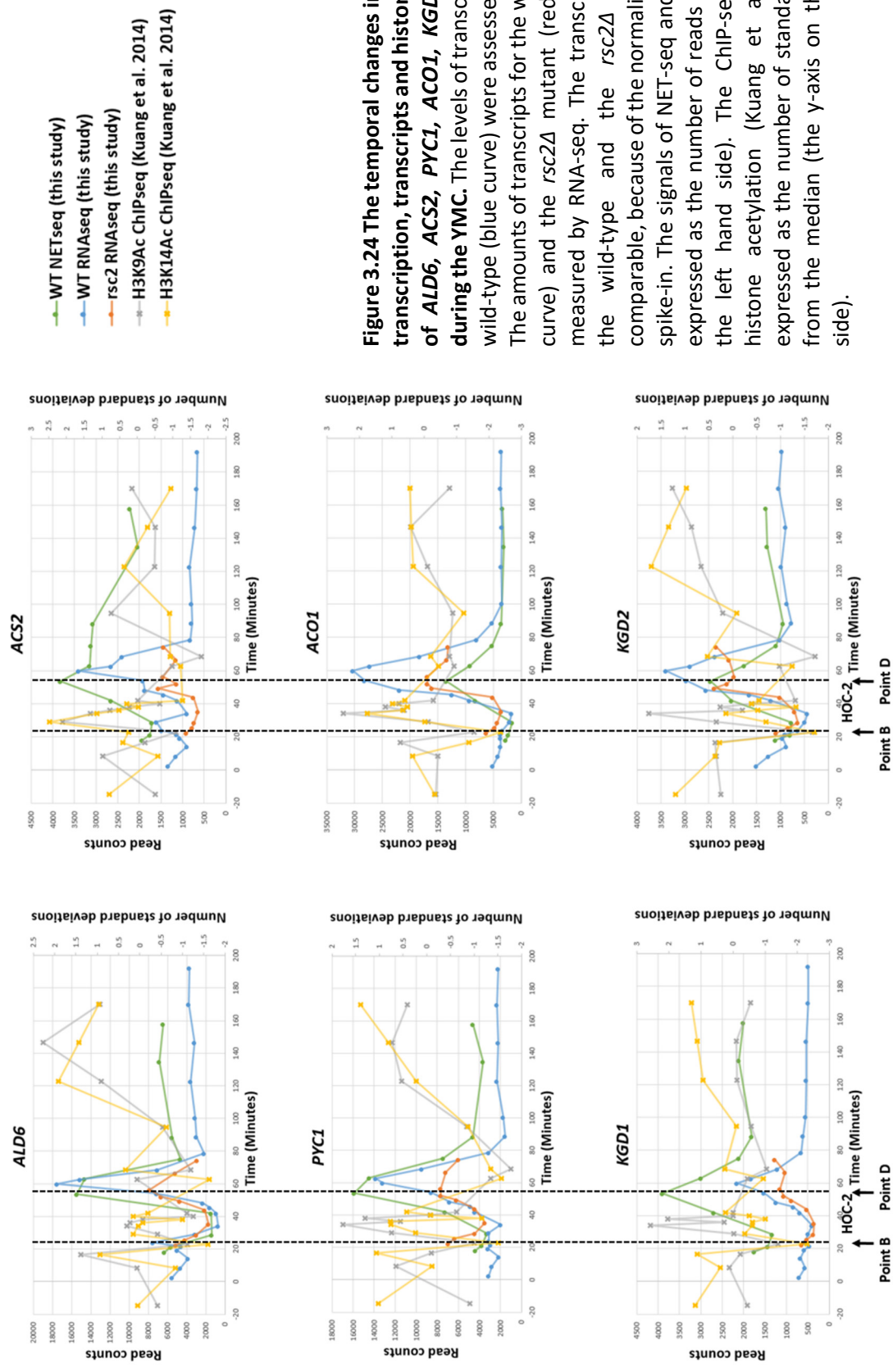


Figure 3.24 The temporal changes in the levels of transcription, transcripts and histone acetylation of *ALD6*, *ACS2*, *PYC1*, *ACO1*, *KGD1* and *KGD2* during the YMC. The levels of transcription for the wild-type (blue curve) were assessed by NET-seq. The amounts of transcripts for the wild-type (blue curve) and the *rsc2Δ* mutant (red curve) were measured by RNA-seq. The transcript levels for the wild-type and the *rsc2Δ* mutant are comparable, because of the normalization by RNA spike-in. The signals of NET-seq and RNA-seq are expressed as the number of reads (the y-axis on the left hand side). The ChIP-seq signals for histone acetylation (Kuang et al. 2014) are expressed as the number of standard deviations from the median (the y-axis on the right hand side).

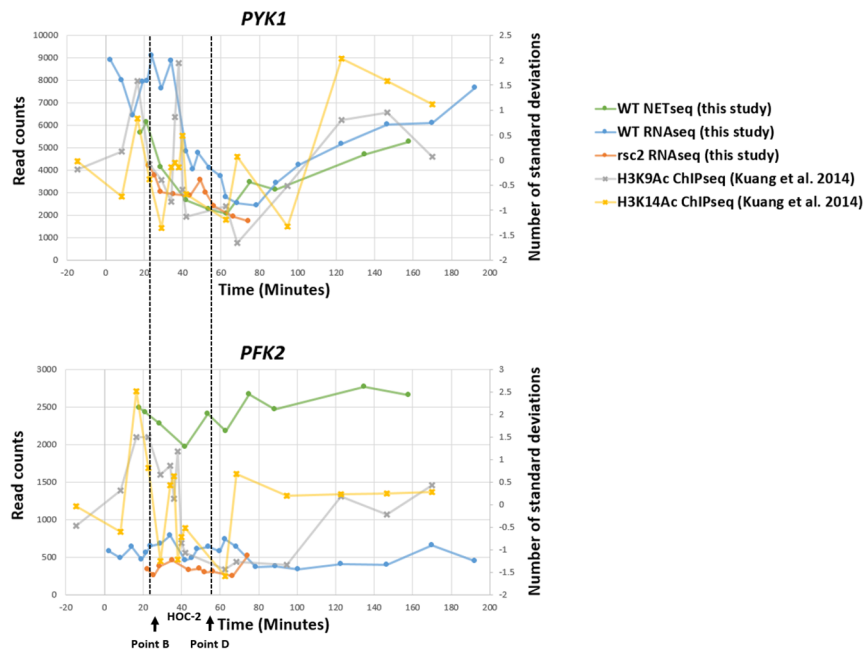


Figure 3.25 The temporal changes in the levels of transcription, transcripts and histone acetylation of *PYK1* and *PFK2* during the YMC. The levels of transcription for the wild-type (blue curve) were assessed by NET-seq. The amounts of transcripts for the wild-type (blue curve) and the *rsc2Δ* mutant (red curve) were measured by RNA-seq. The transcript levels for the wild-type and the *rsc2Δ* mutant are comparable, because of the normalization by RNA spike-in. The signals of NET-seq and RNA-seq are expressed as the number of reads (the y-axis on the left hand side). The ChIP-seq signals for histone acetylation (Kuang et al. 2014) are expressed as the number of standard deviations from the median (the y-axis on the right hand side).

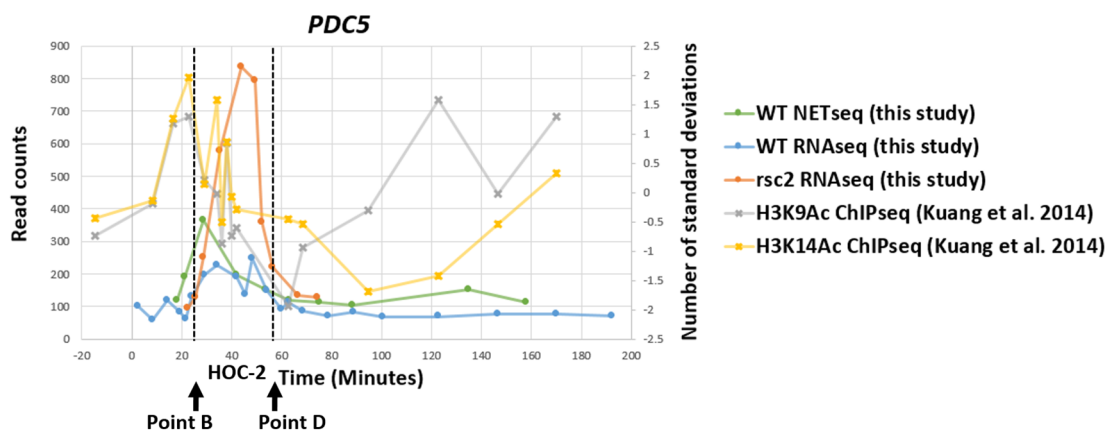


Figure 3.26 The temporal changes in the levels of transcription, transcripts and histone acetylation of *PDC5* during the YMC. The levels of transcription for the wild-type (blue curve) were assessed by NET-seq. The amounts of transcripts for the wild-type (blue curve) and the *rsc2Δ* mutant (red curve) were measured by RNA-seq. The transcript levels for the wild-type and the *rsc2Δ* mutant are comparable, because of the normalization by RNA spike-in. The signals of NET-seq and RNA-seq are expressed as the number of reads (the y-axis on the left hand side). The ChIP-seq signals for histone acetylation (Kuang et al. 2014) are expressed as the number of standard deviations from the median (the y-axis on the right hand side).

3.2.14 The expression of *ACO2* is sensitive to the deletion of genes required for histone H3 acetylation

To further test whether *ACO2* is downstream of the hypothesized acetylation-recruitment pathway, this section investigated whether the expression of *ACO2* was downregulated if genes required for histone H3 acetylation were deleted. Histone H3 can be acetylated by the SAGA/ADA complex. The deletion of several subunits of the SAGA complex, including *Ada2*, *Gcn5*, *Spt20* and *Ngg1*, is known to greatly reduce the cellular levels of histone H3 acetylation (Peng et al. 2008). Additionally, the deletion of *NOT4* has also been shown to cause hypoacetylation of histone H3 at lysine 9, 14, 18 and 23, where the RSC complex can bind (Peng et al. 2008). If *ACO2* is downstream of the hypothesized acetylation-recruitment pathway, the deletion of any one of the five genes (*ADA2*, *GCN5*, *SPT20*, *NGG1* and *NOT4*) should lead to significant downregulation of the transcript level of *ACO2*.

The transcriptomes of the five null mutants were found in the deletome of budding yeast published by Holstege lab in 2014 (Kemmeren et al. 2014). The transcriptome of each mutant strain was measured three times by microarray analysis to determine the significance level for each gene. The log₂ fold changes for genes in the central metabolic pathway are summarized in Figure 3.27. To minimize the false positive rate, only changes with a p-value smaller than 1×10^{-7} were considered significant. Significantly overexpressed or underexpressed genes in each mutant are highlighted with blue and red colours respectively. Strikingly, the expression of *ACO2* is significantly decreased in all five mutants. None of the other genes in glycolysis and TCA cycle have the same behaviour as *ACO2*. Many TCA cycle genes are even significantly overexpressed rather than underexpressed in some of the mutants, which is consistent with their elevated mRNA levels observed in the initial four sampling points of the *rsc2Δ* strain. These data suggest that the expression of *ACO2* is more sensitive to the deletion of gene required for histone H3 acetylation than most of genes in the central metabolic pathway.

		Mutant				
		<i>ada2Δ</i>	<i>gcn5Δ</i>	<i>spt20Δ</i>	<i>ngg1Δ</i>	<i>not4Δ</i>
Glycolysis	HXK1	1.864164	0.137084	0.872088	1.75739	2.480689
	HXK2	0.374511	-0.22148	-0.99717	-0.21374	-0.05825
	GLK1	1.724262	1.106529	-0.27595	1.68503	2.025164
	PGI1	0.813525	0.742558	0.824809	0.817918	0.656011
	PFK1	0.432553	0.388542	-0.38602	0.381416	0.044506
	PFK2	0.340858	0.319745	-0.03516	0.147294	-0.32463
	FBP1	-0.00433	-0.04541	0.070266	0.044534	0.280232
	FBA1	0.578649	0.213553	0.177892	0.88276	0.21867
	TPI1	0.245423	0.164323	-0.64616	-0.06525	-0.24413
	TDH1	0.59907	0.287437	-0.21502	0.374199	-0.07486
	TDH2	1.525092	0.783004	1.095769	0.687777	0.801854
	TDH3	0.769552	0.693159	1.348329	0.493709	0.522267
	PGK1	1.480849	0.816682	0.703827	0.443787	0.726203
	GPM1	1.731344	0.082267	1.76615	1.442721	-2.16044
	GPM2	0.305515	-0.17086	-0.2401	-0.00658	1.060352
	GPM3	-0.09763	-0.04517	-0.46681	-0.1145	-0.66425
	ENO1	0.170002	0.072902	-0.19519	-0.17277	0.351821
	ENO2	1.455607	0.88616	-0.04316	0.811233	0.655141
	PYK1	1.350118	0.531309	0.637898	0.836506	-0.0896
	PYK2	0.227002	-0.27405	0.302196	0.234955	1.268566
	PDA1	0.266001	0.153377	0.233005	0.211585	0.007755
	PDB1	0.004882	-0.08752	0.33026	0.3706	-0.54072
	PDX1	0.181751	0.308233	0.043554	0.012097	0.38228
	LPD1	0.389137	0.445133	0.551777	0.460239	0.090935
	PDC1	0.597278	0.621027	1.062481	0.578726	0.878788
	PDC5	0.626168	0.340676	0.416094	0.369673	0.168295
	PDC6	0.582421	0.585799	0.082675	0.582277	0.140072
	ADH1	0.938007	0.555001	-0.28004	0.3073	0.110951
	ADH2	0.717298	0.344806	-0.98707	0.35035	0.205292
	ADH3	0.063889	0.045611	-1.04878	-0.17242	-0.47867
	ADH4	-0.25968	-0.01634	-0.23981	-0.18384	-0.72187
	ADH5	0.424256	-0.15775	-0.76859	-0.22544	-0.48571
	ALD2	0.194364	0.240852	0.536451	0.271462	0.340312
	ALD3	-0.05021	-0.16568	0.135798	0.706889	2.693663
	ALD4	1.315532	0.966408	-0.15811	1.736687	2.260673
	ALD5	-0.58745	0.142204	-1.29203	0.062923	-1.09036
	ALD6	1.504839	1.200743	0.965574	1.717807	-0.3757
	ACS1	0.290731	0.109708	0.078309	0.172884	-0.16344
	ACS2	-0.44211	0.154026	-0.34613	-0.07841	-0.77884
	PCK1	0.379377	0.43517	0.475313	0.47834	0.187248
	PYC1	0.268542	-0.12662	-0.1194	0.270015	0.763216
	PYC2	0.281894	0.02165	-0.30446	0.309619	-0.86357
TCA cycle	CIT1	-0.08106	-0.30449	0.925939	-0.0233	0.163035
	ACO1	1.588565	0.191803	-0.10188	0.671629	0.444731
	ACO2	-0.69518	-0.42891	-0.49626	-0.62648	-1.29611
	IDP1	0.207206	0.578108	1.099948	0.421555	-0.89603
	IDP2	0.150651	-0.1073	-0.13888	-0.05173	0.517205
	IDP3	-0.12933	-0.19251	0.142479	-0.08553	-0.2801
	IDH1	0.755511	-0.45611	-0.33309	-0.52339	0.274983
	IDH2	1.129685	-0.08313	0.125056	0.221395	0.056486
	KGD1	0.266781	0.379421	0.808039	0.402679	-0.34314
	KGD2	0.101153	0.04235	0.598117	0.17751	0.00992
	LPD1	0.389137	0.445133	0.551777	0.460239	0.090935
	LSC1	0.257988	0.340665	0.647655	0.459169	0.084391
	LSC2	0.928192	0.61105	0.58796	0.930321	0.626972
	SDH1	0.189742	0.639649	1.090447	0.851551	-0.66492
	SDH2	0.103437	0.387296	1.588335	0.830985	-0.25858
	SDH3	0.139181	0.210277	0.807586	0.289551	-0.16438
	SDH4	0.266697	0.406123	1.32671	0.513026	-0.21736
	FUM1	0.10048	0.011896	0.563031	0.023028	-0.27983
	MDH1	0.050843	0.004868	0.236168	0.234771	-0.38481

Figure 3.27 The log2 fold changes in the expression levels of genes in the central metabolic pathways in response to the deletions that perturb histone acetylation. The significantly ($p < 10^{-7}$) overexpressed or underexpressed genes in a certain deletion mutant are colored blue and red respectively.

3.2.15 The protein levels of Aco2p is increased after passing point B and returned to the ground level after point C

So far, the expression of *ACO2* has been examined at the level of transcription and transcripts. However, it is not known whether the protein level of Aco2 increases after passing point B. Some of the literature shows that an increase in mRNA is not always translated into a corresponding increase in its corresponding protein, particularly when its protein is highly abundant (Lahtvee et al. 2017). Hence, it is important to examine the expression of *ACO2* across the YMC at the level of protein.

In this study, 9 samples were taken across a cycle of the HOC phase (Figure 3.28a). The proteome of each sample was measured by mass spectrometry with tandem mass tag (TMT) labelling (Thompson et al. 2003). The TMT system labeled the peptides from the 9 samples with 9 different chemical tags which share the same structure but have different masses due to the presence of isotopes. It allowed the relative quantification of protein levels across different samples. The levels of Aco2 are shown in Figure 3.28b. In the beginning, there are no obvious changes in the protein levels of Aco2 from sample 1 to sample 3, although the oxygen consumption rate of the cells has increased by 38%. After passing point B, sample 4 shows a rapid increase in the protein level of Aco2, which is consistent with the dramatic increase in the *ACO2* mRNA level demonstrated in section 3.2.13. Albeit a slight decrease from sample 4 to 6, Aco2 protein is still maintained at a relatively high level. Following that, both the cellular oxygen consumption rate and the Aco2 protein levels decrease dramatically from sample 6 to 8 and eventually reach their ground levels.

To summarize, the elevation of Aco2 protein levels coincides with the dramatic increase in its transcript level. Furthermore, after passing the point B, the changes in Aco2 protein levels seem to correlate with the changes in cellular oxygen consumption rates. Both indicate that Aco2 might be downstream of the acetylation-recruitment pathway.

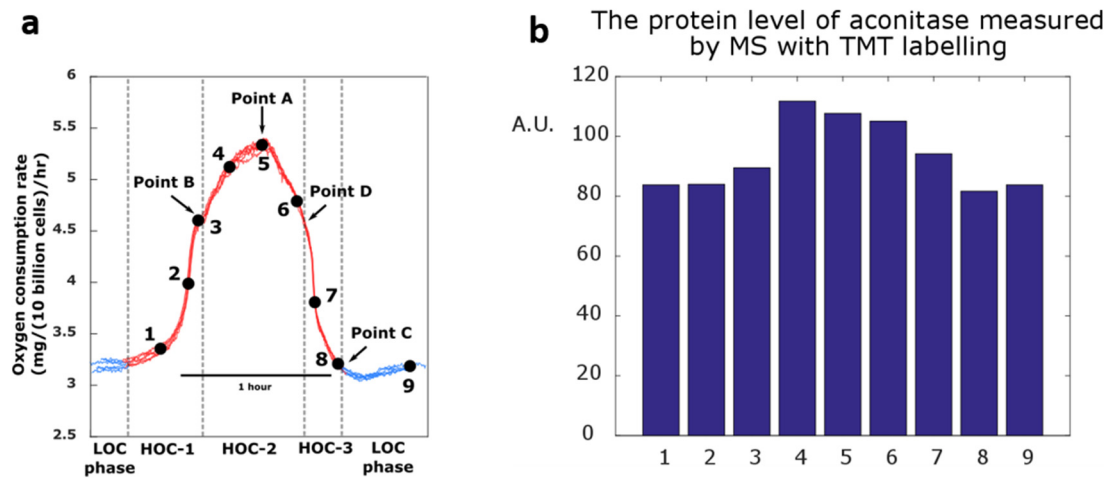


Figure 3.28 The protein levels of Aco2 during the YMC were measured by mass spectrometry with TMT labelling. **a.** The sampling time points during the YMC. **b.** The protein levels of Aco2 during the YMC.

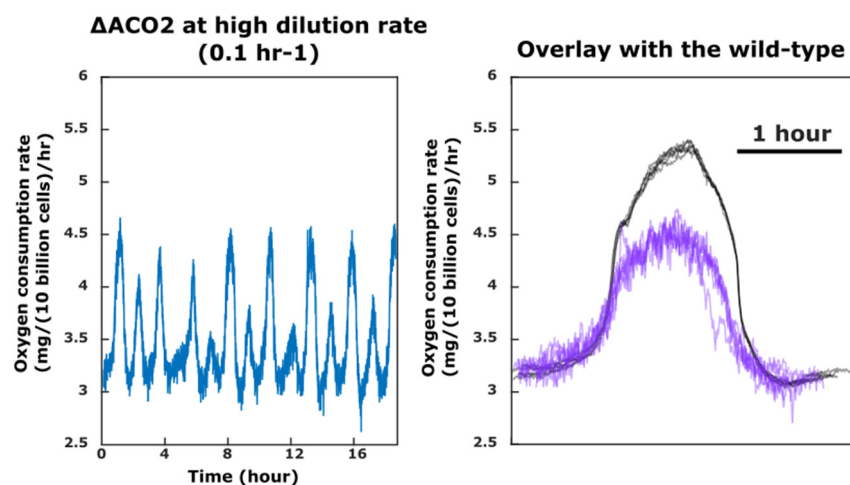


Figure 3.29 The oxygen consumption profiles for the *aco2Δ* strain at the high dilution rate. In the right panel, the curves of the *aco2Δ* strain (indigo) at the high dilution rate is overlaid with that of the wild-type (black) at the same dilution rate. The black bar represents the length of one hour.

3.2.16 The *aco2Δ* strain is unable to further increase cellular oxygen consumption rates after passing point B

To check whether Aco2 has certain functions in the HOC-2 phase, the *aco2Δ* strain was grown in the YMC conditions at the high dilution rate (0.01 hr^{-1}). Its curve of oxygen consumption rate was compared to that of the wild-type (Figure 3.29). Interestingly, the maximal oxygen consumption rate of the *aco2Δ* strain cannot reach the level marked by point B of the wild-type, which is reminiscent of the phenotypes of *rsc2-BD1Δ* and *rsc2-BAHΔ*. In addition, the LOC phase of the *aco2Δ* strain has a small peak of oxygen consumption rate, which is also similar to the *rsc2-BAHΔ* strain. To summarize, Aco2 is required for the HOC-2 phase of the YMC. The similarity between the phenotypes of the *aco2Δ* strain and some *RSC2* mutants supports the view that *ACO2* is one of the downstream targets of the RSC complex.

3.2.17 The defects seen in the *aco2Δ* strain can be partially rescued by adding excess ferrous ions

Aco2 is an aconitase primarily localized in the mitochondria of *S.cerevisiae* and can bind to iron-sulfur clusters (Huh et al. 2003; Bradley et al. 2019). In mammals, it has been shown that the cytosolic aconitase has multiple functions. It can function as a metabolic enzyme when it binds to iron-sulfur clusters, or as a regulator of cellular ferrous iron levels, when it dissociates with iron-sulfur clusters (Narahari et al. 2000; Kennedy et al. 1992). To further investigate the role of Aco2 during the HOC-2 phase, excess ferrous ions were added to the culture medium of the *aco2Δ* strain.

Interestingly, the defective HOC-2 phase seen in the *aco2Δ* strain can be partially rescued by adding excess ferrous ions (Figure 3.30). After the treatment, the maximal oxygen consumption rate gradually increases in the following 10 hours and then the YMC stabilizes. In the presence of excess ferrous ions, the oxygen consumption rate is able to further increase after passing point B. However, before arriving at point A of the wild-type, the oxygen

consumption level of the *aco2Δ* strain starts to decrease prematurely. In the rest of the HOC-2 phase, the oxygen consumption level of the *aco2Δ* strain is always lower than that of the wild-type until the HOC-3 phase starts. All these observations suggest that the function of Aco2 in the early HOC-2 phase might be related to the regulation of intracellular levels of ferrous ions. The partial rescue result indicates that regulating cellular ferrous ion levels is probably not the only way for Aco2 to regulate the oxygen consumption rate during the HOC-2 phase.

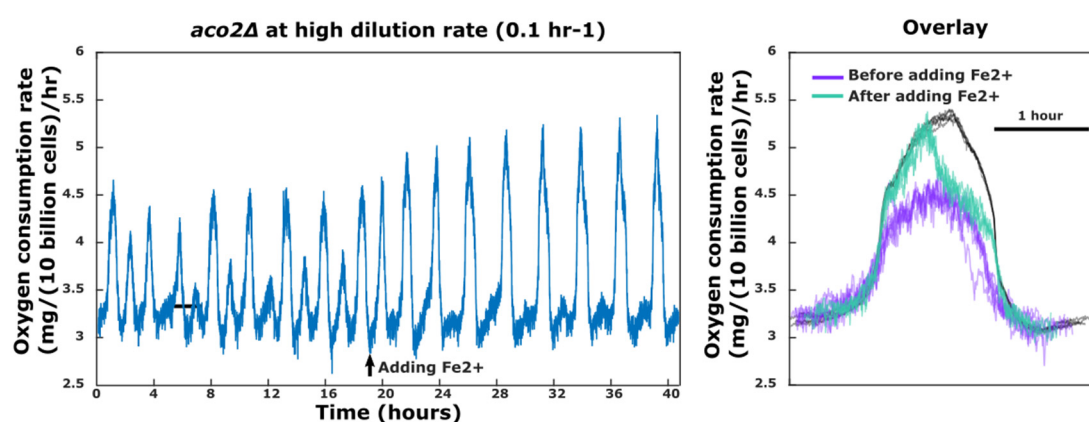


Figure 3.30 The responses of the *aco2Δ* strain to the addition of excess ferrous (Fe^{2+}) ions. In the left panel, the time point when the Fe^{2+} solution was added is marked by a black arrow. In the right panel, the oxygen consumption curves for the *aco2Δ* strain before (indigo) and after (green) adding the Fe^{2+} solution are overlaid with the wild-type curves (black). The black bar represents the length of one hour.

3.2.18 Chelating extracellular $\text{Fe}^{2+}/\text{Fe}^{3+}$ ions in the wild-type strain can mimic the effect of deleting *ACO2*

If enough ferrous ions are required for the further increase in the cellular oxygen consumption rate after passing point B, it is reasonable to hypothesize that removing extracellular $\text{Fe}^{2+}/\text{Fe}^{3+}$ ions may produce a phenotype that is similar to the *aco2Δ* strain. To test this hypothesis, the iron chelator, deferiprone, was added to the culturing medium of the wild-type in a stoichiometric ratio of the extracellular concentration of $\text{Fe}^{2+}/\text{Fe}^{3+}$ ions. Deferiprone can chelate $\text{Fe}^{2+}/\text{Fe}^{3+}$ ions and prevent the cells from uptaking these ions (Hider & Hoffbrand 2018).

After the addition of deferiprone, the maximal oxygen consumption rate is reduced in the next cycle of the YMC, which is similar to the phenotypes of *aco2Δ*, *rsc2-BD1Δ* and *rsc2-BAHΔ*. Apart from that, the cycling frequency of the wild-type also seems to be increased. Interestingly, once excess ferrous ions are added, these defects can be corrected in the following cycle of the YMC (Figure 3.31). Hence, $\text{Fe}^{2+}/\text{Fe}^{3+}$ ions are required for the HOC-2 phase.

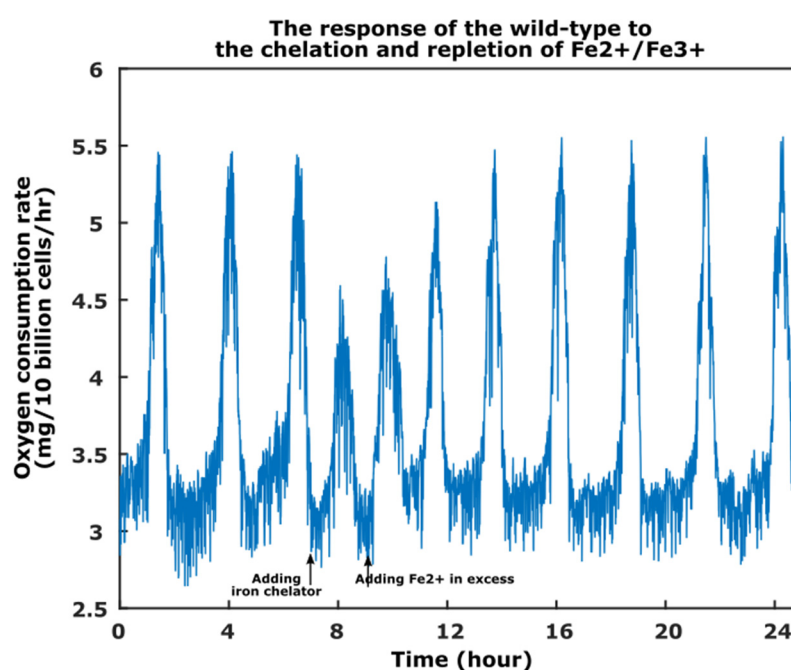


Figure 3.31 The responses of the wild-type to the removal and addition of Fe^{2+} to the culture medium. The points of adding iron chelator and the Fe^{2+} solution are marked by two arrows. This experiment was carried out on an old fermenter. The electrical noise of the DO sensor was too high to perform the overlay procedure.

3.2.19 Excess ferrous ions can rescue *rsc2-BD1Δ* and *rsc2-BAHΔ* as well

aco2Δ, *rsc2-BD1Δ* and *rsc2-BAHΔ* have similar phenotypes during the YMC. Excess ferrous ions can partially rescue *aco2Δ*. It is reasonable to hypothesize that *ACO2* is the downstream target of the RSC complex and the phenotypes of *rsc2-BD1Δ* and *rsc2-BAHΔ* are caused by the disrupted *ACO2* expression. To test this hypothesis, excess ferrous ions were added to the medium of *rsc2-BD1Δ* and *rsc2-BAHΔ* to see if the defective HOC-2 phase could be corrected.

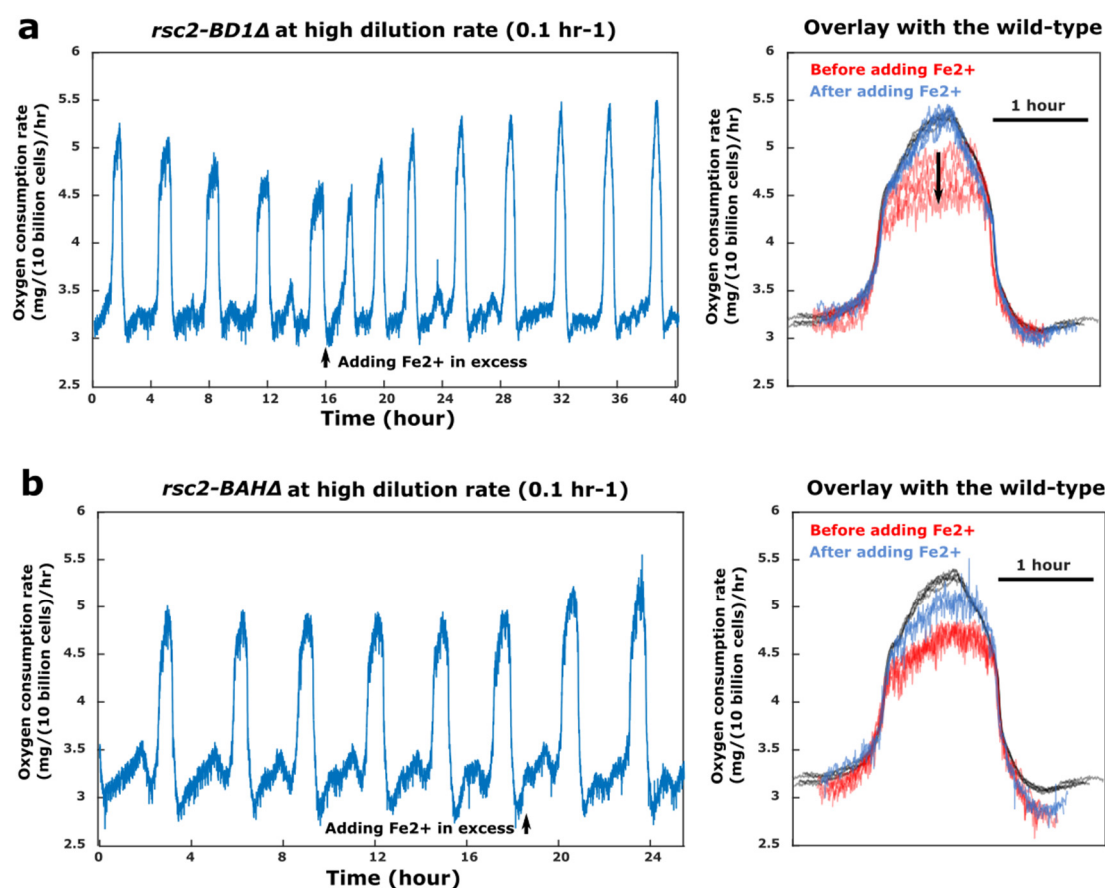


Figure 3.32 The responses of the RSC mutants to the addition of the Fe²⁺ solution. **a.** The responses of the *rsc2-BD1Δ* strain to the addition of the Fe²⁺ solution. **b.** The responses of the *rsc2-BAHΔ* strain to the addition of the Fe²⁺ solution. In the right panels, the oxygen consumption profiles of the RSC mutants before (red) and after (blue) adding Fe²⁺ are compared to the wild-type (black).

The results largely support the hypothesis. For *rsc2-BD1Δ* (Figure 3.32a), adding excess ferrous ions to the culture medium allows the maximal oxygen consumption rate to gradually restore the level of the wild-type, which is similar to the response seen in the *aco2Δ* strain. However, instead of the partial rescuing effect seen in the *aco2Δ* strain, adding excess ferrous ions almost enables the HOC-2 phase of *rsc2-BD1Δ* to overlap with that of the wild-type. The difference can be explained by the residual level of Aco2p in the *rsc2-BD1Δ* strain. For *rsc2-BAHΔ* (Figure 3.32b), the addition of ferrous ions seems to elevate the maximal level of oxygen consumption rate as well. However, the YMC of the cells was stabilized until the end of the experiment. It is hard to conclude whether the addition of ferrous ions can achieve a complete rescue of the *rsc2-BAHΔ* strain. For mutants which show the complete loss of the YMC, such

as *rsc2-CT1Δ* and *rsc2Δ*, adding excess ferrous ions do not seem to produce any effects (data not shown), which indicates that the RSC complex may have some other downstream targets which are critical for the periodic metabolic oscillation and they cannot be rescued by adding ferrous ions.

To conclude, the experimental results support the view that *ACO2* expression is activated via the acetylation recruitment pathway to regulate the entry of the HOC-2 phase.

3.3 Discussion

3.3.1 The possible mechanisms of the RSC complex in bridging acetylation and gene expression

In this study, the importance of the Rsc2 isoform of the RSC complex for the YMC was first identified by testing yeast strains bearing various mutations (domain deletions, full-length deletions and point mutations) on the subunits of the RSC complex in the YMC conditions. The strong YMC defects seen in the *rsc2Δ* strain but not in the *rsc1Δ* strain justified the hypothesis that the Rsc2 subunit of the RSC complex is required for the YMC. Furthermore, by examining the c-terminal domain deletion of Rsc2 (which is responsible for the physical interaction between the Rsc2 subunit and the RSC complex), it was found that the Rsc2 subunit itself can not support the YMC and the rest of the RSC complex was required. Following that, yeast strains carrying individual domain deletions of Rsc2 were tested in the YMC conditions. All of them have YMC defects of various degrees, but none had the same severity as the full-length deletion (For more discussion of this, please refer to section 3.3.3). Particularly, the deletion of either the first bromodomain or the BAH domain of Rsc2 could lead to defects in the HOC-2 phase of the YMC. These results suggest that all domains of Rsc2 are required for the YMC, but none are deterministic. The strong YMC defects of the full-length deletion are likely to be a combinatorial effect of the individual domain deletions of Rsc2. To summarize, the Rsc2 isoform of the RSC complex plays an important role in the YMC.

ACO2 and *CLN3* were found to be potential downstream players of the Rsc2 isoform of the RSC complex during the YMC by comparing the temporally resolved transcriptome of the wild-type and the *rsc2Δ* strain. Further experiments were carried out to validate *ACO2* as one of the downstream targets of the RSC complex as well as its importance for the YMC. Like the *rsc2-BD1Δ* and *rsc2-BAHΔ* strain, the *aco2Δ* strain also had a defect in entering the HOC-2 phase of the YMC. Importantly, this defect of all the three strains could be either completely

or at least partially rescued by adding excess ferrous ions to the culture medium. This strongly supports two arguments. First, *ACO2* is one of the downstream targets of the Rsc2 isoform of the RSC complex during the YMC. Second, Aco2 may be required for the entry of the HOC-2 phase of the YMC or the further increase in cellular oxygen consumption rates. Currently, the exact mechanism downstream of Aco2 is not known. One possibility is that Aco2 may increase cellular oxygen consumption rates by upregulating the cellular availability of ferrous ions. More investigations are required to find out the function of Aco2 during the HOC-2 phase.

In terms of upstream factors of the RSC complex during the YMC, three hypotheses are proposed here. The first hypothesis is that, at the end of the HOC-1 phase, the increased acetylation levels on histone H3 recruits the RSC complex to the target genes. Then the RSC complex activates the expression of these genes. Although currently there is no direct evidence to prove it, this hypothesis is favored by some experiments in the literature and the bioinformatic analysis carried out in this study. First, a lot of studies have shown that the RSC complex can bind to acetylated histone H3, particularly at the acetylated K9, K14, K18 and K23 residues (VanDemark, Kasten, Ferris, Heroux, Hill & Cairns 2007; Kasten et al. 2004; Othman et al. 2007; Chatterjee et al. 2011). Second, both the penta-lysine-to-arginine mutations (mimicking the nonacetylated state) on histone H3 and the deletion of *GCN5* can produce the same phenotype as the *rsc2Δ* strain (Kuang et al. 2014; Cai et al. 2011). Third, there are peaks of histone H3K9, 14 and 18 acetylation in the promoter regions of more than 1000 genes at the end of the HOC-1 phase, which has been revealed by the ChIP-seq data (Cai et al. 2011; Kuang et al. 2014). Fourth, the loss of the bromodomain of Rsc2 can cause defects in the entry of the HOC-2 phase. Fifth, the bioinformatic analysis in this study shows that if a gene has a peak of H3K9, 14 and 18Ac at point B, its peak of transcripts at the beginning of the HOC-2 phase is significantly more likely to be lost in the *rsc2Δ* strain. Although these are five pieces of indirect supporting evidence, the direct evidence is still missing.

The second hypothesis is that the RSC complex is recruited to the chromatin via acetylated non-histone proteins, such as the acetylated SAGA complex or certain acetylated transcription factors. The SAGA complex can also be acetylated by Gcn5 at its Ada3, Sgf73 and Spt7 subunits (Cai et al. 2011). If the acetylation marks on the SAGA complex can always co-occur with histone H3 acetylation during the YMC, five out of the six pieces of the indirect evidence of the first hypothesis can still be easily explained. However, it is hard to use this hypothesis to explain why the YMC is abolished in the strain carrying the penta-lysine-to-arginine mutations on histone H3. Apart from the SAGA complex, transcription factors can potentially be the acetylated non-histone protein in this hypothesis as well. Although not in budding yeast, there are some examples in mammals showing the physical interaction between bromodomains and acetylated transcription factors. For instance, the acetylated transcription factor GATA1 can be recognized by the bromodomain of Brd3, which recruits GATA1 to the chromatin (Lamonica et al. 2011).

The third hypothesis assumes that the recruitment of the RSC complex to the chromatin is not significantly changed during the YMC. Instead, the remodeling activity of the RSC complex is upregulated by histone H3 acetylation at the end of the HOC-1 phase. There is some in vitro evidence showing that the acetylated histone tails can enhance the remodeling activity of the RSC complex (Chatterjee et al. 2011; Skiniotis et al. 2007; Ferreira et al. 2007). Although most of the evidence for the first hypothesis can still be explained, this hypothesis still has a major problem, which makes it become less attractive. The remodeling activity of the RSC complex is only shown to be slightly elevated by the acetylated histone tail (Chatterjee et al. 2011). If it is the major mechanism, it is hard to explain why the transcript level of *ACO2* is dramatically increased upon entering the HOC-2 phase in the wild-type strain. A more likely scenario could be that the enhancement of remodeling activity is an additional but minor mechanism in the former two hypotheses to achieve tighter temporal control of gene expression.

To test the three hypotheses, a ChIP-seq experiment is proposed here. Its experimental design is briefly described below. First, the Rsc2 subunit of the RSC complex in both the wild-type and the histone H3 penta-lysine-to-arginine mutant (the K-to-R strain hereafter) needs to be tagged by the FLAG peptide. Then for the wild-type, the HOC-1 and HOC-2 phases of the YMC are sampled at multiple time points. Each of the samples is split into two parts immediately after being taken from the fermenter. One part is used for making RNA-seq libraries, which can be used to map the sampling time points to a uniform time axis. The other part is immediately crosslinked for 5 mins by formaldehyde. After that, the Rsc2 ChIP-seq library can be constructed by utilizing anti-FLAG M2 antibody. For the K-to-R strain, it does not normally exhibit the YMC. However, it may be possible to induce a YMC-like response by simply removing some medium, like the treatment performed on the *rsc2Δ* strain in section 3.2.8. After the treatment, several samples can be taken in the same way. Both RNA-seq and ChIP-seq libraries are constructed for each sample as well. After sequencing and normalizing all the libraries, it is possible to obtain a temporal profile quantifying the binding of Rsc2 on the chromatin for both strains. By comparing both strains, there could be at least three possible outcomes. If the binding of Rsc2 at the promoter region of *ACO2* does not change for both strains, it indicates that the third hypothesis may be correct. That is, the activation of the downstream gene is not regulated by the recruitment of the RSC complex. If the ChIP-seq signals of Rsc2 show a peak before entering the HOC-2 phase for the wild-type but not for the K-to-R strain, it supports the first hypothesis. The RSC complex is recruited to the chromatin via histone H3 acetylation. If the ChIP-seq signals have a peak for both strains, the second hypothesis may be correct. The recruitment of the RSC complex to the chromatin could be mediated by acetylated non-histone proteins.

3.3.2 The RSC complex may activate genes in three different modes

Apart from *ACO2*, many genes also have their temporal profiles of transcripts changed in the *rsc2Δ* strain. Based on the observations and analysis in the previous sections, these genes may be categorized into at least three classes. Each class will be discussed in the following paragraphs. The three classes of genes may be associated with three potential modes that the RSC complex could influence the expression of genes. The hypothesized acetylation-recruitment pathway could merely be one of the three modes.

In terms of genes in this first class, their expression during the YMC is temporally-controlled by the RSC complex and is associated with histone H3 acetylation. Specifically, their temporal profiles of transcript levels have a narrow peak at TW1 (right after entering the HOC-2 phase) in the wild-type but not in the *rsc2Δ* strain. This peak often has a preceding peak of histone H3 acetylation. Their expression mechanism can be explained by the acetylation-recruitment pathway. That is, the acetylation of certain chromatin bound proteins (which could be histone H3 or other chromatin associated factors) causes the recruitment of the Rsc2 isoform of the RSC complex to the promoter of the genes in the first class, which further activates their expression. *ACO2* is a typical example of genes in this class (Figure 3.23).

Genes in the second class have a peak of transcript levels in TW2 and this peak is also lost in the *rsc2Δ* strain. However, the peak of transcripts does not seem to be associated with a preceding peak of histone H3 acetylation. In other words, the expression of genes in the second class is temporally-controlled by the RSC complex but not associated with histone H3 acetylation. *ALD6* is an example of genes in the second class (Figure 3.24). As mentioned in section 3.2.8, there is a possibility that the differential expression of these genes may just be a natural difference between the normal YMC and the induced YMC-like response, which is not related to the deletion of *RSC2*. However, if the expression of genes in the second class does require Rsc2, there may exist at least three possible explanations for their gene

expression mechanism. The first explanation is that the Rsc2 isoform of the RSC complex is recruited to the promoters of these genes by certain acetylated non-histone-H3 chromatin bound factors and histone H3 does not have a peak of acetylation. It can explain why a peak of histone H3 acetylation is not seen. However, it cannot explain why the non-histone factor is acetylated but histone H3 is not. The second explanation is that the Rsc2 isoform of the RSC complex is recruited to the promoter of these genes via an acetylation independent pathway. Rsc2 also has an AT-hook motif, which can bind DNA directly. The weakness of this explanation is that it cannot explain why the expression of genes in the second class is temporally-controlled. In the third explanation, these genes are indirectly controlled by the Rsc2 isoform of the RSC complex. That is, certain transcription factors are underexpressed in the *rsc2Δ* strain and they are responsible for the temporally controlled expression of the genes in the second class. More experiments are required to test all these explanations.

For genes in the third class, they seem to be underexpressed at most of the time points in the *rsc2Δ* strain. In other words, the expression of these genes is not temporally controlled by the Rsc2 isoform of the RSC complex. A good example is *PFK2* (Figure 3.25). It is possible that the Rsc2 isoform of the RSC complex can bind the promoters of genes in this class and activate their gene expression continuously.

To summarize, the Rsc2 isoform of the RSC complex could potentially activate gene expression in three different modes: H3 acetylation associated and temporally controlled (the first class of genes), not associated with H3 acetylation but temporally-controlled (the second class of genes) and not temporally-controlled (the third class of genes). This classification may help understand the phenotypes of the YMC mutants in this study.

3.3.3 The possible explanations for the RSC mutants showing defects in the YMC

This study has reported 6 strains showing obvious YMC defects. They are *rsc2Δ*, *rsc2-CT1Δ*, *rsc2-BD1Δ*, *rsc2-BD2Δ*, *rsc2-BAHΔ* and *aco2Δ*. According to the severity of the defects, they can be divided into three groups. Starting from the group showing the mildest defects, the first group contains *rsc2-BD1Δ*, *rsc2-BAHΔ* and *aco2Δ*. They have periodic oscillations between the HOC and LOC phases, but they seem to have difficulties in entering the HOC-2 phase. This defect can be rescued by adding excess ferrous ions. *rsc2-BD2Δ* belongs to the second group. It can exhibit the periodic oscillations at the medium but not the high dilution rate. The third group contains *rsc2Δ* and *rsc2-CT1Δ*. They completely lose the periodic oscillations. Their maximal oxygen consumption rates cannot even reach the level at point B. The defects in the second and the third group cannot be rescued by adding ferrous ions.

For mutants in the first group, it seems that the loss of *ACO2* expression after passing point B is enough to explain their phenotypes. *ACO2* belongs to the first class of gene described in the previous section. It implies that both BD1 and BAH domains might be required for the first mode of gene activation by the RSC complex (H3 acetylation associated and temporally controlled). For mutants in the second group, currently the molecular mechanism for producing such a phenotype is completely elusive. However, based on the analysis on the genes in the central metabolic pathways, one possible explanation is that the loss of the BD2 domain of Rsc2 may lead to the underexpression of certain glycolytic genes (such as *PFK1* and *PFK2*, which belong to the third class of genes in the previous section). The underexpression of these genes may limit and disrupt the flux of glucose through glycolysis during the HOC phase at the high dilution rate, which might further cause the loss of the periodicity. At the medium dilution rate, the flux of glucose through glycolysis during the HOC phase may be lower. Even if these genes are slightly underexpressed, they do not limit the flux at the medium dilution rate. In terms of the third group, these glycolytic genes may be

underexpressed to a greater extent, probably due to the loss of more domains of Rsc2. Consequently, they may become the limit of the flux of glucose through glycolysis at all three dilution rates. Based on this explanation, the BD2 domain of Rsc2 could be the major requirement for the normal expression of these glycolytic genes, while the BD1 and BAH domains may only have minor contributions.

It is not known why the deletion of the different domains of Rsc2 can influence the expression of different genes and lead to different phenotypes. It is possible that the BD1, BD2 and BAH domains of Rsc2 have different contributions to the three different modes of gene activation described in the previous section. Based on the limited information, it is too early to make more interpretations.

3.3.4 The ‘Sozu’ model and the molecular mechanism behind the YMC

Although the YMC has been studied by many high throughput techniques, the mechanism of the YMC itself is still largely elusive. Some mathematical models and the potentially important factors have been proposed for the YMC (Sánchez-Gaya et al. 2018; Rao & Pellegrini 2011), however, an integrative mechanistic explanation for the YMC is still missing.

To explain the mechanism of the YMC, there are two major questions that need to be answered. First, how can the cells in the YMC maintain periodic oscillations in a constant environment lacking obvious external periodical entraining factors? During the YMC, both the metabolic cycle and the cell cycle exhibit periodicity at the population level. Normally, the periodicity of a cell population requires the entrainment and synchronization by a periodical environmental cue. In other periodical oscillatory systems such as the circadian rhythm, a periodic environmental factor such as the light is required to set the phases of oscillators in different cells back to zeros at the beginning of a new cycle, which synchronizes the oscillators in different cells (Golombek & Rosenstein 2010; Husse et al. 2015). Without such entraining factors, the oscillators in different cells may start to accumulate phase differences, which can

eventually lead to the loss of the synchronization and the macroscopic periodic behaviour (Welsh et al. 2004; Carr & Whitmore 2005; Nagoshi et al. 2004). However, the fermenter provides constant aeration, pH, volume, temperature and dilution rate. There is no obvious periodical environmental factor. How is the accumulation of cell-by-cell phase differences controlled during the YMC? Second, what can cause the cells to adopt higher oxygen consumption rates during the HOC phase of the YMC? It can be further broken down into three smaller questions. That is, what causes the entry of the HOC-1 phase from the LOC phase? What is happening at the point B? And what is happening before the exit of the HOC phase? Based on all the results in this study and the findings in the literature, here I attempt to propose a testable model for the YMC, called the 'Sozu' model. Many parts of the model are based on indirect evidence, so the model is still a hypothesis and requires more experiments to validate. The following paragraphs start by introducing the 'Sozu' and how it is analogous to the YMC, followed by the molecular mechanisms at different phases of the YMC.

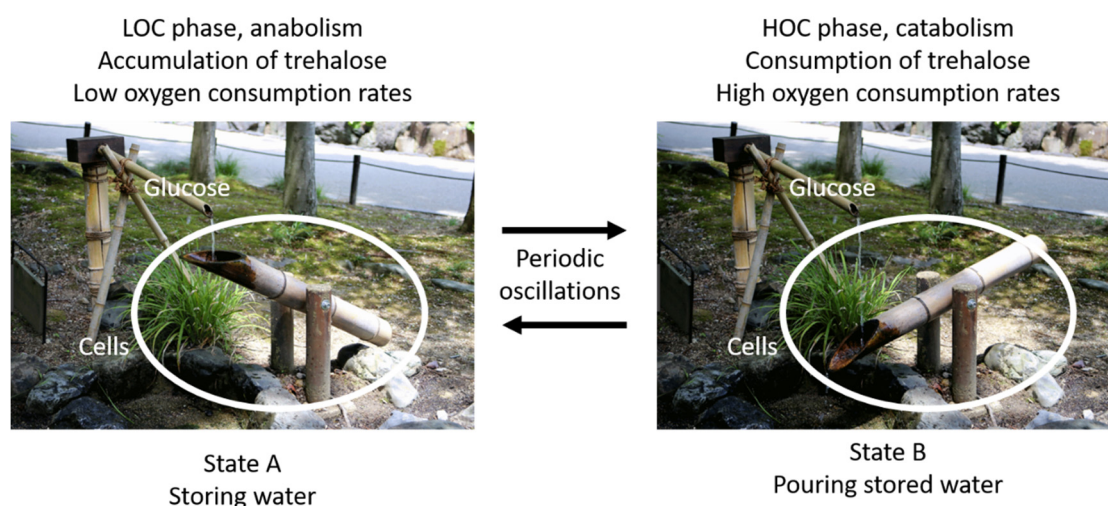


Figure 3.33 The yeast metabolic cycle is analogous to the 'Sozu', which can oscillate between two states (A and B) periodically without external entraining factors. State A is similar to the LOC phase of the YMC, while state B is similar to the HOC phase of the YMC. The 'Sozu' is analogous to the cells in the culture vessel of the fermenter.

The Sozu is a water powered device commonly used for decorating Japanese gardens (Figure 3.33). It is also referred to as 'Shishi-odoshi' or 'scaredeer', because it can frighten away animals like deers by hitting a stone and making a sound periodically. The Sozu is comprised of two major parts, a rotatable bamboo container resting on a rack and a bamboo pipe which directs water with a constant flow rate into the bamboo container. The gravity centre of the bamboo container and its rotation axis are carefully designed, so that the bamboo container can adopt two states of balance (A and B) depending on whether the amount of water in the container is above a certain threshold or not. To compare the Sozu and the YMC, the working process of the Sozu is described in the next paragraph and the corresponding concepts in the YMC are put in brackets.

At the balance state A, which is analogous to the LOC phase of the YMC, the bamboo container (the cells in the fermenter) adopts the balance state which allows water (freshly added glucose) to flow into the container. Then the water level in the container gradually increases. (Glucose is converted into trehalose and glycogen, which gradually accumulates in the cells.) In state A, the water is largely stored (anabolism and a relatively low oxygen consumption rate) rather than consumed (catabolism). After a fixed amount of time, the water level (carbon source storage) in the container will reach the threshold and change the balance to state B, which allows the bamboo container to rotate and pour the stored water out (catabolism and the increased oxygen consumption rates). Now, the balance state B is analogous to the HOC phase of the cell. After a while, as the amount of water in the bamboo container is reduced, it will stop further pouring water and restores the balance state A and start accumulating water again (Exiting the HOC phase and entering the LOC phase), which initiates a new cycle.

The Sozu is analogous to the YMC in three major aspects. First, both the Sozu and the YMC are exposed to a constant environment, but both of them can exhibit periodicity. Second, the length of the LOC phase of the YMC has been shown to be inversely proportional to the

feeding rate, while the length of the HOC phase is relatively constant at various feeding rates (Burnetti et al. 2016). Similarly, for the Sozu, the length of the balance state A is also inversely proportional to the water flow rate, while the length of the balance state B is relatively independent of the water flow rate. Finally, the cellular trehalose and glycogen levels gradually increase during the LOC phase and reach the maximum before entering the HOC phase. After that, it keeps decreasing until it reaches the minimum just before exiting the HOC phase. The increased flow of carbon sources into the central metabolic pathway can explain the increased oxygen consumption rate during the HOC phase (Shi et al. 2010). In terms of the Sozu, the temporal changes in the water levels in the bamboo container are also similar to the changes in the cellular trehalose and glycogen level during the YMC. Next, the molecular mechanism of the 'Sozu' model is discussed (Figure 3.34).

Throughout the YMC, the glucose within the fresh medium is immediately absorbed by the cells, after being added to the cell culture. It is evidenced by the fact that the glucose level in the culture is extremely low during the YMC (Porro et al. 1988; Müller et al. 2003). In the LOC phase of the YMC, part of the absorbed glucose is converted to trehalose or glycogen, which are stored in the vacuole. It is supported by the increased levels of trehalose during the LOC phase (Shi et al. 2010; Müller et al. 2003). The rest of the glucose is fed into the central metabolic pathway, which can explain the relatively low oxygen consumption rate during the LOC phase. The cellular capacity of storing trehalose has an upper limit. It was originally proposed by Futcher in 2006, albeit the existence of such a limit still requires more direct evidence to prove (Futcher 2006). Once the limited capacity is reached, a part of the absorbed glucose that is originally destined to be stored may lead to a spike of the cellular glucose level in the HOC-1 phase. As expected, such a spike of glucose can be clearly seen from both the glucose assay data and the metabolome data of YMC (Müller et al. 2003; Tu et al. 2007). The spike of glucose in the cell can trigger the cAMP signalling pathway and activate protein kinase A, which further phosphorylates and activates Nth1 (neutral trehalose hydrolase 1) (Gancedo

2008; Tamaki 2007; Robertson & Fink 1998; *Cerevisiae* Graphical et al. 2016). It is also supported by the spike of cAMP concentration and the upregulated Nth1 activity observed in the HOC-1 phase (Müller et al. 2003; Shi et al. 2010). The activated Nth1 can degrade trehalose and release more glucose from it, which may further activate the cAMP signalling and trigger a positive feedback loop. The sharp increase in the oxygen consumption rate during the HOC-1 phase may be caused by the suddenly increased flux of glucose for aerobic metabolism. Here, the increased substrate level caused by post-translational regulation is enough to explain the increased oxygen consumption rate at this stage (Note that section 3.2.9 shows that the entry of the HOC-1 phase is not accompanied by an obvious change in transcript levels).

The increased glucose flux also raises the intracellular levels of acetyl-CoA. Apart from being the substrate for the TCA cycle, it is also the substrate of histone acetyltransferase, such as Gcn5 and Esa1. Section 3.1.2.1 has introduced that the increased level of acetyl-CoA can lead to the acetylation of histone H3 and H4 at the promoters of a large number of genes (Cai et al. 2011). Based on the hypothesized acetylation-recruitment pathway, the increases in the acetylation of H3K9, 14 and 18 at the promoter regions may recruit the RSC complex. The RSC complex can then activate the transcription of a set of genes, including *CLN3* and *ACO2*. They may upregulate cellular oxygen consumption rates and allow the entry of the HOC-2 phase. However, as is shown in supplementary Figure S5 (In Appendix), there is a short time delay between the peak of transcription and the peak of transcripts. Even though the transcription of *CLN3* and *ACO2* has been activated, the cells cannot upregulate the oxygen consumption rate and enter the HOC-2 phase immediately.

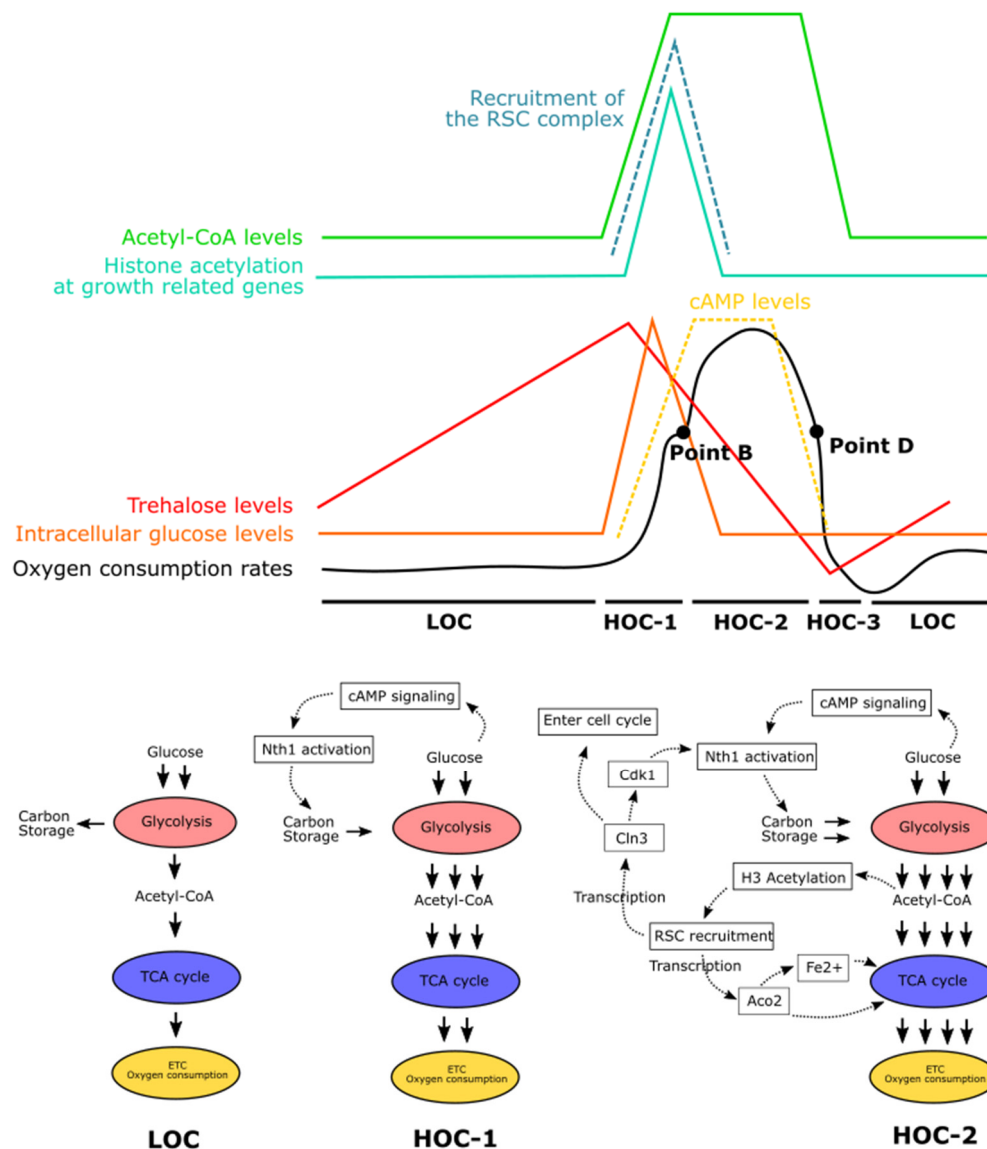


Figure 3.34 The 'Sozu' model for the molecular mechanism behind the yeast metabolic cycle. In the top and middle panels, the known or measured changes in chromatin and cellular metabolites are plotted with the solid lines in different colors, while the predicted or hypothesized changes are plotted in dashed lines. In the bottom panel, the three diagrams illustrate the key interplays between cellular metabolism, epigenetics and gene expression during the YMC

Then the cells encounter point B, where the increases in oxygen consumption rate experience a brief stagnation. In this model, it can be explained by the limited capacity of the TCA cycle. Although the flux of glucose through glycolysis is still increasing, the enzymatic activities of one or more TCA enzymes are saturated, which limits the further increase in the oxygen consumption rate. One of its direct outcomes is that certain upstream metabolites may accumulate before encountering point B. Actually, the level of ethanol was found to increase

sharply just before encountering the point B (Porro et al. 1988). Ethanol is converted from pyruvate via acetaldehyde. The production of ethanol in the aerobic conditions is referred to as the Crabtree effect (De Deken 1966). Both the overload of glucose and the limited capacity of the TCA cycle have been shown to be able to cause the Crabtree effect (Boulton & Quain 2001). Although the Crabtree phenotype normally happens at a relatively high glucose concentration, a pulse of glucose in the glucose limited environment has also been shown to induce the short term Crabtree effect (Hagman et al. 2014). Once the time delay between the transcription and transcripts of *ACO2* and *CLN3* has passed, the cells will leave this stagnation point and enter the HOC-2 phase.

The expression of *ACO2* can explain the further increase in the oxygen consumption rate after the point B, which is evidenced by the fact that the deletion of *ACO2* can prevent the further increase in the oxygen consumption rate after the point B. Aco2 is one of the two aconitases in budding yeast that can catalyse the conversion from citrate to isocitrate at the beginning of the TCA cycle (Bradley et al. 2019). In mammals, the cytosolic aconitase can also regulate the levels of ferrous ions in the cells (Narahari et al. 2000; Kennedy et al. 1992). During the HOC-2 phase of the YMC, the exact mechanism of Aco2 upregulating the oxygen consumption rate is not known. However, experiments in this study suggest that Aco2 may function both as the iron regulator and the aconitase, because excess ferrous ions in the medium can but only partially rescue the phenotype of *aco2Δ*. More experiments are needed to uncover the exact mechanisms.

Cln3 is one of the earliest G1/M cyclins. It can bind to and activate Cdk1, which further activates the expression of CLN1 and CLN2. Here, the activated Cdk1 has two main functions. First, based on the findings in the literature, it can further phosphorylate and activate Nth1, which further promotes the consumption of the stored trehalose (Ewald et al. 2016; Cerevisiae Graphical et al. 2016). Second, it allows the cells to pass the START point and enter

the cell cycle (Wijnen et al. 2002). Actually, its timing is consistent with the start point of DNA replication revealed by the flow cytometer results of propidium iodide staining (Burnetti et al. 2016). However, one weakness here is that it cannot explain why only less than 30% of the cells enter the cell cycle in each cycle of the YMC.

Currently, it is unknown how the cells exit the HOC phase and re-enter the LOC phase. One possibility is that the exit of the HOC phase is triggered by the low cAMP level. The low cAMP signalling might limit the degradation of trehalose by reducing the activity of Tpk1 which can phosphorylate and activate Nth1. All the explanations above need to be tested by more experiments.

3.3.5 Other experiments that may help understand the mechanism of the YMC

3.3.5.1 Checking other bromodomain containing chromatin remodelling complexes

So far, the importance of the RSC complex in budding yeast has been tested during the YMC. However, the RSC complex is not the only chromatin remodelling complex in yeast that contains bromodomains. The SWR1 complex also contains two bromodomains in each of its Bdf1 and Bdf2 subunit. The SWR1 complex is able to perform nucleosome editing by replacing the canonical H2A-H2B dimers with the H2A.Z-H2B dimers, which makes the nucleosomes become less stable (Kobor et al. 2004; Krogan et al. 2003; Zhang et al. 2005b; Mizuguchi et al. 2004b). Some evidence suggests that Bdf1 mainly binds to acetylated histone H4 (Matangkasombut & Buratowski 2003). The batch culture experiments have shown that the yeast strain having *BDF1* deletion is unable to grow on nonfermentable carbon sources (Fu et al. 2013). In other words, the deletion of *BDF1* may cause defects in aerobic respiration. In fact, the deletion of *RSC2* can also reduce the ability of the cells to undergo aerobic respiration in batch cultures (Tigano et al. 2015). Due to this similarity, it is still worth testing the phenotype of *BDF1* deletion under the YMC conditions. The SWI/SNF complex belongs to the same family of remodeling complexes as the RSC complex. It has a bromodomain in its

catalytic subunit Snf2. In this study, the Rsc2 isoform of the RSC complex has been shown to be required for the YMC, however it is not known whether the RSC complex is enough for the YMC. Although the function of the SWI/SNF complex seems to be less important for budding yeast than the RSC complex (The deletion of *STH1* is lethal, however the strain carrying *SNF2* deletion is viable (Cairns 1998)), an early work indicates that its bromodomain can anchor it to the acetylated promoter nucleosome (Awad & Hassan 2008a). Hence, it is still worth examining whether the SWI/SNF complex is also required for the YMC by deleting *SNF2* directly. If the full-length deletion can cause YMC defects, it may be even necessary to check the phenotype of the Snf2 bromodomain deletion mutant.

It may also be worth investigating other bromodomain containing proteins as well, particularly Gcn5, which is the lysine acetyltransferase and is also a subunit of the SAGA complex. Although the deletion of *GCN5* can abolish the YMC (Cai et al. 2011), it is not known whether deleting its bromodomain is enough to cause defects for the YMC. There is some evidence suggesting the bromodomain of Gcn5 can target the SAGA complex to the acetylated promoter nucleosome (Hassan et al. 2002b; Li & Shogren-Knaak 2009). Hence, it is worth deleting the bromodomain of Gcn5 and checking the phenotype in the fermenter.

3.3.5.2 High throughput screening of mutants that show defects in the YMC

In the past 15 years, a few YMC mutants have been identified by individual experiments, which are summarized in a recent review (Mellor 2016). They have pointed out genes that are important for the YMC and illuminated some aspects of the mechanism of the YMC. However, due to the limited number of YMC mutants, it is hard to draft an overall picture of the YMC in a bias-free manner. The ‘Sozu’ model proposed in the last section is still largely based on indirect evidence. One way is to develop a high throughput assay to screen for mutants showing defects in the YMC.

However, unlike other phenotypes, the YMC is hard to test in a high throughput manner. Currently, in order to check the YMC phenotype, the strain needs to be grown in a 1L culture in the fermenter with the growth conditions carefully controlled. Only one strain can be tested at a time. The whole process, including the preparation of the fermenter, takes at least one week. Although some fermenter models have multiple (often 4) growth chambers, the increase in the throughput is still limited and the cost of such models are significantly higher. Hence, in the current setting, it is not practical to test a large number of strains within a reasonable amount of time, energy and spending.

In order to develop a high throughput assay, it is necessary to find a simpler method to observe the YMC. It has been showed that a YMC-like periodic oscillation of cellular oxygen consumption can be observed in a batch culture mode (Slavov & Botstein 2011). Instead of growing cells in the normal YMC medium (containing 1% of glucose), the cells were inoculated into the ethanol containing phosphate-limited media. After a period of exponential growth, the cells started to exhibit oscillations of oxygen consumption. Importantly, during the whole process, no fresh medium was added. Although is not known whether such periodic oscillations are the same as the normal YMC observed in this study, it provides a solution to observe the phenotype of the YMC with a relatively simple setting and by a potentially fermenter-independent manner. However, in the paper, the YMC was still reported as a trace of readings from the DO sensor. A better reporter method is needed in order to fit the high throughput settings.

A possible solution is to utilize fluorescence to monitor and report the YMC. Here, I propose three different methods to achieve this. The first one is the direct monitoring of oxygen levels in the medium. It could be achieved by special fluorescent dyes which can respond to the levels of molecular oxygen by changing their fluorescence spectra. One example is [Ru-(dpp)₃]²⁺. It has been applied to monitor the intracellular levels of molecular oxygen in rat C6

glioma cells (Miller et al. 2002). The second method is to monitor the fluorescence of certain metabolites. The levels of NADH have been shown to oscillate during the YMC and can reflect the cellular redox states. It can absorb light of 340nm and emit fluorescence at 460nm. The oscillatory fluorescence levels of NADH can indicate its intracellular levels and further reflect the phases of the YMC (Papagiannakis et al. 2017). Finally, the phases of the YMC can also be indicated by short half-life fluorescent proteins. Two fluorescent protein genes (such as GFP and RFP) can be placed under the regulation of a HOC phase gene promoter and a LOC phase gene promoter respectively. The relative intensities of fluorescence from two channels can be utilized to indicate the phase of the YMC. Actually, as long as there are enough channels to record the fluorescence, the three methods are not mutually exclusive. They can potentially be combined to provide a more reliable indication of the YMC phases and even reflect more cell physiological states.

With the batch mode YMC and the fluorescent reporter, the experimental design for the high throughput screening of YMC related genes is proposed below. The prototrophic yeast deletion mutant collection is commercially available from Euroscarf. Individual deletion mutants can be inoculated into each well of the 96-well plate containing the ethanol YMC media. The cells on the 96-well plate can be grown at 30°C in the PHERAstar microplate reader (a device to measure the fluorescence levels in a high throughput manner) with constant shaking. Accordingly, this condition should be enough for the cells to exhibit the YMC (Slavov & Botstein 2011). The oxygen consumption rates or the phases of the YMC in each well of the 96-well plate can be monitored in real time by recording the fluorescence in various channels. Eventually, multiple traces of signals can be obtained from each well of the plate. They are further compared to the normal YMC signals recorded from the wild-type. Abnormal signals imply that the deleted gene in the corresponding well on the plate could be YMC relevant. Further validation experiments can then be performed on that mutant under the normal YMC conditions. With this approach, it may be possible to screen all viable deletions within a few

months. Once a large number of YMC relevant genes are identified, it may be even possible to perform data mining in an unbiased way. For example, certain signalling pathways could be important for the YMC, if they are significantly enriched for the YMC relevant genes.

3.3.5.3 C¹³ metabolic flux analysis

The 'Sozu' model proposed here describes an oscillation between the anabolic and catabolic metabolism. Specifically, there are two carbon sources, the constantly-added fresh glucose-containing medium and the cells own carbon source storage (trehalose and glycogen). During the LOC phase, the cells limit their consumption of glucose from the fresh medium and store it in the form of trehalose and glycogen. Once the carbon storage reaches a certain threshold, the cells enter the HOC phase, where the cells switch from anabolism to catabolism and consume the carbon source storage in addition to the glucose in the fresh media, in order to fuel the cell cycle and other energy demanding cellular processes. Currently, it is still a hypothesis built on indirect evidence. To test this hypothesis, it is critical to know how the cells utilize the two carbon sources in different phases of the YMC. The C¹³ metabolic flux analysis might be a method to answer this question. The general experimental design is described below.

In the middle of the LOC phase, the supply of the fresh medium is transiently switched from the normal glucose containing medium to the C¹³ labelled glucose containing medium. After that, a series of samples are taken every two minutes in the initial 20 minutes. Then samples are taken every 15 mins before entering the HOC phase. Once the oxygen consumption rate starts to increase, the samples are taken every 2 mins during the HOC phase. The abundances of C¹³ in the major metabolites in each sample are quantified by mass spectrometry. This experiment can potentially illuminate two things. First, it can tell how the glucose is fluxed into various metabolic pathways during the LOC phase by measuring the changes in the abundances of C¹³ for each metabolite in the initial 20 mins. Second, it can tell how carbon

storage is utilized during the HOC phase. After the addition of C^{13} labelled glucose in the LOC phase, some trehalose and glycogen may be labelled with C^{13} as well. Once they are utilized in the HOC phase, the abundances of C^{13} in the central metabolic pathway may increase, even if C^{13} labelled glucose is no longer added. The above is one C^{13} metabolic flux analysis on the LOC phase. The same analysis can then be performed on the HOC-1, HOC-2 and HOC-3 phases by adding the C^{13} labelled glucose containing medium to the corresponding phases.

3.3.5.4 Phosphoproteome and Acetylome

The YMC mechanism described by the 'Sozu' model involves a lot of post-translational modifications, such as the activation of the cAMP signaling pathway, the phosphorylation of Nth1 and the upregulation of lysine acetylation. To test this hypothesis in an unbiased manner, it may be important to obtain the phosphoproteome and acetylome and study their temporal changes during the YMC. In this study, the phosphoproteome of 9 samples during the YMC was measured once in collaboration with Kessler lab. However, limited by the depth of the phosphoproteome, the phosphorylation levels of only few proteins could be quantified during the YMC, so it was difficult to get meaningful insights from that dataset. To improve the depth, there are two methods. First, the number of samples needs to be reduced. Because the samples need to be pooled together before quantification by mass spectrometry, the complexity of the final sample is greatly increased, if the more samples are pooled together. The high complexity can reduce the depth of the phosphoproteome. In this study, the 9 samples covered the whole YMC. It may be better to use fewer samples to study one phase of the YMC at a time. Second, during the fractionation of the samples by the HILIC column, the eluate needs to be split into more fractions. Mass spectrometry needs to be performed on each of the fractions.

Chapter 4.

Technical development of yeast NET-seq

4 Technical development of yeast NET-seq

4.1 Ultra-low background NET-seq, a new protocol for yeast NET-seq

4.1.1 Introduction

4.1.1.1 The high background is a major technical limitation for yeast NET-seq

In this chapter, NET-seq specifically refers to the yeast NET-seq developed in Churchman lab (Churchman & Weissman 2011). Mammalian NET-seq developed in Proudfoot lab does not have the problem of the high background (Nojima et al. 2015). In NET-seq, the nascent RNA is extracted from the whole cell lysate by immunoprecipitating RNAPII. Then the extracted nascent RNA proceeds to library construction, sequencing and downstream bioinformatic analysis. As nascent RNA only constitutes a small proportion of the total RNA in the cell, efficient removal of the non-nascent background is critical for using NET-seq to study transcription.

The major problem of the current yeast NET-seq protocol is that the final library still contains a significant amount of non-nascent background. It is evidenced by both the results of the bioanalyzer and the analysis of sequencing reads (Figure 4.1a,b). When the RNA extracted by RNAPII immunoprecipitation is examined by a bioanalyzer, the strong peaks of 5.8s, 18s and 25s rRNA can be clearly seen. It implies that mature RNA still constitutes a vast majority of the extracted RNA. On the contrary, the peak of U1 snRNA is almost negligible. U1 snRNA is one of the spliceosomal RNA species. Spliceosomes can bind to the CTD of RNAPII, so U1 snRNA normally co-precipitates with RNAPII and nascent RNA (Nojima et al. 2018). The small peak of U1 snRNA indicates that the proportion of nascent RNA in the extracted RNA is relatively low in the library, at least by comparing their fluorescence signals. Furthermore, when the sequencing reads of NET-seq is examined, 73% of all the reads actually map to rRNA and tRNA. These reads normally cannot be uniquely aligned to the genome, so they are filtered at the alignment stage. Among the uniquely aligned reads, around 33% corresponds to mature RNA, which is estimated by checking the number of *S.cerevisiae* reads in a spike-in normalized mock

IP library. (The mock IP library is produced by performing immunoprecipitation on the same strain except having untagged RNAPII, so it is purely non-nascent background. The mock IP library can be normalized to the real IP library by read counts of either the *SCR1* gene (transcribed by RNAPIII rather than RNAPII, so the reads mapping to it are purely background) or the total reads of *S.pombe*.) In other words, only 18% of all the sequenced reads of NET-seq are actually nascent. Even if the rRNA and tRNA reads can be filtered, 33% of the final uniquely aligned library is still the background of mature RNA. The amount of background can be seen in Figure 4.1c.

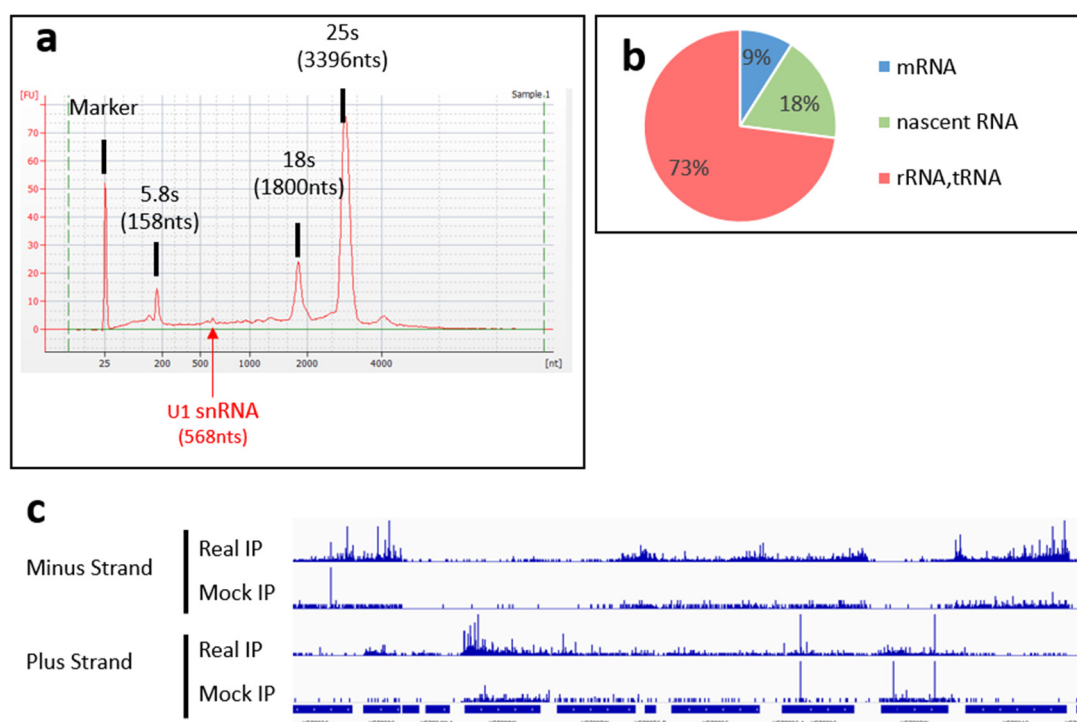


Figure 4.1 The original protocol of yeast NET-seq has a high background. a. The Bioanalyzer result of the eluted RNA after immunoprecipitating RNAPII with the original protocol. **b.** The composition of the reads in a NET-seq library prepared by the original protocol. **c.** The sequenced libraries prepared by the real IP and the mock IP are compared in IGV.

Currently, to remove the 33% of the background in the final sequencing data, the signals in each of the real IP library needs to be subtracted by the signals of its corresponding mock IP library. Although this method can remove the background of mature RNA, there are still several caveats.

First, it greatly increases the complexity, labour and expense of NET-seq. In order to check a mutant with NET-seq, its corresponding untagged strain has to be constructed. Then the same NET-seq library construction and sequencing procedures need to be repeated on that strain as well, which directly doubles the time and expense of NET-seq. After the subtraction between the two libraries, there is only one library with 18% of reads left. Because of the low usage rate of reads, the NET-seq libraries have to be sequenced more deeply, which further increases the cost of NET-seq.

Second, it is critical but also difficult to perform a proper normalization between the real IP and mock IP libraries for the same strain as well as between libraries of different strains. As the background subtraction procedure removes a large proportion of reads from the library, the results after background subtraction can be strongly dependent on the normalization, particularly when the actual signals of nascent RNA are weak and the signal to noise ratio is low. Also, after subtraction, a large number of nucleotide positions on the genome will have negative read counts, which can cause problems in the downstream bioinformatic analysis. On the other hand, a perfect normalization can be difficult to achieve, because it is subject to a lot of variation. For example, the batch differences in the efficiency of the washing step of NET-seq immunoprecipitation can directly lead to the variations of background in different samples. As a result, the sensitivity and difficulty of normalization can be a limiting factor for NET-seq to be used in a more quantitative scenario.

Third, the high background limits the development of other NET-seq variants. TEF-seq (Transcription elongation factor sequencing) is a variant of NET-seq developed in Mellor lab (Fischl et al. 2017). It pulls down various transcription elongation factors which can co-precipitate with RNAPII and nascent RNA, so it can be a powerful tool to study the interaction between RNAPII and transcription elongation factors with single nucleotide resolution during the process of transcription elongation. However, because of the indirect

immunoprecipitation of RNAPII, the signals in TEF-seq are normally lower than those in NET-seq, while its background level is still similar to NET-seq. As a result, for certain transcription elongation factors which only weakly interact with RNAPII, the signal to noise ratio will be too low to produce a usable TEF-seq library. Hence, the high background levels restrict the application of TEF-seq.

4.1.1.2 The possible causes of the high background and the potential solutions

After cell lysis, the original cellular structures are disturbed. RNAPII can be exposed to many other proteins and RNA molecules which RNAPII normally does not contact. At the same time, RNA molecules have the ability to bind to various surfaces non-specifically. The non-specific RNA eventually becomes the background in RNA immunoprecipitation. Hence, to minimize the background, there are two general strategies. First, the background can be reduced by retaining the original cellular and molecular structures, which reduces the exposure to non-specific RNA. Second, the background can be reduced by lowering the capability of non-specific RNA binding. The two strategies will be discussed separately.

Nuclear or chromatin extraction is a common method of the first strategy. The majority of non-specific RNA species in NET-seq experiments are tRNA, rRNA and mRNA. Most of these RNA molecules are present in the cytoplasm rather than the nucleus, while the nucleus is the major place of transcription. In theory, by extracting the nucleus rather than lysing and homogenizing the whole cell, the chance of RNAPII getting in contact with non-specific RNA from the cytoplasm is minimized. Then, the background can be reduced. For example, mammalian NET-seq developed in the Proudfoot lab has cell fractionation steps to extract the cell nuclei (Nojima et al. 2015). However, this method is not suitable for yeast NET-seq experiment, because yeast cells have cell walls, which makes nuclear extraction difficult.

UV crosslinking can also be considered as a method of the first strategy. Performing UV crosslinking while the cells are still intact can retain those in vivo physical interactions. The

non-specific RNA binding during cell lysis and immunoprecipitation can be removed by using a more stringent wash buffer. Furthermore, the crosslinking can produce CIMS (crosslinking induced mutation sites) on RNA, which can be easily identified by sequencing (Zhang & Darnell 2011). It is utilized in CLIP experiments to identify *in vivo* RNA binding sites. However, similar to nuclear extraction, transcription may be changed during the UV crosslinking step. Hence, this method may not be desirable for NET-seq.

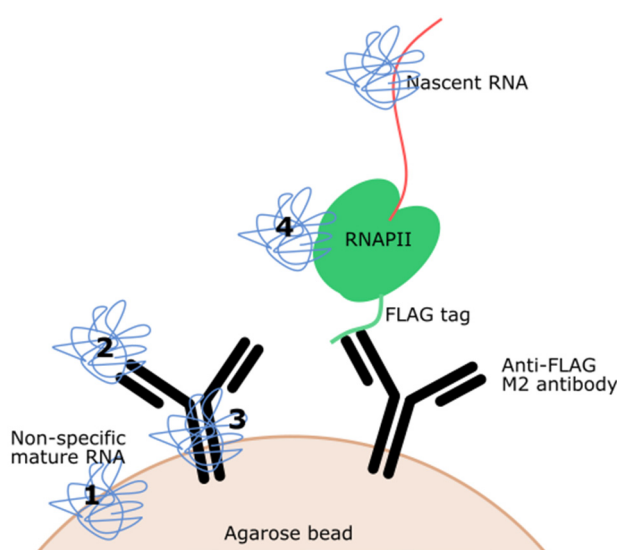


Figure 4.2 A schematic diagram showing the potential binding sites for the non-specific mature RNA. The non-specific mature RNA is represented by a ball of string. Its four potential binding sites are: 1. the surface of the agarose beads 2. The body of the attached antibody 3. The complementarity determining regions of the antibody 4. the surface of RNAPII and nascent RNA.

In terms of the second strategy, there are many different methods to reduce the binding capability of non-specific RNA. The choice of methods also depends on the major binding places of the non-specific RNA. In the NET-seq experiment, the binding places are demonstrated in Figure 4.2. They can be the surface of agarose which is the material of the immunoprecipitation beads, the complementarity determining regions (CDRs) of anti-FLAG M2 antibody, the remaining part of the antibody and the outer surface of RNAPII along with its associated factors. There have been many attempts to reduce this non-specific binding.

First, it was hypothesized that the bead material (agarose) was the major source of nonspecific RNA binding. The reason was that the agarose beads were porous, which could give rise to a high binding capacity, but might also increase non-specific binding (suggested by the manufacturer). If this hypothesis was correct, the background should be reduced by performing the immunoprecipitation with Dynabeads, which generally had less nonspecific binding than agarose beads. The Dynabeads were tested in the NET-seq experiments, however, no obvious improvement in terms of the background could be observed (unpublished data from Vasilieva lab in University of Oxford). It indicates that the surface of the bead material is unlikely to be the primary binding place for the non-specific RNA.

Second, the CDR of the anti-FLAG M2 antibody might have the potential to bind RNA non-specifically. As the FLAG epitope is heavily negatively charged, there is a possibility that the CDR of anti-FLAG M2 antibody might be positively charged, which may bind the negatively charged RNA non-specifically during immunoprecipitation. If this hypothesis was correct, changing the epitope might reduce the background. However, the anti-FLAG and anti-HA double pull down experiment does not support this hypothesis. The background level was even increased, compared to single pull down NET-seq (unpublished data from Mellor lab in University of Oxford). In this experiment, RNAPII was tagged by both FLAG and HA epitope tags. Unlike the FLAG peptide, the HA peptide is uncharged. The CDR of anti-HA antibody is unlikely to have the same spectrum of non-specific binding as the anti-FLAG M2 antibody. It suggests the CDR of the antibodies are not the major source of the background of NET-seq.

Third, the non-CDR region of the antibody could be a binding site for non-specific RNA as well. As the anti-HA and anti-FLAG M2 antibodies in the double pull-down experiment share a similar non-CDR region, it cannot exclude the possibility that the non-CDR region is the major source of the non-specific background. If this is true, affinity purification methods that are not based on antibodies need to be used. There are a range of non-antibody purification systems,

such as SpyTag-SpyCatcher system, StrepTag-StrepTactin system, PolyhistidineTag-IMAC system and CBP-calmodulin system. After balancing all the factors, twin-strep tag was chosen to purify RNAPII. It has a high affinity to StrepTactinXT and can also be eluted in a mild condition with neutral pH. The Twin-Strep tag is only 28 amino acids long, which reduces the chance of disrupting the functions of RNAPII. However, performing double-pull down with both anti-FLAG M2 antibody and the TwinStrepTag-StrepTactinXT system only produced a small reduction in the background (unpublished data from Mellor lab in University of Oxford). This indicates that the background may be hard to reduce by merely changing the protein purification system without increasing the stringency of the wash buffer.

The last potential place for non-specific RNA binding is the RNAPII itself. During the process of transcription, RNAPII can also be associated with a lot of factors, such as Dst1, Spt4/5, Paf1 complex, spliceosome and FACT complex (Mason & Struhl 2003; Jaehning 2010; Martinez-Rucobo et al. 2011; Kim et al. 2007; Nojima et al. 2018). Many of them have a large surface area and some even have RNA binding capabilities, so RNAPII itself can be a good candidate for the binding site of non-specific RNA. If this is true, there are two options. One is to use a certain nuclease to digest all RNA exposed and leave a footprint (16-20nts) of RNAPII at the 3' end of nascent RNA. The other option is to use more stringent wash buffers, such as buffers containing SDS, high concentrations of salt or chaotropic agents. This study tested the second option. Specifically, in order to reduce the binding of non-specific RNA, the stringency of the NET-seq wash buffer was increased by adding the chaotropic agent, urea.

4.1.1.3 Urea can potentially be used in NET-seq to increase the stringency of the wash step

After checking all the common reagents for increasing the stringency of the wash step, urea appeared to be a good choice, for three main reasons. First, wash buffers containing 1M of urea has been used in other RNA immunoprecipitation experiments such as iCLIP (Huppertz

et al. 2014). It has been demonstrated to reduce the background. Second, the FLAG peptide and anti-FLAG M2 antibody were also used in those CLIP experiments, so the interaction between FLAG peptide and the antibody was not disrupted in the presence of 1M urea (Huppertz et al. 2014). Third, RNAPII for both *S.cerevisiae* and *S.pombe* has been shown to be able to tolerate at least 2M of urea without detectable disintegration (Kimura et al. n.d.; Edwards et al. 1991). This indicates that nascent RNA could be immunoprecipitated after washing with buffers containing 2M of urea. To conclude, adding 1M of urea to the wash buffer of NET-seq can potentially be a solution to the binding of non-specific RNA in NET-seq.

4.1.1.4 Aims and plans

The aim of this chapter is to test if the background of NET-seq can be reduced by adding 1M of urea to its wash buffer (the new wash buffer). In the later discussion, this modification to the original NET-seq protocol is referred to as the 'new protocol'. First of all, western blot was performed to check if RNAPII and the interaction between the FLAG epitope and its antibody could withstand the increased stringency during the wash step. Then the background level was examined by comparing the *S.pombe* normalized sequencing data of the real NET-seq (RNAPII tagged by FLAG) and the mock NET-seq (RNAPII untagged). If the new protocol could reduce the background, the signals in the mock NET-seq library should be reduced. Finally, other than the background reduction, the new protocol should not introduce obvious unexpected changes to the data. To test this, the metagenes were compared between the NET-seq data prepared by the new protocol and the background subtracted data prepared by the original protocol.

4.1.2 Results

4.1.2.1 Both the FLAG-epitope-antibody interaction and RNAPII can tolerate the stringency of the new wash buffer

To check if RNAPII and the binding between the FLAG epitope and its antibody can tolerate the stringency of 1M urea during the wash step, untagged *S.cerevisiae* and Rpb3-FLAG tagged *S.cerevisiae* samples were processed with the new immunoprecipitation protocol. For details of the new protocol, please refer to section 2.3.2. A western blot was performed to check the presence of Rpb3-FLAG and Rpb1 in different fractions of the immunoprecipitation (Both Rpb1 and Rpb3 are subunits of RNAPII). The loading volume of each well was proportional to the total volume of its corresponding fraction so that the signals in the western blot should be proportional to the total amounts of their corresponding proteins in the fractions. Rpb1 and Rpb3-FLAG were probed with anti-Rpb1S5P (04-1672-I, Sigma) and anti-FLAG M2 (F3165, Sigma) antibody respectively.

In Figure 4.3, both Rpb1 and Rpb3 give clear bands in the input fraction, while the flow through does not show any signal. It indicates that RNAPII has been efficiently captured by the anti-FLAG M2 beads rather than being lost in the flow through. In the new wash buffer, none of the two proteins can be detected. However, they can be seen in the elution buffer. It means that the new wash buffer does not significantly disturb the interaction between the FLAG epitope and anti-FLAG M2 antibody. Furthermore, the co-precipitation of Rpb1 suggests the nascent RNA is likely to co-precipitate as well after the wash steps.

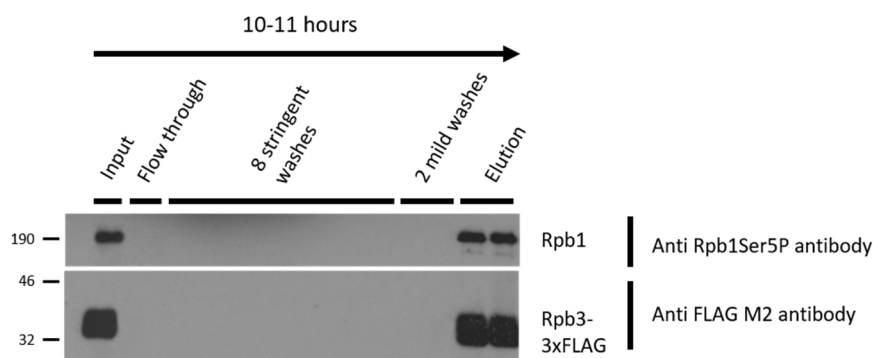


Figure 4.3 The presence of Rpb1 and Rpb3 in different fractions of new NET-seq immunoprecipitation procedure was checked by western blot. Rpb1 was probed by anti-Rpb1Ser5P antibody. Rpb3 was tagged at its c-terminus with the 3xFLAG epitope. It was probed by the anti-FLAG M2 antibody. The whole immunoprecipitation procedure took 10-11 hours. The wash buffer used in the 8 stringent wash steps contained 1M of urea. In the 2 mild wash steps, the original NET-seq wash buffer was used.

To sum up, the FLAG-epitope-antibody interaction can tolerate the stringency of the new wash buffer and the nascent RNA is likely to co-precipitate after the stringent washes. It is worth to proceed to the following tests.

4.1.2.2 The new protocol can greatly reduce the background of NET-seq

To check if the nascent RNA can be co-precipitated and the non-specific RNA background can be reduced by the new protocol, both the Rpb3-FLAG tagged *S.cerevisiae* (Sample A hereafter) and the untagged *S.cerevisiae* (Sample B hereafter) were spiked-in with the Rpb9-FLAG tagged *S.pombe* at the same ratio. Considering the signals from the untagged *S.cerevisiae* sample may be low if the background was efficiently removed, a very low spike-in ratio was used. That is, 0.05g of *S.pombe* grindate was added to every gram of *S.cerevisiae*. Both samples were processed with the new NET-seq procedures, including the immunoprecipitation, library preparation and sequencing.

Before library construction, the yields of RNA from the eluates after the immunoprecipitation step were first examined. It was found that if the new protocol was used, the total yields of RNA after immunoprecipitation were reduced and the yield for the tagged strain was higher

than that for the untagged one. Normally, with the original protocol, 7-9ug of RNA was expected from the same amount of cells for both Rpb3 tagged and untagged *S.cerevisiae* and the yields of RNA did not show a dramatic difference between the tagged and untagged strains. However, the new protocol produced 1.1ug of RNA for sample A (tagged) and 0.07ug of RNA for sample B (untagged). During the library construction, the RNA from sample A and B was alkaline fragmented into small pieces respectively, which gave the smear shown in Figure 4.4a. The fluorescence intensity of the smear for sample A was much stronger than that for sample B, which was consistent with the results of the Nanodrop. Although *S.pombe* in sample B was tagged at Rpb9, due to the low spike-in ratio, it could not contribute a significant amount of RNA. To sum up, as the background is known to constitute the majority of the RNA prepared by the original protocol, the lower yield for the new protocol and the differential yields between sample A and B suggest that the new protocol can remove the background more efficiently.

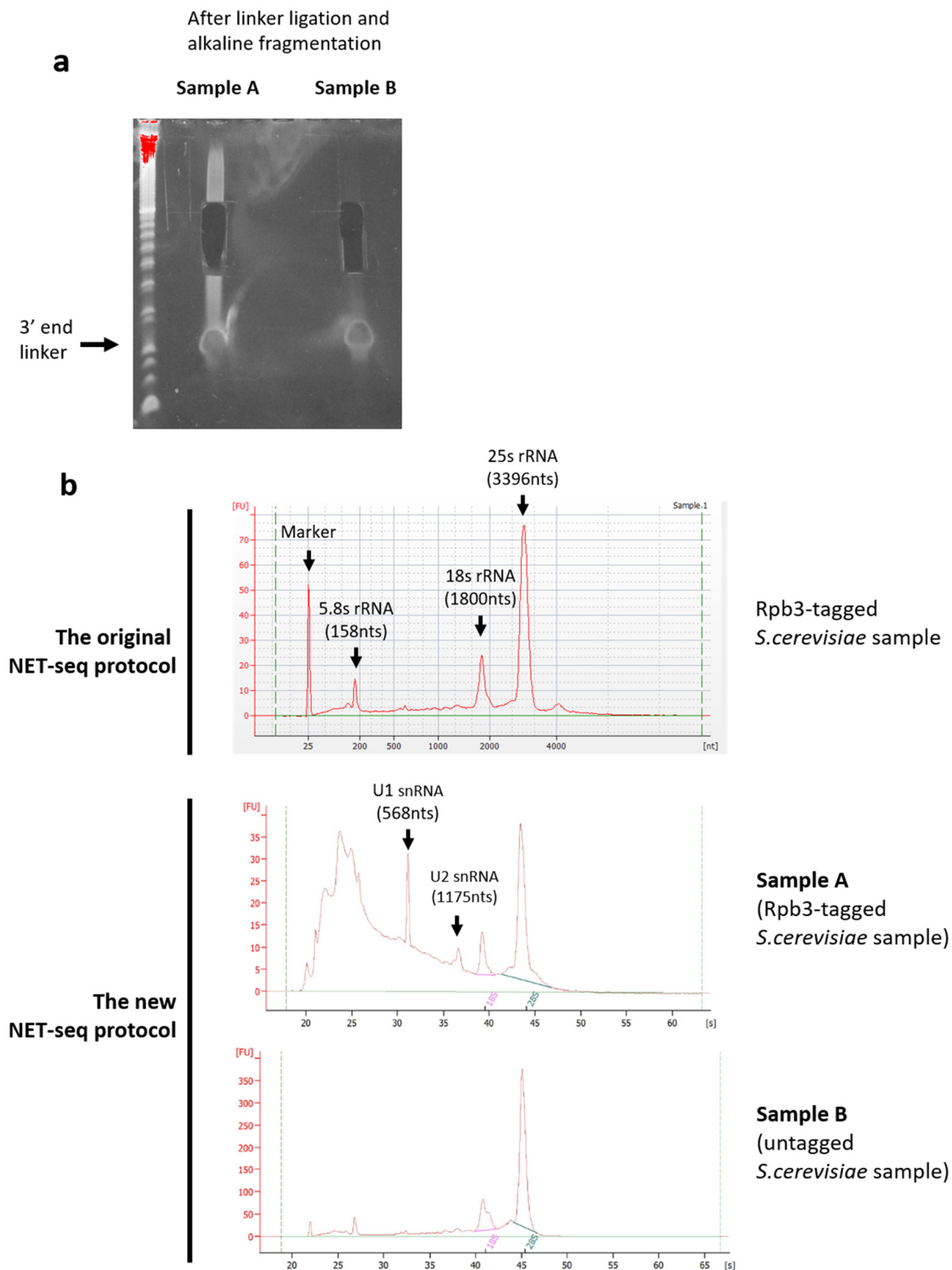


Figure 4.4 a. The yields of RNA after immunoprecipitation were revealed at the alkaline fragmentation steps. Sample A contained tagged *S.cerevisiae* RNAPII, while sample B had untagged *S.cerevisiae* RNAPII. **b.** The Bioanalyzer results of the immunoprecipitated RNA samples prepared by the original protocol and the new protocol were compared. The sizes and identities of some peaks have been marked.

As most of the nascent RNA is being transcribed, its sizes are generally smaller than those for mature RNA (the non-specific background). If the background was reduced by the new

protocol, a larger proportion of small size RNA should be seen. To check this, the RNA was further examined by the Bioanalyzer (Figure 4.4b, the middle and bottom panels). For comparison, the Bioanalyzer result for the tagged sample prepared by the original protocol is also displayed (Figure 4.4b, the top panel). Generally, sample B is similar to the tagged sample prepared by the original protocol. High peaks of rRNA can be seen, while the peak that has the same size as U1 snRNA is low. This supports the view that the majority of RNA from the original protocol is non-specific background. Sample A contains a much larger proportion of RNA with smaller size (below 500bps) than the other two samples. This is unlikely to be caused by RNA degradation, because sample B does not have such enrichment in small size RNA. Moreover, in sample A, the proportion of rRNA is also much lower and the peaks that have the same sizes as U1 and U2 snRNA can be clearly seen. These results indicate that the background can be reduced by the new protocol.

To validate the fore findings, library A and B were constructed from sample A and B respectively. After sequencing, reads were aligned to both *S.cerevisiae* and *S.pombe* genomes. Only uniquely aligned reads were retained. As *S.cerevisiae* cells in sample B were untagged, library B was predicted to mainly consist of reads *S.pombe*, which was tagged at Rpb9 (a subunit of RNAPII). Actually, the results are consistent with this prediction. Although *S.pombe* cells are added to sample A and B at the same ratio, the percentage of *S.pombe* reads in library A is 4.8%, which is much smaller than that in sample B (87.6%). The 12.4% *S.cerevisiae* reads in library B should be completely non-specific background. In order to obtain the proportion of background in *S.cerevisiae* reads of library A, the *S.pombe* read counts of library A and B were utilized to normalize their *S.cerevisiae* read counts. After normalization, sample A, prepared by the new NET-seq protocol, is estimated to contain 0.7% of non-specific background (Figure 4.5a). For comparison, the same analysis was performed on two libraries (tagged and untagged *S.cerevisiae*, both spiked-in with tagged *S.pombe*) prepare by the original protocol in the past. The overall background level for the original NET-seq protocol is

32% (Figure 4.6a). Hence, the new protocol can reduce the background of NET-seq from 32% to 0.7%.

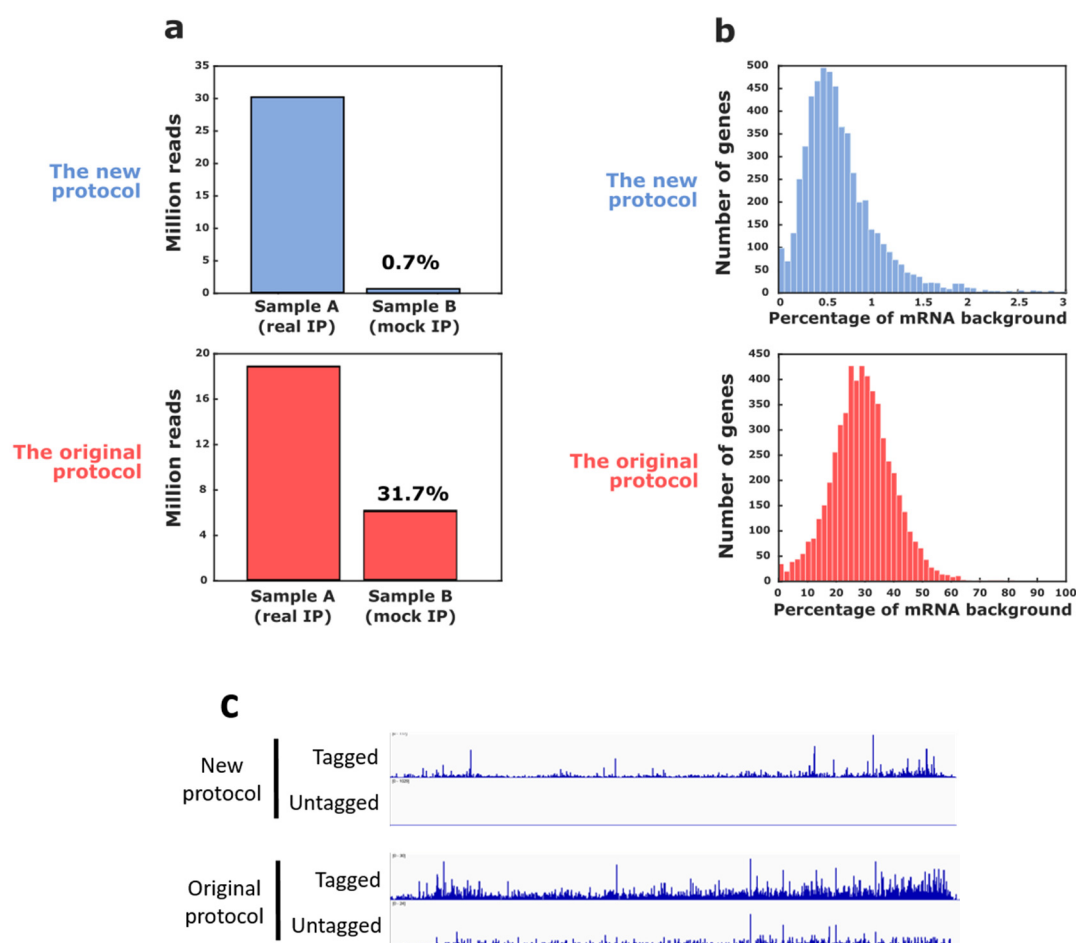


Figure 4.5 Comparing the background levels of the new and original protocols. The upper and lower panels are the results for the new and original protocols respectively. **a.** The total amount of background reads in the NET-seq library is revealed by the mock IP. **b.** The histogram for the percentage of background reads on individual genes. **c.** The NET-seq data is visualized by IGV. The untagged library shows the amount of background. Both the tagged and untagged libraries have been normalized.

To investigate the background levels in more detail, the background levels of individual genes were calculated and displayed by histograms (Figure 4.5b). The distribution profiles of the background levels of both protocols show a single and nearly symmetrical peak. For the new protocol, the median background level of all protein coding genes is 0.5%. Also, 99% of the genes have background levels lower than 2.6%. The highest background level, 13%, is found on the gene 'SNF6'. Except for this gene, all the background levels are lower than 10%. For the

original protocol, the background levels are much higher. The median background level for all protein coding genes is 29%. Furthermore, 95% of genes have their background levels exceeding 10%. There even exists 182 genes whose background levels are higher than 50%.

In order to get a more direct understanding of the background levels, the libraries are visualized in IGV (Integrative Genome Viewer). It is clearly seen that the background of the new protocol is extremely low, which is indicated by an almost flat line on its background track. On the contrary, a significant amount of reads can still be seen in the background track for the original protocol (Figure 4.5c).

To conclude, the results from Nanodrop, Bioanalyzer and sequencing show that the new NET-seq protocol has a much lower background than the original protocol.

4.1.2.3 The metagene profile of the library prepared by the new protocol is similar to that of the background subtracted library prepared by the conventional protocol

Without the new protocol, the conventional method for removing the background was to perform a background subtraction with the mock-IP library, as mentioned in section 4.1.1.1. To investigate whether the library prepared by the new protocol was similar to the background subtracted library prepared by the original protocol, their metagene profiles were plotted and compared. A metagene is one of the most basic ways of analysing NET-seq data. However, it can be heavily influenced by the background. This is because the distribution of RNAPII is biased towards the proximal regions of genes, while the background is evenly distributed on genes. If the library contains a higher level of background, the final metagene profile can become more flattened.

As is shown in Figure 4.6, if there is no background subtraction, the library prepared by the original protocol produces a more flattened metagene profile than the one prepared by the

new protocol. Interestingly, after background subtraction, the two metagene profiles almost completely overlap. It suggests that the new protocol does not introduce obvious systematic errors or at least the error is not seen on the metagene.

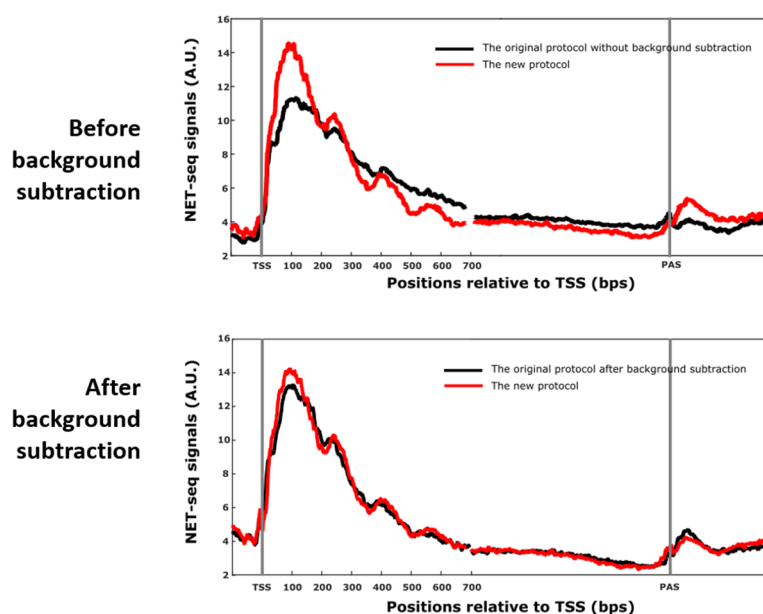


Figure 4.6 Comparing the metagene profiles of libraries prepared by the new (red) and the original (black) protocols. The upper and low panels show the metagene profiles before and after background subtraction. TSS (Transcription start sites); PAS (Polyadenylation sites)

4.1.3 Discussion

4.1.3.1 The advantages and possible applications of the new protocol

First of all, the current test of the new protocol suggests that it can eliminate the hassle of making the mock IP libraries and performing background subtractions at least for log-phase batch culture experiments. Obviously, the new protocol has an advantage on saving a great amount of time and energy spent on constructing the corresponding untagged strains, growing the untagged cells, performing the mock immunoprecipitation, constructing and sequencing the mock IP library. More than that, the mock IP NET-seq libraries can sometimes be very difficult to obtain in some time-resolved experiments. The YMC is a typical example. In the YMC, the levels of transcripts can change by several folds within few mins particularly in the HOC-2 phase. It means that the background levels of NET-seq may also vary in a highly

temporally sensitive way. To study the transcription during the YMC with NET-seq, the mock IP libraries can be critical, if the original protocol of NET-seq is used. To get the mock IP NET-seq library for background subtraction, the corresponding untagged strain needs to be grown in the YMC conditions as well and all the samplings have to be performed again on exactly the same time points. In practice, this is hard to achieve. The same problem might also exist for performing NET-seq in other time-resolved experimental systems, such as cell cycles, Glu-Gal shift and temperature shift experiments. Although the new protocol is only tested for log-phase batch culture cells, it has the potential to be applied to other systems and save the trouble of making the mock IP libraries.

Second, the new protocol allows NET-seq to achieve a higher signal-to-noise ratio by lowering the background. It can potentially be utilized to improve other NET-seq based techniques, such as TEF-seq (Fischl et al. 2017) and size-fractionated NET-seq (discussed in Chapter 4.2). As mentioned before, the high background of TEF-seq prevents weakly associated transcription factors to be studied by this technique. For factors that have a relatively strong association with RNAPII, the metagene analysis, which combines the features from a large number of genes, can still be performed. However, it may still be challenging to carry out investigations on individual genes or particular positions on the genome. To improve TEF-seq with this new protocol, considering the binding between the transcription elongation factor and RNAPII may not be strong enough to tolerate the new urea containing wash buffer, the HA-tagged transcription elongation factor together with its associated RNAPII should be enriched first from the whole cell lysate by the original NET-seq protocol. Then in the second immunoprecipitation against the FLAG tag on Rpb3, RNAPII is further purified from the previous eluate with the new protocol. The other technique that can be improved is size-fractionated NET-seq. The mature RNA background often exists as RNA fragments of variable sizes. They can be captured by each size fraction and eventually mapped to random places of a gene, which complicates the latter bioinformatic analysis. As the size selection step is

performed at the downstream of the new protocol, the new protocol can be directly applied to size-fraction NET-seq.

Third, the new protocol can potentially be applied to other species. Although the background reduction efficiency is not assessed for *S.pombe*, the enrichment of *S.pombe* reads in library B indicates that the new protocol may also work for *S.pombe*.

4.1.3.2 The limitations and possible improvements of the new protocol

In terms of the protocol itself, there are two main weakness. First, because of the eight extra urea wash steps, the new protocol is more time consuming and labour intensive than the original one. Reducing the number of urea washes and increasing the volume of urea wash buffer used may solve this problem without compromising the quality of the library. Second, as the new protocol removes the majority of the background, the final yield of RNA after immunoprecipitation is much lower than the yield of the original protocol, so all the downstream library construction steps have to be performed more carefully. Particularly, after reverse transcription, cDNA need to be size selected on a polyacrylamide gel. Due to the low RNA input at the beginning of the library construction, the fluorescence of cDNA may not be visible on the gel. Hence, the region of the gel to be excised can only be estimated. It may be challenging for people with little experience in NET-seq. One way to circumvent it is to increase the initial amount of cells. In this case, all the reagents in the immunoprecipitation may have to be scaled up as well.

In terms of the data, one concern is whether RNAPII of all the states (including pausing, elongating and backtracking states) are equally represented after the stringent washes. If RNAPII in a certain state is less stable during the urea wash steps, it may get lost. The NET-seq signals originated from RNAPII in that state may be underrepresented in the final sequencing data. Currently, based on the western blot and the comparison between metagene profiles, it does not seem to be the case.

4.2 Size-fractionated NET-seq, a new method for mapping transcription start sites

4.2.1 Introduction

There are still a lot of unannotated transcription units on the genome of *S.cerevisiae*. To study their expression patterns and regulation, it is important to obtain their accurate transcription start sites (TSS). In this section, I used a new NET-seq based sequencing method to map the TSS for those unannotated transcription units in a relatively unbiased way.

4.2.1.1 The current methods for mapping TSS

Currently, there are a variety of sequencing methods for genome-wide TSS mapping. They normally involve three relatively independent steps. First, RNA is isolated from the relevant biological material. Second, the 5' ends of RNA is captured for library construction and sequencing. Third, since the 5' ends of RNA are sequenced, a peak of sequencing reads is produced on the TSS of genes. The TSS can then be mapped by looking for the positions of the peaks of sequencing reads. The first step defines the type of the input RNA material for library construction. It determines the ranges of TSS that can be mapped. In the second step, the method for capturing RNA 5' ends can largely influence the sensitivity and precision of TSS mapping. The two steps will be discussed separately in this section.

The input RNA material for library construction can be either nascent RNA or mature RNA. Mature RNA is easy to obtain, but the drawback is that the TSS of many unstable transcripts may be underrepresented, if mature RNA is chosen to be the input material. It is because those unstable transcripts, such as CUTs, XUTs and NUTs, are quickly degraded and are only present in small quantities in the mature RNA fraction (Schulz et al. 2013a; van Dijk et al. 2011; Wyers et al. 2005). Although disturbing the RNA nuclear surveillance pathway or cytosolic degradation system by certain deletions or mutations can allow a subset of unstable transcripts to increase their cellular concentrations, it may not be applicable in some experiments. Nascent RNA can be isolated from total RNA by three main methods, including

nucleotide labelling (4SU-seq, TT-seq, GRO-seq and PRO-seq), immunoprecipitation (NET-seq, mNET-seq) and direct chromatin isolation (nascent-seq) (Core et al. 2008; H. Kwak et al. 2013; Rabani et al. 2011; Nojima et al. 2015; Churchman & Weissman 2011; Khodor et al. 2011; Schwalb et al. 2016). Although nascent RNA requires many more steps for isolation, it allows the TSS for both stable and unstable transcripts to be mapped in a relatively unbiased way. Here, the word ‘relatively’ is used, because even nascent RNA is limited by the growth conditions of the cell. Different growth conditions allow different subsets of genes to be expressed. For example, log phase fast growing cells have different transcription profiles from stationary phase cells or cells in stress conditions. *HXT6*, which encodes the high-affinity glucose transporter, is not expressed in log phase cells in the YPD medium (Liang & Gaber 1996). *GAL* genes are expressed in the galactose medium (Hovland et al. 1989). Obviously, if a gene is not expressed in a certain condition, it is challenging to map its TSS in that condition. To obtain a comprehensive mapping of transcription start sites, sequencing might need to be performed in several different growth conditions. The YMC may be an ideal system to study gene expression. During a typical cycle of the YMC, the cells can experience various metabolic states. As a result, more genes have the chance to be expressed at the different time points during the YMC.

In terms of the second step, which is the library construction, there are three major strategies. They are oligo-capping, cap-trapping and template switching (Harbers & Kahl 2011). All the strategies aim to capture and sequence the 5' ends of RNA. Normally, the isolated RNA is comprised of two species, the 5' capped and the 5' uncapped RNA. The former reflects the positions of TSS and needs to be retained. The latter needs to be removed, because it is mostly RNA degradation products and does not reflect the real TSS positions. The oligo-capping and cap-trapping strategies can distinguish the two species by enzymatic reactions and direct chemical modifications respectively. The template switching strategy does not distinguish

them, so it normally has more artefacts and is less used. The three strategies are briefly introduced below.

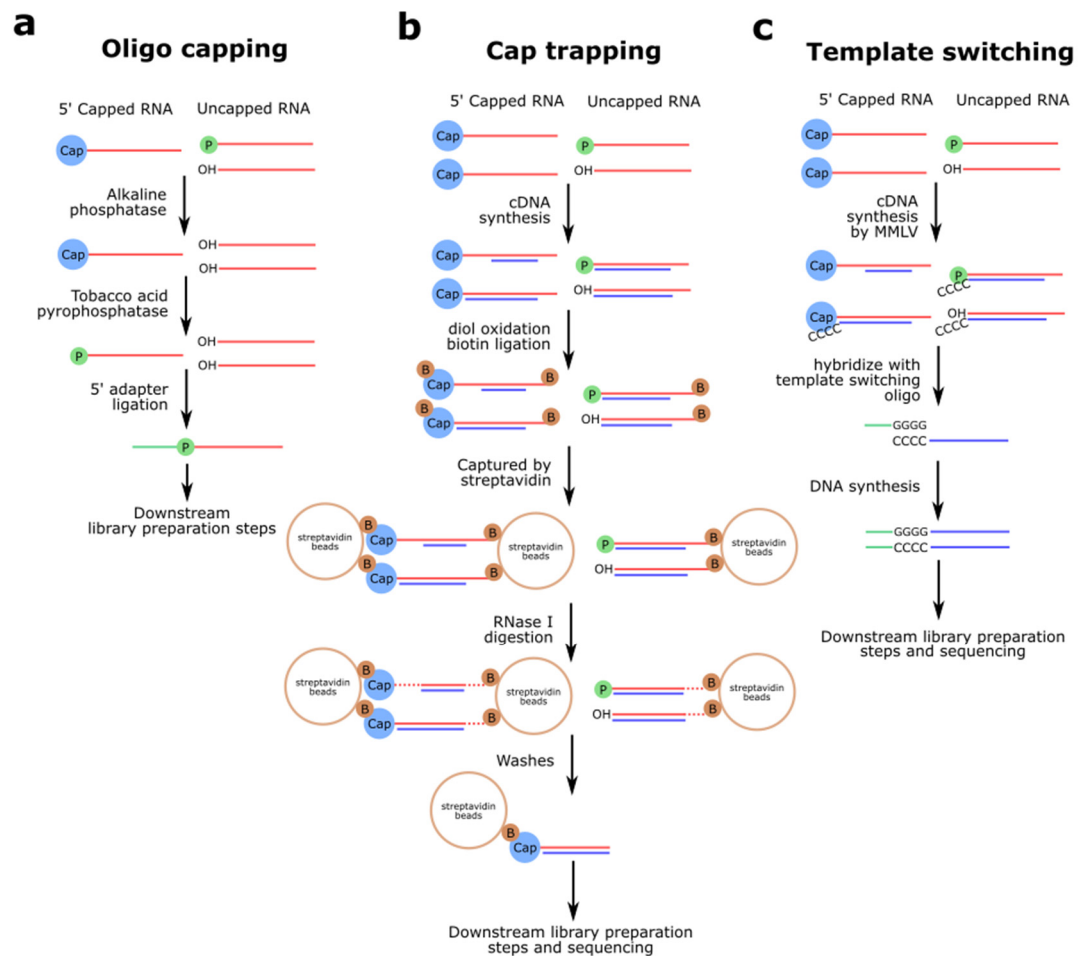


Figure 4.7 The three existing library construction strategies for mapping TSS. a. Oligo capping. b. Cap trapping c. Template switching. The RNA, DNA and adapters are represented by red, blue and green lines. The brown bubble with a capital 'B' denotes the covalently linked biotin molecules.

Oligo-capping

This method for capturing the 5' ends of RNA was initially published in 1994 by Maruyama and Sugano (Maruyama & Sugano 1994). Currently it is one of the most frequently used methods. A lot of sequencing methods for annotating TSS are based on it. For example, GRO-cap (Core et al. 2014), PRO-cap (Hojoong Kwak et al. 2013), START-seq (Nechaev et al. 2010) and 5'GRO-seq (Lam et al. 2013) use nascent RNA as the input for oligo-capping, whereas TSS-seq

(Tsuchihara et al. 2009), TIF-seq (Pelechano et al. 2013), CapSeq (Gu et al. 2012) and TL-seq (Arribere & Gilbert 2013) use mature RNA.

The general workflow for oligo-capping is illustrated in Figure 4.7a. Different sequencing techniques may vary slightly in different ways, but the principle is the same. First of all, this method removes the uncapped RNA by treating the sample with either alkaline phosphatase. The capped RNA is not affected by the enzymatic reaction due to protection by the 5' cap. However, the uncapped RNA is dephosphorylated by the alkaline phosphatase. As the following 5' adapter ligation requires a monophosphate group at the 5' end of the RNA, this enzymatic treatment prevents the uncapped RNA from proceeding to the following library construction steps. Then the sample is treated with tobacco acid pyrophosphatase (TAP), which removes the 5' cap from the capped RNA and leaves a monophosphate group at the 5' end. This makes the originally capped RNA permissive for 5' adapter ligation. After ligation, the adapter provides an anchor for the 5' end of the originally capped RNA, which allows a library to be constructed based on the 5' end of the capped RNA. After sequencing the library, the TSS can be found by mapping the positions of the RNA 5' end on the genome.

Cap-trapping

Cap-trapping is a collection of methods that can pull down the 5' caps of RNA, which allows the 5' ends of RNA to be sequenced. Then the TSS can be mapped based on those RNA 5' end tags. Diol oxidation followed by biotinylation is the most common way to pull down the 5' caps of RNA. It is utilized by Cap analysis of gene expression (CAGE). CAGE was initially developed by Shiraki et al in 2003 (Arribere & Gilbert 2013). In recent years, some improved versions of CAGE have been developed, such as nAnt-iCAGE (Murata et al. 2014) and SLIC-CAGE (Cvetesic et al. 2018), but their general principles are the same as the original CAGE.

Figure 4.7b illustrates the workflow of CAGE. First of all, the cDNA is synthesized by random priming. As reverse transcription is not 100% efficient, not all the 3' ends of cDNA can reach

the cap. Those cDNA molecules that fail to reach the cap will be removed in the later steps. After cDNA synthesis, the RNA-DNA hybrids are treated with NaIO₄, which can oxidize the diol group on the RNA 5' cap to the dialdehyde group. The dialdehyde group can further react with biotin hydrazide, which attaches biotin molecules to the RNA 5' caps. After this step, if an RNA molecule has a 5' cap, it will be biotinylated at two places, its cap and its 3' end. Otherwise, it will only be biotinylated at its 3' end. DNA is not biotinylated at all. In the next step, biotinylated RNA-DNA hybrids are captured by streptavidin beads and treated with RNase I. After several rounds of washes, only those RNA-DNA hybrids in which the RNA has the 5' cap and cDNA reaches the 5' end of the RNA will remain on the streptavidin beads. As a result, CAGE can selectively sequence the 5' end of the capped RNA, which reflects the positions of TSS.

Template switching

Moloney murine leukaemia virus (MMLV) is a reverse transcriptase, which synthesizes cDNA from RNA. When it reaches the 5' end of RNA, it will exhibit the terminal transferase activity and add a short stretch of deoxycytidine to the 3' end of cDNA. This 3' overhang of cDNA will be further hybridized to a template switching oligo, which allows the downstream library construction steps to happen (Figure 4.7c). Template switching does not distinguish the capped and uncapped RNA, so more artefacts are expected. The typical sequencing techniques that utilize template switching are nanoCAGE (Although it is called nanoCAGE, it actually uses template switching), 5' SAGE and SMART-seq (Ramsköld et al. 2012; Hashimoto et al. 2004; Salimullah et al. 2011). There is some evidence showing that template switching has sequence biases (Zajac et al. 2013; Tang et al. 2013). TSS starts with G are more likely to be found in the library prepared by template switching. Some studies have made more detailed comparisons between template switching and other methods (Adiconis et al. 2018). Generally, template switching is less frequently used for mapping TSS.

4.2.1.2 The limitation for the current TSS mapping methods and the potential solution

Most of the existing methods map TSS by sequencing RNA 5' ends, which can produce a single peak of sequencing reads at the TSS of genes. Then the TSS can be identified by finding those peaks. To distinguish them from the background peaks, a threshold is normally applied. That is, a peak is regarded as a TSS, only when its height exceeds a certain threshold. As a result, the quality or the confidence of the TSS mapping can be strongly dependent on gene expression levels. Highly expressed genes will produce a high peak at their TSS, which can be easily distinguished from the background. However, lowly expressed genes do not have enough reads at the TSS. Distinguishing them from the background peaks can be hard. Raising the threshold of peak height may cause the loss of the TSS of lowly expressed genes. On the other hand, reducing the threshold will face the risk of artefacts. Hence, the current TSS mapping methods have a bias towards the highly expressed genes.

Is it possible to find a way of mapping TSS which is more independent of gene expression levels? One of the essential features of an active TSS is that RNAPII can initiate transcription from it and move further away in one direction. If the nascent RNA is not spliced, the distance between the TSS and each RNAPII should be equal to the size of the nascent RNA associated with that RNAPII. Actually, it is particularly true for budding yeast genome, because most budding yeast genes do not have introns. There are more than 6000 annotated genes in the yeast genome and only 345 of them contain introns (based on the statistics from SGD). Most of the intron containing genes are ribosomal protein genes. Hence, once the position of RNAPII and the length of the associated nascent RNA are known, the position of the TSS can be deduced. The position of RNAPII can be mapped accurately with NET-seq. The difficulty is to know the length of nascent RNA for each of the corresponding RNAPII. To assign the RNA size information to the RNAPII signals in NET-seq, a size selection step was added to the NET-

seq protocol. After isolating nascent RNA by RNAPII immunoprecipitation, PAGE gel electrophoresis was utilized to split the nascent RNA into several fractions. Each fraction had a defined size range. A NET-seq library was constructed from the nascent RNA in each of the fractions. Although the size information for each RNAPII position was a range rather than a defined number, a characteristic pattern of NET-seq signals should be seen downstream of each TSS, if the NET-seq signals for all the size fractions were visualized (Figure 4.8). Then bioinformatic methods could be applied to map TSS by finding and matching the characteristic pattern of NET-seq signals on the genome.

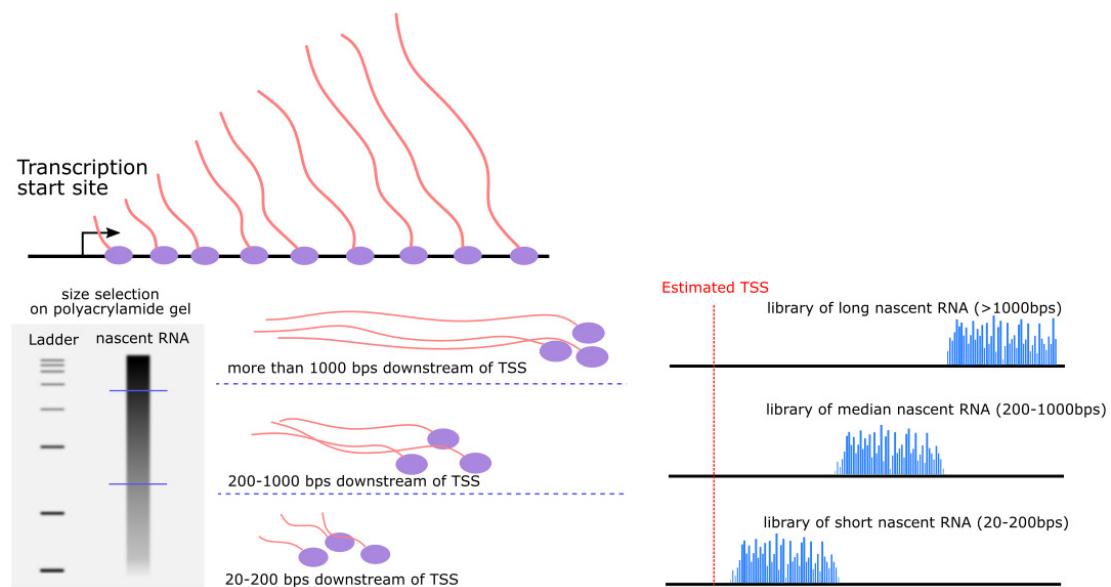


Figure 4.8 A schematic diagram showing the mechanism of the size-fractionated NET-seq. Size-fractionated NET-seq maps TSS by encoding the information of TSS positions into different size fractions and constructing NET-seq libraries from each fraction.

This method has two potential advantages. First, this method may allow the TSS of the lowly expressed genes to be mapped more easily, because it finds and matches the pattern of NET-seq signals associated with a TSS rather than finding single peaks. The pattern of NET-seq signals is largely independent of the gene expression levels. Second, this method is based on NET-seq, so it uses nascent RNA as the input material. It has the potential to map the TSS of those unstable transcripts.

4.2.1.3 Aims and plans

This chapter aims to report size-fractionated NET-seq as a new TSS annotation method. The result sections can be divided into two parts. The first part (section 4.2.2.1-4.2.2.3) is to characterize the technique and compare the TSS annotated by this technique to existing annotations of TSS. In the second part (section 4.2.2.4-4.2.2.8), three different bioinformatic analysis will be carried out to show how the new TSS annotation can be used, including the de novo identification of enriched DNA motifs, the screening of enriched genomic features and the metagene analysis. The analysis in this chapter is mainly data-driven.

4.2.2 Results

4.2.2.1 The workflow of size-fractionated NET-seq for annotating TSS

The workflow of size-fractionated NET-seq was divided into two parts, the wet lab part and the dry lab part (Figure 4.9). In the wet lab part, the nascent RNA was first isolated from the cell lysate by the original NET-seq immunoprecipitation protocols. The 3' ends of the nascent RNA molecules were immediately ligated with the 3' adapter, which marks the original 3' ends of nascent RNA. Then, by performing polyacrylamide gel electrophoresis, the 3' adapter ligated nascent RNA molecules were split into 5 fractions with different size ranges (Figure 4.9a). Next, the rest of the NET-seq library construction procedures were performed on each of the 5 fractions respectively. Finally, all the five barcoded libraries were pooled and sequenced on an illumina sequencer, which was the end of the wet lab part. The data output was the nucleotide sequences at the 3' ends of nascent RNA. Because the sequencing depths of fraction IV and V were too low, this study only used the sequencing data from fraction I, II and III. The three fractions were found to be sufficient for generating a high quality TSS map.

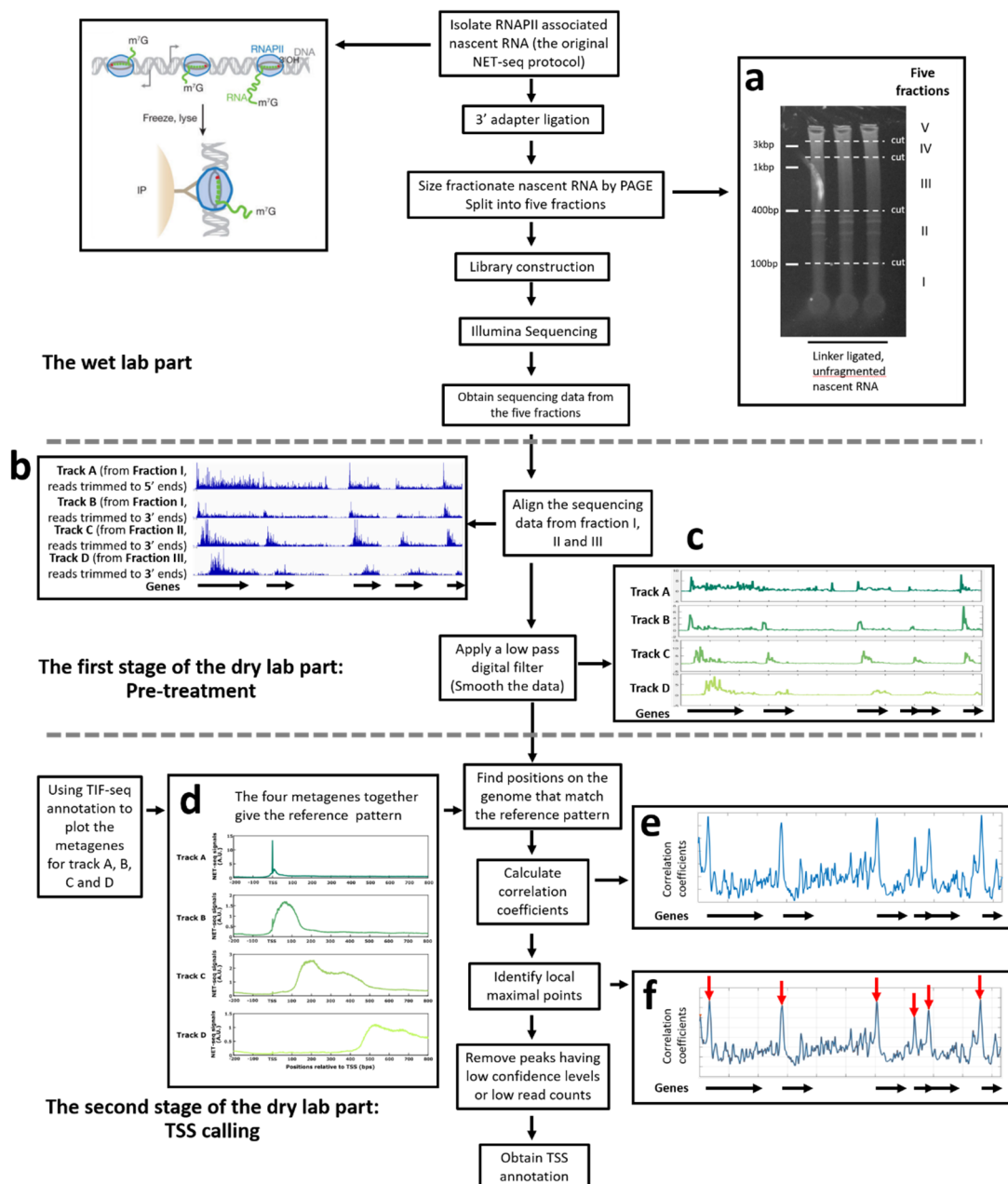


Figure 4.9 The workflow of the size-fractionated NET-seq. **a.** After linker ligation, the RNA molecules of different sizes were separated by the polyacrylamide gel and divided into five fractions (from I to V). **b.** The sequencing data from fraction I to III were visualized by IGV after alignment. Track A and B were both from fraction I. Track A represents the 5' ends of reads, while track B represents the 3' ends of reads. **c.** The data after being smoothed by the digital filter are visualized. **d.** The annotations of TSS produced by TIF-seq were utilized to plot the metagenes for track A, B, C and D. The four metagenes constituted the reference pattern. **e.** The correlation coefficients on all genome positions are visualized, which is a smooth curve with peaks. The positions of the peaks are the genomic positions of TSS. **f.** Peak calling algorithms were applied to find the positions of the peaks, which gave the positions of TSS.

The dry lab part was further divided into two stages, the pre-treatment and the TSS calling. In the pre-treatment, all the reads were first processed by the standard NET-seq data processing protocols (Churchman & Weissman 2011). The reads from fraction I were trimmed to either 5' ends or 3' ends, which generated two data tracks, A and B, respectively. The reads from fraction II and III were both trimmed to 3' ends, which generated the corresponding data track C and D (Figure 4.9b). Actually, because the RNA size of fraction I was shorter than the maximal read length of the illumina sequencer, peaks in track A should directly map to the TSS. In other data tracks, a cluster of NET-seq signals can be seen at a certain distance downstream of each TSS. Before starting TSS calling, each data track was smoothed by applying a low pass digital filter. It can remove the spikes but retain the general shapes of the peaks, which can reduce artefacts in the following TSS calling steps. After that, each data track should look like Figure 4.9c.

In the TSS calling stage, a metagene was first plotted for each of the four data track by using the existing TSS annotation created by TIF-seq (Pelechano et al. 2013) (Figure 4.9d). All the four metagenes together gave the reference pattern. If a certain nucleotide position on the genome was a TSS, the reference pattern was expected to be seen in its downstream positions. To search for the TSS on the genome, downstream NET-seq signals of each nucleotide position were compared to the reference pattern by calculating the Pearson correlation coefficient. After this step, each nucleotide position on the genome was given a Pearson correlation coefficient. When all the Pearson correlation coefficients were plotted, a smooth curve should be seen (Figure 4.9e). The TSS can be mapped by finding the peaks of this smooth curve (Figure 4.9f). Here, the heights of the peaks were largely independent of the gene expression levels. Certain thresholds can be applied to balance the sensitivity and specificity of TSS calling. In this study, the false discovery rate was set at 10^{-3} . In other words, one false positive was expected in every 1000 mapped TSS.

4.2.2.2 Size fractionated NET-seq can identify different type of TSS accurately, specifically and efficiently

The size fractionated NET-seq identified 9637 putative TSS in total from log phase cells grown in YPD. To check the identities of the TSS found by size fractionated NET-seq, the TSS positions were first compared to several existing TSS annotations. Specifically, if a putative TSS fell within a ± 70 bps window of its closest existing annotation, the putative TSS was assigned the identity of that existing annotation. It is found that 6430 of those putative TSS have been annotated. 4509 of them are TSS of protein coding genes (Figure 4.10a), while 1921 of them are TSS of known non-coding transcripts, including 344 CUTs (cryptic unstable transcript), 374 SUTs (stable uncharacterized transcripts), 704 XUTs (Xrn1-sensitive unstable transcript), 470 NUTs (Nrd1-dependent unterminated transcript) and 29 snRNA (Figure 4.10b) (Pelechano et al. 2013; Wery et al. 2016; Schulz et al. 2013b; Xu et al. 2009).

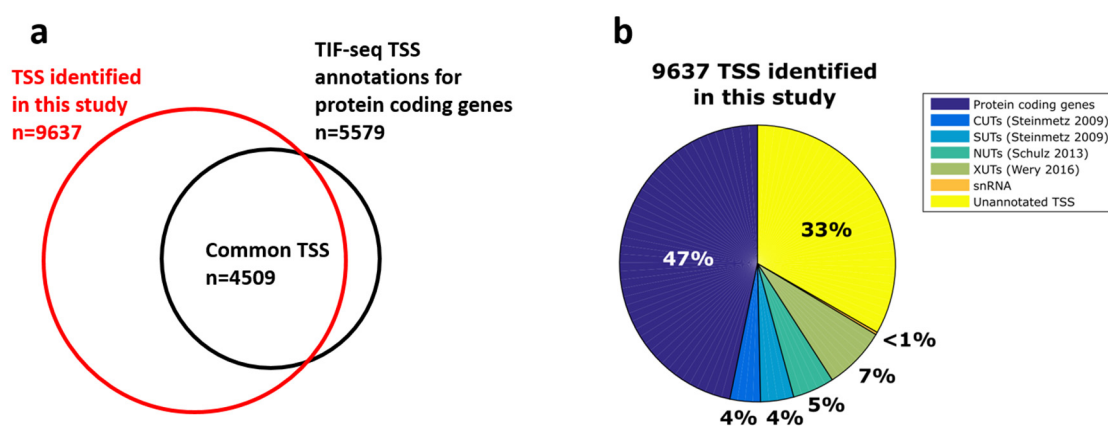


Figure 4.10 The identities of the TSS mapped by the size-fractionated NET-seq. a. The Venn diagram shows the degree of overlap between the TSS identified by size-fractionated NET-seq and the TSS found by TIF-seq. **b.** The pie chart reveals the composition of the TSS of different types of transcription units. 33% of the 9637 TSS found by size-fractionated NET-seq are not annotated.

The annotation accuracy was examined by calculating the relative distances between the TSS positions determined in this study and the existing annotations. As is shown in Figure 4.11, the deviations in positions are plotted as histograms for various types of transcription units.

The positive deviations mean the TSS determined in this study are downstream of their existing annotations, while the negative deviations mean the opposite. Generally, all histograms have a unimodal distribution centred near the zero point. The distributions for protein coding genes, SUTs and CUTs are nearly symmetrical. Although XUTs and NUTs have slightly biased distribution, their degrees of bias are not high and the centres of the peaks are still close to the zero point. These results indicate that there is no obvious systematic error for the TSS positions determined by size fractionated NET-seq. Furthermore, it is worth to note that there are only few transcription units located at both extremities of the distribution. It means that it is unlikely to assign more existing annotations to the putative TSS identified in this study via further increasing the size of the window (the current size is ± 70 bps). It suggests that the small window size is not an explanation for the remaining 3208 unannotated TSS, which comprises 33% of all TSS identified in this study. To sum up, the TSS positions determined by size fractionated NET-seq is accurate.

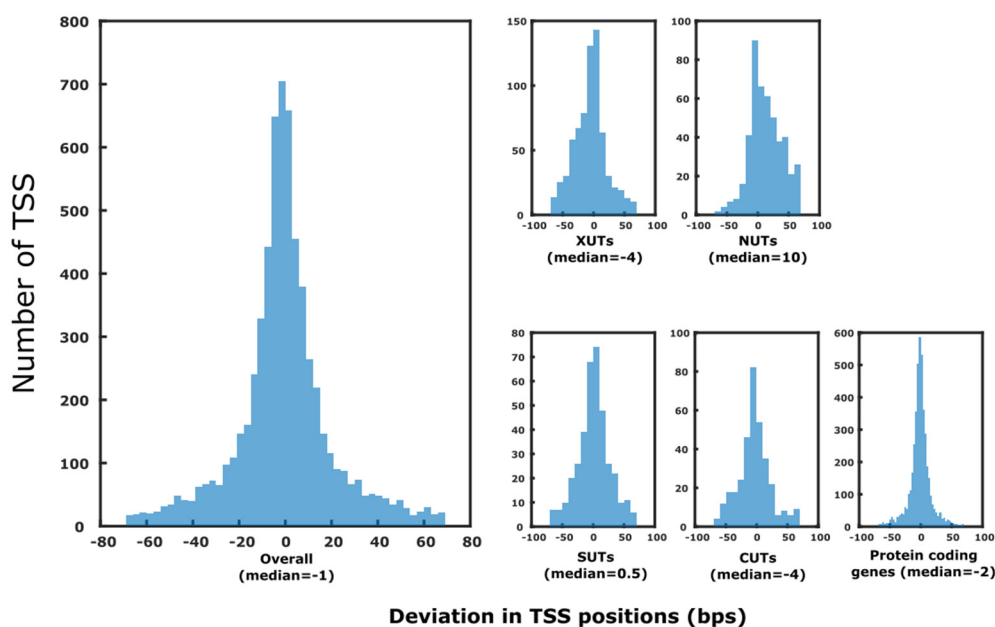


Figure 4.11 Comparing the TSS positons mapped by size-fractionated NET-seq to the existing annotations. The x-axis of each histogram is the deviations of TSS positions. The positive deviations mean the TSS determined in this study are downstream of their existing annotations, while the negative deviations mean the opposite. The median deviation for each type of TSS is written in brackets.

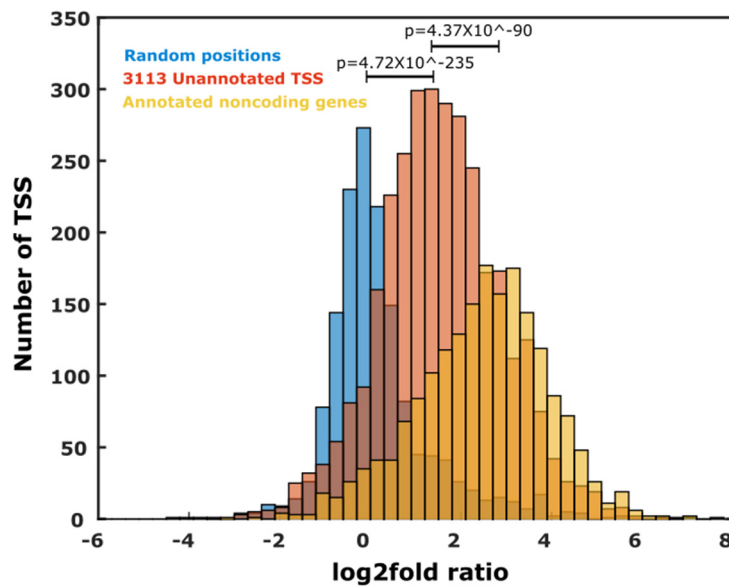


Figure 4.12 Histograms of the log₂fold ratios between the downstream reads and the upstream reads of the TSS. The annotated noncoding TSS and the unannotated noncoding TSS (identified in this study) are colored yellow and red respectively. For comparison, the distribution for random genomic positions is plotted with the blue color. The Kolmogorov-Smirnov test is utilized to check whether two distributions are significantly different. The p-values are marked between the corresponding distributions.

As size fractionated NET-seq is not 5'cap-dependent, it is worth checking whether those unannotated TSS are RNA cleavage points or random artefacts. PAS (polyadenylation sites) and splicing sites are the major known RNA cleavage points in budding yeast. Actually, among all the 3208 unannotated TSS, only 68 and 27 of them can be assigned to PAS and splicing sites respectively. It implies that the vast majority of those unannotated TSS are not RNA cleavage points, although there is still a possibility that they are cleavage points which are not reported before. To further check whether these unannotated TSS were real TSS or RNA cleavage points, the reads of ultra-low background NET-seq were counted on both sides of those unannotated TSS. A real TSS should have more reads on its downstream side. The log transformed ratios of downstream to upstream reads for all the unannotated TSS are plotted as a histogram in Figure 4.12. Although the ratios for the unannotated TSS are not as high as the ratios for the TSS of annotated noncoding genes, they are significantly higher than the ratios for random positions on the genome ($p=4.72 \times 10^{-235}$). It indicates that size fractionated NET-seq can

identify TSS specifically and most of the remaining 3113 unannotated TSS are real TSS rather than RNA cleavage points or random artefacts.

To check the efficiency of TSS mapping, size fractionated NET-seq data are visualized in IGV and compared to PRO-cap. As was mentioned in section 4.2.1.1, PRO-cap is a TSS mapping method that is also based on the nascent RNA and has been applied to yeast *S.cerevisiae* before (Mahat et al. 2016). TSS for both protein coding genes and various types of non-coding transcription units, including those lowly expressed transcription units, have been annotated efficiently by size fractionated NET-seq (Figure 4.13). Some previously unannotated TSS (highlighted with red circles) are also clearly identified. The green lane, which is the normal NET-seq, shows that the *S.cerevisiae* genome is pervasively transcribed and the genes are close to each other. Hence, it can be difficult to tell where each RNAPII signal in normal NET-seq originates (Figure 4.14). After size fractionation, many adjacent or overlapping transcription units can be separated (Figure 4.14). In comparison, PRO-cap almost has signals appearing everywhere on the genome. The signals in the regions where TSS are expected to be found is only moderately higher than those in other regions. For lowly expressed genes, the signals can be indistinguishable from the background. For some TSS, the PRO-cap signals are even completely absent in the expected region. For highly expressed genes, strong background covers the whole gene body (Figure 4.13). To conclude, size fractionated NET-seq performs better than PRO-cap and can map different types of TSS efficiently.

To summarize, size fractionated NET-seq can identify different type of TSS accurately, specifically and efficiently. In the following sections, several analysis will be performed on these TSS to explore their distributions and properties.

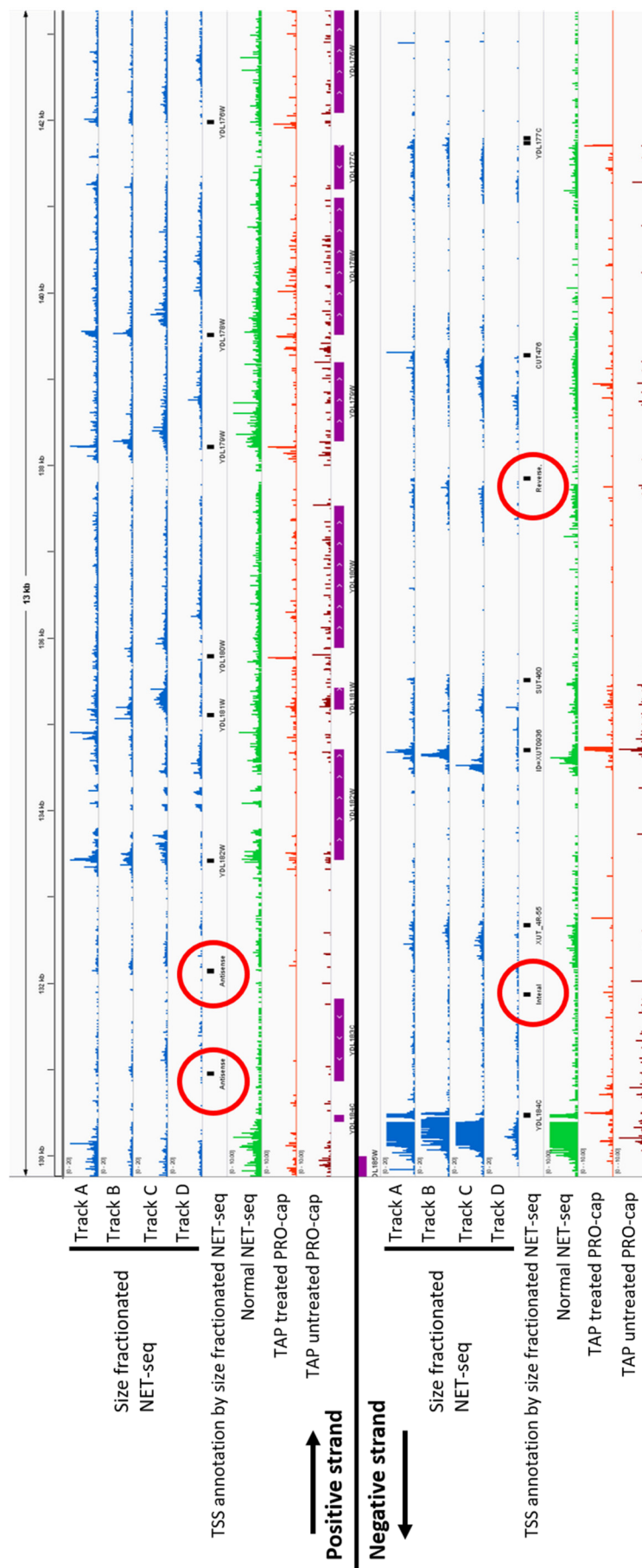


Figure 4.13. The size-fractionated NET-seq, normal NET-seq and PRO-seq libraries are visualized and compared in IGV. Previously unannotated TSS are marked with red circles. The middle purple lane shows the existing annotations for the open reading frames of all protein coding genes. It separates the figure into two parts. The upper part is the positive strand of the genome, while the lower part is the negative strand. The data of size fractionated NET-seq are the four light blue lanes on each part. The fractions of different sizes are arranged in order, followed by a black lane, which is the TSS annotations made by size fractionated NET-seq. As the technique does not tell the termination point of transcription, the length of all transcripts are set at 50bps by default. The green lane shows the normal NET-seq data, which is not size selected. For comparison, PRO-cap libraries are also attached. The red lane is TAP treated PRO-cap, which annotates cap-dependent TSS (Booth et al. 2016). The crimson lane is PRO-cap without TAP treatment, which gives a measurement of the background of PRO-cap (Booth et al. 2016).

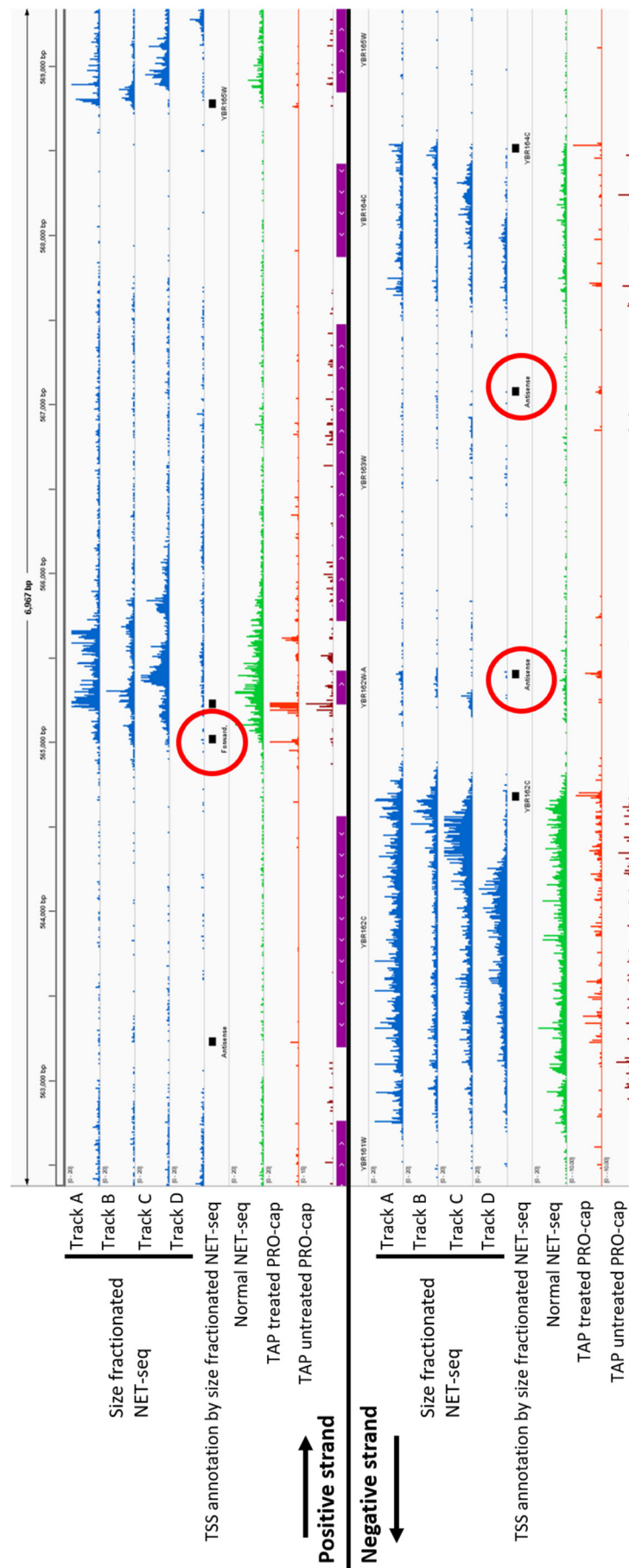


Figure 4.13 (continued)

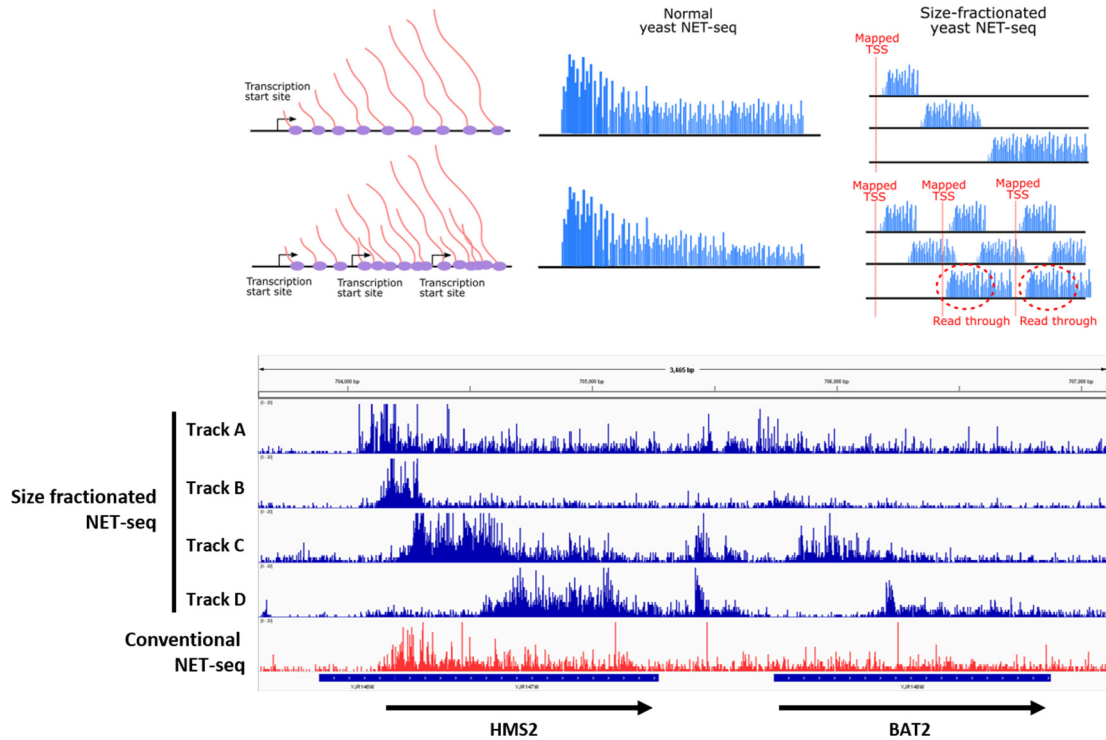


Figure 4.14 The size fractionated NET-seq can map the TSS of tandemly arranged genes. The read through events from the upstream gene make it difficult to tell the TSS of the downstream gene by normal NET-seq

4.2.2.3 The distributions of TSS around protein coding genes can be revealed by size fractionated NET-seq

All the 9637 TSS were divided into three groups, TSS for protein coding genes, TSS for annotated noncoding transcription units (previously identified in the literature) and TSS for unannotated noncoding transcription units (identified in this study by size fractionated NET-seq). Their distributions on both strands near the TSS and PAS of protein coding genes are displayed in Figure 4.15. Seven regions (from region A to G) of protein coding genes (n=5579) were examined. Regions A-D are on the sense strand of the protein coding genes, while regions E-G are on the antisense strand.

Generally, the distributions of annotated and unannotated noncoding TSS are similar, except in region A and C. The similar distribution indicates that most of the unannotated noncoding TSS are real TSS rather than random artefacts. Those unannotated noncoding TSS might

belong to some of the categories of already annotated noncoding TSS, but for some reasons, they failed to be detected previously. Major differences can be seen between the distributions of TSS for protein coding genes and noncoding transcription units in most of the regions, except in region E. Next, the distributions near the TSS and the PAS of protein coding genes are discussed sequentially.

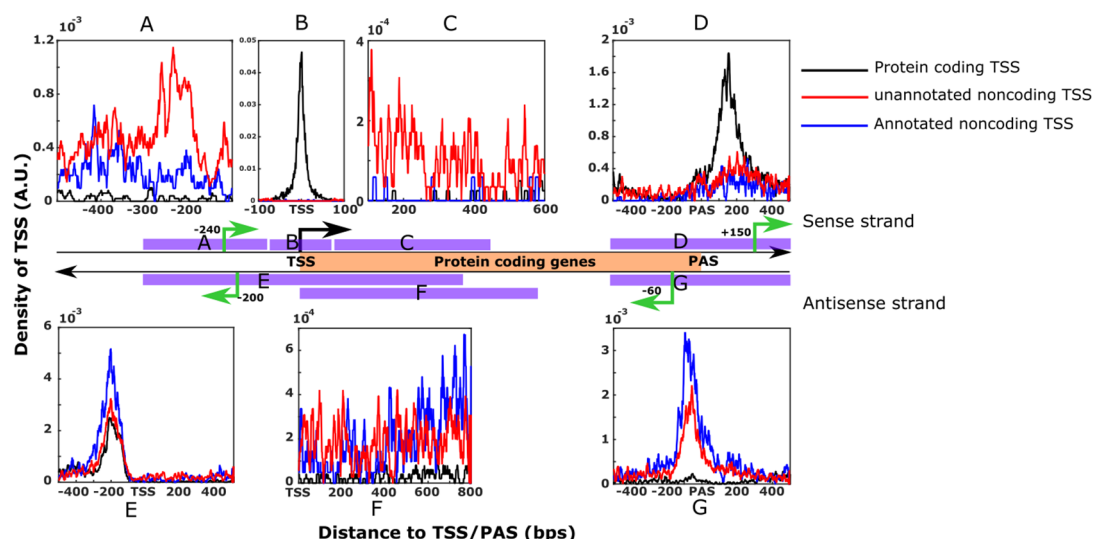


Figure 4.15 Metagenes showing the distributions of TSS around protein coding genes. The upstream and downstream of TSS and PAS on both strands were examined, which produced the metagenes plots for the 7 regions from A to G. The coverage of each region on the protein coding genes is represented with the purple bars. The frequently found TSS positions relative to the protein coding genes are represented with green arrows. For each plot, the y-axis is the percentage of protein coding genes having TSS at each single nucleotide position. The x-axis is the relative distance to TSS or PAS. The distances for downstream and upstream positions are denoted with positive and negative values respectively. It is worth to note that the scales of y-axis are variable for each small plot.

Region A, B, C, E and F are near the TSS for protein coding genes. In region A (Figure 4.15), the distribution for those unannotated noncoding TSS (red line) has a peak spanning a region between -275bps and -190bps, followed by a trough between -190bps and -125bps. This feature is not seen for the annotated noncoding TSS (blue line). This result indicates that there might be a previously unidentified class of noncoding TSS, which is located at about 240bp upstream of some protein coding genes. On the opposite strand, in region E, all three distributions are similar. They all have a peak at around -200bps. The peak for annotated

noncoding TSS (blue line) is slightly higher. In region C and F, which are the gene body, many more noncoding TSS (blue and red lines) can be found than coding TSS (black line), although there are not many TSS in either region. Interestingly, the unannotated noncoding TSS in region C seem to be arranged in clusters. The clusters seem to have a periodicity of 160bps, which is similar to the periodicity of nucleosome dyads. A similar pattern of TSS is also observed in mammalian cells by CoPRO (Tome et al. 2018).

Region D and G map to the PAS of protein coding genes. In region D, the TSS for protein coding genes (black line) give a high peak centring at +150bps of the PAS, while the peaks for noncoding TSS (blue and red lines) are much lower and positioned slightly downstream. As those false 'TSS' mapping to PAS were not excluded in this analysis, all three distributions have a small peak at the PAS in region D. In region G, noncoding TSS (blue and red lines) but not coding TSS (black line) have a peak spanning the region between -140bps and -60bps. To summarize, the TSS for noncoding antisense and coding sense transcripts are preferentially found at particular positions upstream and downstream of PAS respectively.

Now, there are two questions regarding region D and G. First, is there any relationship between the peaks in both regions? Second, what genomic features can be associated with these peaks? The following sections attempt to answer the two questions.

4.2.2.4 In the region flanking PAS, an upstream noncoding antisense TSS and a downstream coding sense TSS have a strong tendency to co-occur

To investigate the potential relationship between the peaks of coding and noncoding TSS seen in region D and G respectively, a single-sided Fisher's test was performed to examine whether there exists certain co-occurrence relationship between different types of TSS flanking the PAS of protein coding genes. As the distributions for annotated and unannotated coding TSS are similar, they were pooled together as a single group in all the following analysis. The null hypothesis for the statistical test is that there is no co-occurrence relationship between the upstream TSS and the downstream TSS. The p-values for each of the four combinations are summarized in Table 4.1.

P-values	Downstream coding	Downstream noncoding
Upstream coding	5.11×10^{-05}	1.32×10^{-08}
Upstream noncoding	1.51×10^{-311}	9.15×10^{-12}

Table 4.1 The positive correlations between an upstream (relative to the PAS) coding/noncoding TSS and a downstream (relative to the PAS) coding/noncoding TSS were tested by single sided Fisher's tests. The p-values for all four combinations are shown in the table. The combination with a highly significantly positive correlation is highlighted in bold.

P-values	Downstream coding	Downstream annotated noncoding	Downstream unannotated noncoding
Upstream coding	5.11×10^{-05}	1.57×10^{-06}	0.000907585
Upstream annotated noncoding	4.67×10^{-159}	0.140176503	0.000390385
Upstream unannotated noncoding	4.09×10^{-96}	0.01240602	3.38×10^{-06}

Table 4.2 Similar to Table 4.1, but the noncoding TSS was further divided into the annotated and unannotated noncoding TSS, which produced nine combinations in total. The p-values for all nine combinations are shown in the table. The combination with a highly significantly positive correlation is highlighted in bold.

All four combinations give low p-values, which means that there is a general correlation between the upstream and the downstream TSS. In other words, the TSS flanking a PAS has a tendency to occur in pairs. Notably, the p-value for the combination of an upstream noncoding TSS and a downstream coding TSS is much lower than the other three combinations. In fact, if the annotated and unannotated noncoding TSS are not pooled together, the results remain the same (Table 4.2). Both the upstream annotated and the upstream unannotated noncoding TSS strongly correlate with the downstream coding TSS. It also supports the view that the unannotated noncoding TSS identified in this study are real TSS rather than artefacts. All these results show that in the region flanking a PAS, an upstream noncoding TSS has a strong tendency to co-occur with a downstream coding TSS. Although currently it is not known whether such pairs of TSS are divergent promoters, the word 'divergent promoter' is still used, in order to denote such an arrangement of TSS near the PAS of protein coding genes.

4.2.2.5 The DNA sequences of divergent promoters are significantly enriched for the motif of the Reb1 binding site

To check whether there are any DNA sequences preferentially associated with the PAS regions having divergent promoters, the MEME (Multiple Em for Motif Elicitation) server was utilized to perform a de novo discovery of enriched DNA sequences for the PAS regions having divergent promoters (Bailey et al. 2009). MEME is based on the Hidden Markov Model (HMM) and does not rely on the prior knowledge of DNA binding sequences of DNA binding factors. Although its sensitivity is low, it can discover motifs in a de novo and unbiased way.

Specifically, a total of 5579 PAS regions of protein coding genes were considered. For comparison, these PAS regions were divided into 4 groups, depending on whether they had an upstream noncoding TSS and whether they had a downstream coding TSS. The MEME was performed on the DNA sequences in each group to discover the enriched DNA motifs relative to the total 5579 PAS regions (Figure 4.16). To further control for the false discovery of motifs,

a fifth group of 400 randomly selected PAS regions from the total 5579 PAS regions was analysed by the MEME server in the same way (Figure 4.17).

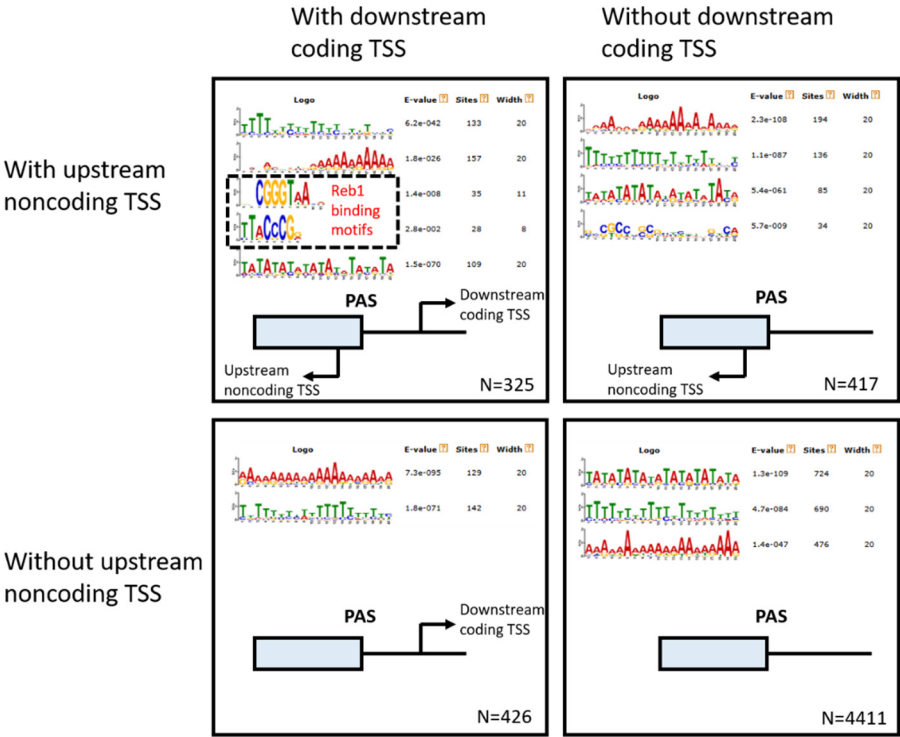


Figure 4.16 The enriched motifs for the four different types of PAS regions. Depending on whether the PAS regions had an upstream noncoding TSS and/or a downstream coding TSS, all the PAS regions were divided into four different types. The number of PAS regions for each type is indicated in the bottom-right corner. The upper left is the PAS regions associated with the divergent promoter (both an upstream noncoding TSS and a downstream coding TSS).

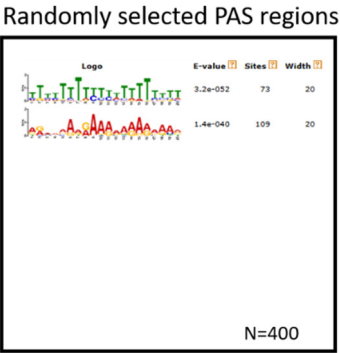


Figure 4.17 The enriched motifs for randomly selected PAS regions.

Several significantly enriched motifs are found for all five groups. Among them, the polyadenosine and polythymidine motifs are found in all five groups, including the fifth random control group, which suggests that these two motifs are false discoveries or simply

the sequences of the polyadenylation signals. Although the AT alternating motif is not found in the control group, it lacks specificity in the four experimental groups, so it is not considered for further analysis. However, there are three motifs that are only found in individual groups. They are CGGGTAA, TTACCCG and a GC-rich motif. Interestingly, the first and the second motifs are actually complementary sequences and are only found in the group for the divergent promoters. Both of them have very high information contents indicated by the heights of their sequence logos, which implies that they are well conserved. By searching the sequence of the complementary motifs in the YEASTRACT (Yeast Search for Transcriptional Regulators and Consensus Tracking) database (Teixeira et al. 2018), it is found that the complementary motifs are actually Reb1 binding sites. Reb1 is a transcription factor that has been shown to act as a roadblock on the DNA to terminate the transcription of RNAPII (Candelli et al. 2018; Colin et al. 2014). It can also create a nucleosome depleted region by recruiting chromatin remodelers (Krietenstein et al. 2016; Hartley & Madhani 2009; Bai et al. 2011). The median positions of the two complementary motifs are 28bps and 35bps downstream of the PAS respectively. Interestingly, both motifs are very close to the middle position between the upstream peak of the noncoding TSS (-60bps) and the downstream peak of the coding TSS (+140bps). The third GC-rich motif is only found in the group with an upstream noncoding TSS but without a downstream coding TSS. Its median starting position is about 44bps downstream of the PAS. Unlike the Reb1 binding site, this motif is not well conserved, as its information content is relatively low.

To conclude, the divergent promoters are significantly associated with the motif of the Reb1 binding site.

4.2.2.6 The divergent promoters are associated with Reb1 binding

To further examine whether the binding of Reb1 protein was significantly associated with the PAS regions with divergent promoters, the Reb1 ChIP-seq signals (Bosio et al. 2017) were

calculated for each of the four experimental groups in Figure 4.16. The metagene of Reb1 ChIP-seq signals flanking the PAS was plotted for each of the four groups (red lines in Figure 4.18). To indicate the significance of the peaks for each group, the same number of PAS regions were randomly selected from the total 5579 PAS regions. For each group, this process was repeated 500 times to simulate the background levels. The median background levels are represented with the blue line, while the top and bottom 1% of the background simulations are indicated by the black and green lines.

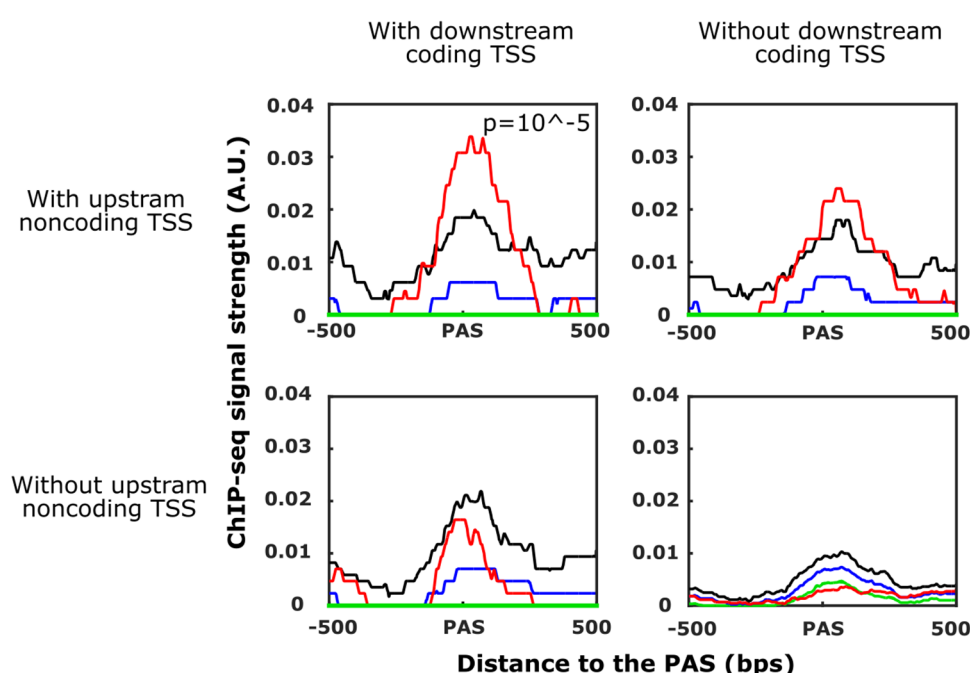


Figure 4.18 The Reb1 ChIP-seq signals in each of the four types of PAS regions. The upper left one is the PAS regions associated with a divergent promoter. It is significantly enriched for Reb1 ChIP-seq signals ($p=10^{-5}$). The black, blue and green curves represent the ChIP-seq signals of the top 1%, median and bottom 1% of the one million random simulations. The ChIPseq data was retrieved from the GEO database of NCBI (accession number: GSM2143115, (Bosio et al. 2017)).

For the group of PAS regions with divergent promoters, its Reb1 ChIP-seq signals are much stronger than the top 1% random simulation. Actually, after simulating the background for 1000000 times, it is estimated that the p value for this group is 10^{-5} . The ChIP-seq signal of Reb1 peaks at 40bps downstream of the PAS, which is consistent with the positions of Reb1 binding motifs reported in the previous section. For the groups only having a downstream

coding TSS or an upstream noncoding TSS, their signals are not significantly higher than the background ($p>0.01$). Interestingly, for the fourth group, its signal is close to the bottom 1% of random simulation, which indicates that the binding of Reb1 is less likely to be associated with a PAS without surrounding TSS. To sum up, the results suggest that PAS regions with divergent promoters are associated with the Reb1 binding.

4.2.2.7 Coding and noncoding TSS are associated with different genomic features

In section 4.2.2.3, region E shows that both coding and noncoding TSS can appear at 200bps upstream of protein coding TSS. One question is whether these coding and noncoding TSS in region E are associated with different genomic features in their promoter regions. To answer this question, this section uses statistical tests to check the differential enrichment of various genomic features in the promoter regions of the coding and noncoding TSS in region E.

In region E of Figure 4.15, there are 1015 and 739 promoter regions (from -300 to -100 in region E) that are associated with the coding and noncoding TSS respectively. The noncoding transcription units are much more likely to have a PAS nearby (Figure 4.15, region G). To avoid finding genomic features that were actually related to PAS rather than noncoding TSS, if the TSS had PAS within their promoter regions, it was excluded from the analysis. In the analysis, there were 1453 different genomic features to be tested (based on the datasets collected and organized by Mellor lab), including various histone modifications and chromatin associated factors. Each genomic feature has a value to indicate its amount in the promoter region for each TSS. As a result, both coding and noncoding TSS have a distribution of values for each genomic feature in their promoter regions. For a certain genomic feature, if its distribution of values is significantly higher or lower for the coding or noncoding TSS, this genomic feature may be of interest, because it indicates that this genomic feature is enriched/depleted for coding or noncoding TSS. A two-sample Kolmogorov-Smirnov test was utilized to examine whether a genomic feature was enriched or depleted in the promoter regions of coding or

noncoding TSS. The threshold for the significance test was set at the p-value of 1×10^{-5} . Any genomic features having their p-values lower than that were considered significantly enriched for coding or noncoding TSS. This p-value was equivalent to an E-value of 0.015. In other words, it was expected to get 0.015 genomic feature by chance from the analysis.

a	GO	p-values (Enriched for coding)
	H3K4me3	1.12E-102
	H3K79me3	8.76E-68
	H2BK123ub	3.09E-32
	Htz1	1.89E-26
	H3K14ac	1.01E-20
	Swr1	2.79E-19
	Bdf1	2.14E-12
	H3K36me3	1.89E-11
	H4K12ac	2.63E-10
	H4K5ac	1.66E-09
	Aor1	2.88E-09
	NFR sizes	2.17E-08
	H2AK5ac	2.25E-08
	TAF1	3.10E-08
	Bye1	1.18E-07
	H4K8ac	1.21E-07
	Spp1	2.64E-07
	H3K79me2	2.94E-07
	Tfb1	6.70E-07
	TFIIB	7.06E-07
	H3K23ac	8.38E-07
	H4ac	9.27E-07
	Vps72	1.24E-06
	Sua7	2.20E-06
	Ssl2	2.85E-06
	Itc1	2.98E-06
	Tfg1	3.61E-06
	Cet1	4.51E-06
	Paf1	9.57E-06

b	GO	p-values (Enriched for noncoding)
	H3K4me1	1.18E-70
	H3R2me2	4.00E-30
	H3 acetylation	1.54E-27
	H3K4me2	3.28E-26
	Isw1	3.55E-25
	H3K4ac	4.22E-21
	H4K20me	1.27E-20
	H3K56ac	1.83E-20
	Rsc1	3.87E-18
	Spn1	1.64E-14
	Histone turnover	9.04E-14
	H4 acetylation	1.35E-13
	Elf1	4.39E-13
	Nrd	1.12E-12
	H3K27ac	1.16E-07
	Rsc9	1.35E-06
	H2A.ZK14ac	1.87E-06
	Spt4	3.93E-06
	Nhp6a	4.84E-06
	H3K9ac	6.63E-06

Table 4.3 The promoter regions of the coding (a) and noncoding (b) transcription units are enriched for different genomic features. The different types of genomic features are marked with different colours. Red, blue and yellow represent chromatin remodelling factors, histone modifications and RNAPII transcription related factors respectively.

Table 4.3 a and b show the genomic features that are significantly enriched for coding TSS and noncoding TSS respectively. Based on the current threshold of significance levels, there are 29 and 20 genomic features significantly enriched for coding and noncoding TSS respectively. For coding TSS, they are most strongly enriched for H3K4me3. Bye1, which has been shown to interact with H3K4me3, also has a higher level for coding TSS (Kinkelin et al. 2013). In addition,

coding TSS are also associated with H3K79me3, H3K36me3, H2bK123ub, H3K14Ac, H4K12Ac and Htz1. In terms of chromatin remodelling complexes, coding genes also enrich for several components of SWR1 complex, which exchanges Htz1 for H2A. In terms of transcription, coding TSS have more Paf1, which is consistent with the literature (Fischl et al. 2017). In terms of noncoding TSS, they are enriched for H3K4me1/2 rather than H3K4me3. Apart from that, H3R2me2a, H4K20me, H3K4Ac and H3K56Ac are also strongly associated with noncoding TSS. In terms of chromatin remodelling factors, Isw1 and Rsc1 are enriched for noncoding TSS. For transcription associated factors, noncoding TSS have slightly more Spt4 binding. Spt4 has been shown to promote RNAPII transcription elongation and the *spt4Δ* strain has a higher promoter proximal pausing than the wild-type (Churchman & Weissman 2011). The enrichment of Spt4 for noncoding TSS suggests that noncoding transcription units might have less promoter proximal pausing for RNAPII transcription. It is tested in the next section by examining the metagene profiles of NET-seq. Finally, Nrd1 is also found to be enriched for those noncoding TSS, which is as expected, because those noncoding TSS contain a significant proportion of NUTs.

To conclude, coding and noncoding TSS are associated with different genomic features, including histone modifications, chromatin remodelling factors and transcription factors.

4.2.2.8 Coding and noncoding transcription units have a similar distribution of RNAPII in their promoter proximal regions

Many transcription elongation factors, such as Paf1, Dst1 and Spt4 have been shown to influence the distribution of RNAPII in the proximal region of genes (Churchman & Weissman 2011; Fischl et al. 2017). The last section shows that coding and noncoding TSS have differential enrichment for elongation factors and other genomic features. To explore whether these differential enrichment of genomic features were associated with different RNAPII distribution profiles, metagene profiles of NET-seq were plotted for both coding and

noncoding TSS in various regions of Figure 4.15. Section 4.2.2.3 has already shown that noncoding TSS is more likely to overlap with other transcription units, which may influence the metagene profiles for noncoding TSS. Hence, during the metagene analysis, TSS were excluded if they have other downstream transcription units on either strand.

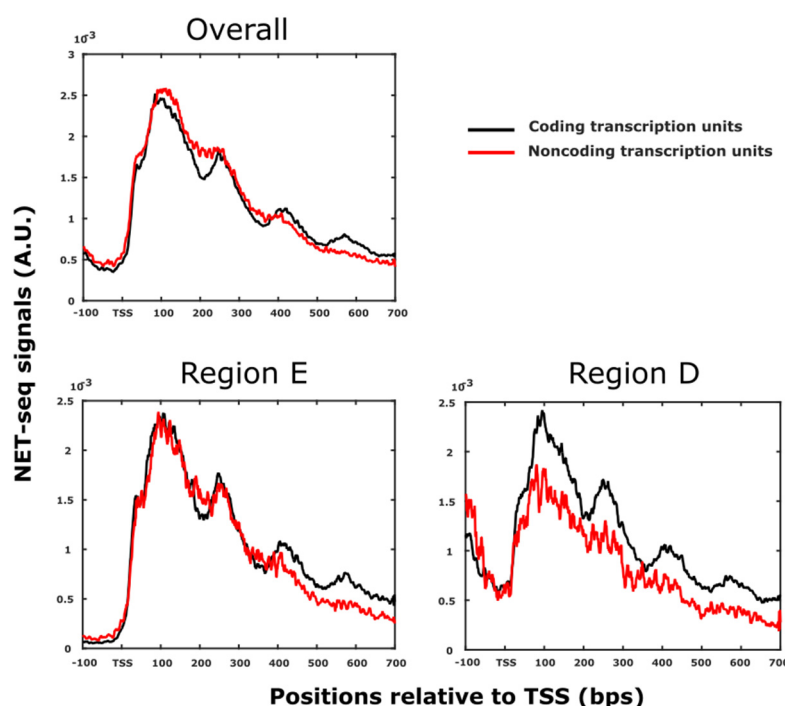


Figure 4.19 The comparison between the metagene profiles of the coding and noncoding transcription units. ‘Region E’ and ‘Region D’ mean that the corresponding plot is assessing the metagene profiles of TSS in ‘Region E’ or ‘Region D’ of Figure 4.15.

Interestingly, the metagene profiles of the downstream regions of coding and noncoding TSS are similar. The metagenes for all coding and noncoding TSS are plotted in Figure 4.19. Generally, they have a highly similar proximal biased RNAPII distribution with periodic peaks. However, the peaks downstream of noncoding TSS are fuzzier than those for coding TSS and start to diminish after 3 peaks. It could be explained by the shorter lengths and lower expression levels of noncoding transcription units. As is shown in 4.2.2.3, TSS may have different surrounding environments. A TSS can be next to a PAS or another TSS. Hence, the overall metagene plots may not be able to represent the situations in different regions of Figure 4.15. To further examine the distributions of RNAPII downstream of coding and

noncoding TSS, the NET-seq metagene profiles were plotted for TSS in regions E and D respectively. In region E, no obvious differences can be found between coding and noncoding TSS, which is consistent with the observation made from the overall metagenes. In region D, the noncoding TSS has slightly fewer signals in the gene body than the coding TSS. However, the profile of the noncoding TSS does not seem to be more flattened.

To sum up, except the noncoding TSS in region D, the NET-seq metagene profiles for both coding and noncoding TSS are highly similar. It implies that the proximal biased RNAPII distribution profile is relatively robust and cannot be significantly influenced by the differential enrichment of genomic features between coding and noncoding TSS as well as the different surrounding environments of TSS.

4.2.3 Discussion

4.2.3.1 The advantages of the size fractionated NET-seq

Size fractionated NET-seq maps TSS by looking for a particular pattern among several libraries rather than merely finding peaks higher than a threshold in a single library. It allows lowly expressed transcription units to be detected with equally high sensitivity and confidence. In fact, the majority of the 3113 newly identified TSS are noncoding TSS and have lower transcription levels in the log phase batch culture condition than other annotated TSS. It might be the reason why they are not annotated in previous studies.

As the starting material of size fractionated NET-seq is nascent RNA, the identification of TSS is relatively less influenced by the nuclear surveillance and cytoplasmic degradation pathways. The majority of noncoding TSS, including NUTs, SUTs, CUTs and XUTs, can be detected in the wild-type condition. It allows the TSS of unstable transcripts to be determined without mutating the cellular RNA degradation systems, which can avoid the potential artefacts and stress responses caused by mutating RNA degradation factors. Although PRO-cap also uses

nascent RNA, size fractionated NET-seq outperforms it in terms of the specificity and sensitivity.

4.2.3.2 The limitations and future improvements

The size fractionated NET-seq is not dependent on the 5'cap, which can make it difficult to distinguish the real TSS from co-transcriptional cleavage events, such as splicing at the intron-exon boundaries and the cleavage at the PAS. In this study, they are not major issues, because the genome of budding yeast does not have many genes containing introns and section 4.2.2.2 has found that only few putative TSS correspond to the two types of cleavage. However, there is still a possibility that the unannotated TSS identified in this study are some unknown RNA cleavage points. To perform a cross check, it may be possible to divide the starting material of size-fractionated NET-seq into two parts. One part is performed with the normal size fractionated NET-seq. The other part can be used for constructing the CAGE library. The CAGE library only maps the 5' ends of nascent RNA that have a cap. The TSS mapped by both parts can be compared to further increase the accuracy and specificity of TSS annotation.

As the size-fractionated NET-seq technique was developed before the ultra-low background protocol of NET-seq, the size fractionated NET-seq discussed in this chapter is still based on the original NET-seq protocol, which has a high background. The direct consequence of the high background was that the longest fraction (fraction V) failed to align uniquely. Actually, more than 95% of the reads in fraction V mapped to rRNA, because its size range overlapped with rRNA. If the new NET-seq immunoprecipitation protocol can be applied, the longest fraction may produce enough usable reads to study the transcription read through events (Nguyen et al. 2014; Mellor et al. 2016). For example, if a significant number of reads from the longest fraction are found in the promoter proximal region of a gene and the gene has an upstream gene in the same direction, those reads are very likely to reflect the read through transcription events from the upstream gene. Without size fractionation, it can be difficult to

tell whether the signals of NET-seq in the promoter proximal region of a gene actually originate from its TSS or the TSS of its upstream gene. In addition, as the new NET-seq protocol greatly reduces the background and increases the proportion of nascent RNA in the starting material, the sensitivity of size fractionated NET-seq may be further improved. Because the size fractionation step is performed after immunoprecipitation, there is no obvious obstruction for applying the new NET-seq protocol to size-fractionated NET-seq.

The growth condition of yeast cells can largely determine the spectrum of TSS to be mapped. It can be a limitation for the TSS annotation produced in this study, although it is not a limit for the technique itself. Currently, the cells are harvested from the log phase batch culture in the rich medium. The expression of some transcription units may only happen in special culturing conditions, such as various stress conditions. To generate a more comprehensive map of TSS, size fractionated NET-seq may need to be performed on cells grown in various culture conditions. One solution is to perform this technique on samples taken from various phases of the YMC, where yeast cells experience different physiological states and have dramatically different transcription profiles.

4.2.3.3 The possibility of using the same size selection method to annotate the mammalian genome

Currently, size fractionated NET-seq is mainly designed for mapping TSS on the yeast genome. To further apply it to other genomes, such as the mammalian genome, there are two major problems. The first problem is that unlike budding yeast genome, introns are much more prevalent on the mammalian genome. The presence of introns not only increases the false discovery rate but also makes the length of nascent RNA not equal to the distance between RNAPII and the TSS, due to co-transcriptional splicing. The potential solution is to take fractions of small sizes to make the library, such as 50nts-75nts, 75nts-100nts and 100nts-125nts. In these short fractions, the nascent RNA for most genes have not encountered the

first exon-intron boundary. The length of the nascent RNA is still equal to the distance between RNAPII and the TSS. Although those intron-exon junctions will still be mapped, they can be excluded either bioinformatically or via comparing to the CAGE library as mentioned in section 4.2.3.2. The second problem is that mammalian genes are much longer than the yeast genes. It means there is much less mammalian nascent RNA falling into the same size range than yeast nascent RNA. For example, if 1 μ g of nascent RNA is isolated from the mammalian cells and the yeast cells respectively and the same size range of 100-200nts is selected after PAGE electrophoresis, much less RNA will be obtained from the mammalian nascent RNA, because the 1 μ g of mammalian nascent RNA contains any sizes from dozens of nucleotides to hundreds of thousands of nucleotides. On the contrary, the median length of yeast genes is only 1000 nucleotides. To solve this problem, the more mammalian cells need to be harvested and the library construction method has to be improved to handle a small quantity of nascent RNA.

If the two problems above can be solved, the potential way of using the size selection method to annotate the mammalian genome is to combine it with 4sU-seq. The nascent RNA isolated by 4sU labelling can be fractionated into different size ranges. After 3' end linker ligation, libraries can be constructed from each of the size ranges to sequence to 3' ends of the 4sU labelled RNA.

Chapter 5.

Discussion

5 Discussion

5.1 Future plan

5.1.1 Developing precise run-on NET-seq/TEF-seq

PRO-seq and NET-seq are currently the two major ways of studying transcription with single nucleotide resolution (Churchman & Weissman 2012; Hojoong Kwak et al. 2013). The information they provide is complementary and allows the assessment of different aspects of transcription. NET-seq utilizes immunoprecipitation to isolate nascent RNA. By using different antibodies, it can selectively enrich signals from a particular state of RNAPII. For example, it can select RNAPII with different post-translational modifications on its CTD by using antibodies, such as anti-Rpb1Ser5P and anti-Rpb1Ser2P (Nojima et al. 2015). Moreover, by tagging a certain transcription elongation factor with an epitope tag, RNAPII in a state where it is associated with that particular elongation factor can be isolated, which is the TEF-seq (transcription elongation factor sequencing) developed in Mellor lab (Fischl et al. 2017). However, NET-seq normally cannot tell whether the RNAPII is pausing or elongating. On the contrary, PRO-seq uses run-on reactions to tag elongating RNAPII with biotinylated nucleotides, so it mainly enriches elongation competent RNAPII (Hojoong Kwak et al. 2013), albeit there is a debate on, to what extent, the in situ run-on reaction conditions of PRO-seq can reflect the in vivo transcription elongation conditions. Without immunoprecipitation, it can be hard to directly relate the elongating or pausing states to different CTD modifications or transcription elongation factors. When the two techniques are performed on budding yeast, the metagene profiles from the two techniques can be quite different. NET-seq produces a metagene profile with a broad promoter proximal peak, while the profile for PRO-seq is almost flat (Booth et al. 2016; Churchman & Weissman 2011; Fischl et al. 2017). It can be explained by RNAPII being more likely in the pausing state in the promoter proximal region. However,

the protocols for the two techniques are also dramatically different. There is a chance that the different procedures of the two techniques may also lead to different metagene profiles.

Here I propose a technique called ‘Precise run-on NET-seq’, which can be a potential way of unifying and drawing on the strengths of both techniques (Figure 5.1). In the first step, the biotin run-on of PRO-seq and the cell harvesting of NET-seq are performed separately, which produces two batches of cells. Ideally, all the cells should be subjected to the biotin run-on reaction. However, it might be expensive to do so, because NET-seq requires many more cells as the starting material. In the second step, the cells from both batches are mixed together and proceeded to the normal immunoprecipitation procedures of NET-seq or TEF-seq. The isolated nascent RNA will be ligated with the 3’ end linker. As the elongating competent RNAPII produces 3’ end biotin tagged nascent RNA during the biotin run-on, these nascent RNA can be separated by streptavidin beads after the 3’ end linker ligation. The flow through of this purification step will also be retained. Then two libraries will be constructed separately from the eluate and the flow through. After sequencing both libraries, the proportion of elongating RNAPII at each nucleotide position can be revealed by the ratios of signals from the two libraries. In this way, the technical differences between NET-seq and PRO-seq can be minimized.

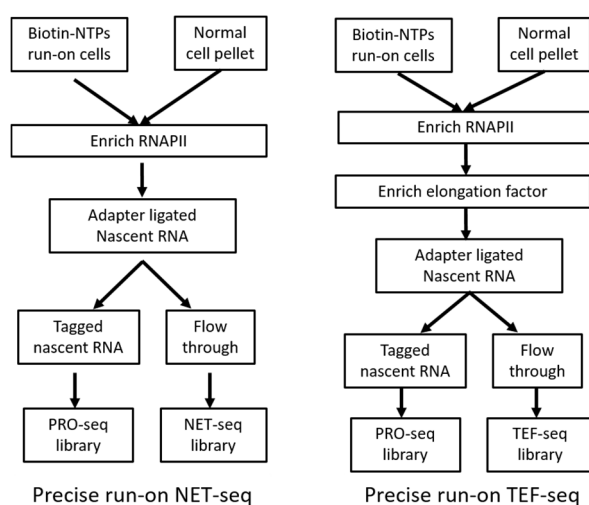


Figure 5.1 The general workflows of the precise run-on NET-seq and the precise run-on TEF-seq

Furthermore, precise run-on NET-seq/TEF-seq can potentially study the relationship between RNAPII kinetic states (pausing or elongating) and RNAPII protein associations or modifications (the binding of transcription elongation factors and the modifications on Rpb1-CTD). For example, certain antibodies can be used during the immunoprecipitation step to enrich RNAPII associated with a particular transcription elongation factor. The ratios of the two libraries (from the eluate and the flow through mentioned in the last paragraph) can be compared to those obtained by using general anti RNAPII antibodies. If the proportion of elongating RNAPII is significantly higher or lower in the libraries enriched for the transcription elongation factor. The elongation factor may be associated with elongating or pausing RNAPII. Different genes such as coding and noncoding genes can be clustered to see whether there is a difference in the function of a particular transcription elongation factor on different genes.

5.2 Concluding remarks

This study investigates the relationship between acetyl-CoA metabolism, chromatin acetylation and gene expression. Four main achievements associated with this topic have been made. First, this study uses the YMC system as an example to illustrate that a chromatin remodelling complex can play an important role in the interplay between cellular metabolism and gene expression. Second, the identification of *ACO2* as the downstream effector of the hypothesized acetylation-recruitment pathway shows that there exists a reciprocal regulation between cellular metabolism and gene expression. Third, although it has not been proved, the results in this study support the view that histone acetylation marks can activate gene expression via recruiting the bromodomain containing chromatin remodelling complex, the RSC complex. Finally, this study also integrates the experimental results from this study and the literature to propose the 'Sozu' model, which can shed light on the molecular mechanism behind the phenomenon of the YMC.

Two new techniques are also developed in this study. First, the ultra-low background protocol of yeast NET-seq can successfully suppress the background of yeast NET-seq, which can reduce the cost and increase the usability of yeast NET-seq. Second, the size fractionated NET-seq brings a new strategy of mapping TSS and has successfully identified 9637 TSS in budding yeast genome. More than 30% of them are new. Both techniques can pave the way for using the YMC to study the relationship between transcription and transcripts.

Appendix

S1 At the global level, there is no obvious change in the distribution of RNA polymerase II on genes

Section 3.2.11 has shown that genes having a peak of histone H3 acetylation at point B are significantly more likely to lose the peak of transcripts in TW1 by deleting *RSC2*. Currently, it is not known whether the RSC complex activates gene expression via increasing transcription initiation or promoting transcription elongation. In section 1.7.4, some evidence suggests that the RSC complex can regulate transcription elongation, such as helping RNAPII overcome the barrier of nucleosomes. Normally, if RNAPII is affected at the elongation stage, the metagene profiles of NET-seq would show some changes (Fischl et al. 2017; Churchman & Weissman 2011). In this study, the YMC was sampled at 11 time points (Figure S1a). To enable the quantitative comparison between samples, the same batch of *S.pombe* cells was added to each of the samples at the same ratio. NET-seq libraries were prepared from the 11 *S.pombe* spiked samples. It is worth to note here that the yeast NET-seq libraries were prepared before the ultra-low background NET-seq was developed, so these NET-seq libraries may contain 10-30% non-specific background from mature RNA.

The metagene profiles of *S.cerevisiae* reads for all the samples are plotted in Figure S1b. To control for the technical variations, the metagene profiles of *S.pombe* reads for all the samples are plotted in Figure S1c. Generally, at the global level, no obvious differences in the distribution of RNAPII can be seen on the metagene plot at neither the 5' nor the 3' ends of genes. The small differences in the metagene plot are mainly due to the technical variations. At the 5' end, the distribution of RNAPII is biased towards the transcription start sites. Several peaks can be seen in the first 600bps. They seem to have the same spacing as the nucleosomes. Some samples have slight variations in the heights of peaks. However, the same differences can be seen in the *S.pombe* control as well, which means they are technical variations rather

than real differences in the distribution of RNAPII. At the 3' end, the only obvious variation is in the region just before the PAS. However, like the 5' end, it can also be seen in the *S.pombe* control, so it is probably a technical variation as well. The heat map (Figure S1d) also agrees with the observations on the metagene. To sum up, at least at a global level, there are no obvious changes in the distribution of RNAPII on genes, which does not support the RSC complex regulating the global transcription elongation during the YMC.

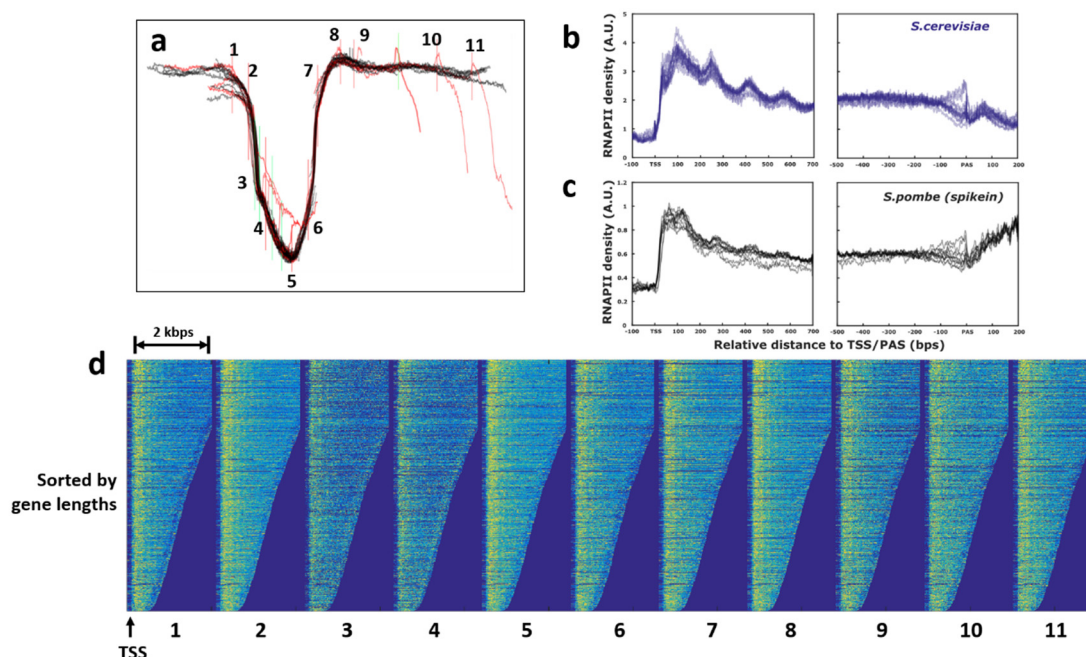


Figure S1 The distributions of RNAPII on genes were examined by NET-seq across the YMC. **a.** The YMC was sampled at 11 time points. The NET-seq and RNA-seq libraries were constructed from each sample. **b.** The metagene profiles based on the *S.cerevisiae* reads of NET-seq. The metagenes of the 11 NET-seq libraries were overlaid. **c.** The metagene profiles based on the *S.pombe* reads of NET-seq. The metagenes of the 11 NET-seq libraries were overlaid. **d.** A heatmap showing the distribution of RNAPII on each *S.cerevisiae* gene.

S2 The levels of transcripts are dominantly regulated by transcription

Section 1.9 suggests that the YMC may also be utilized to study the relationship between transcription and transcripts. In this study, a preliminary experiment was performed on this research topic by using NET-seq and RNA-seq to check the levels of transcription and transcripts during the YMC. It is worth noting here that this experiment was performed before

the development of the ultra-low background protocol of NET-seq, so the NET-seq reads may still contain a significant amount of signals from mature RNA.

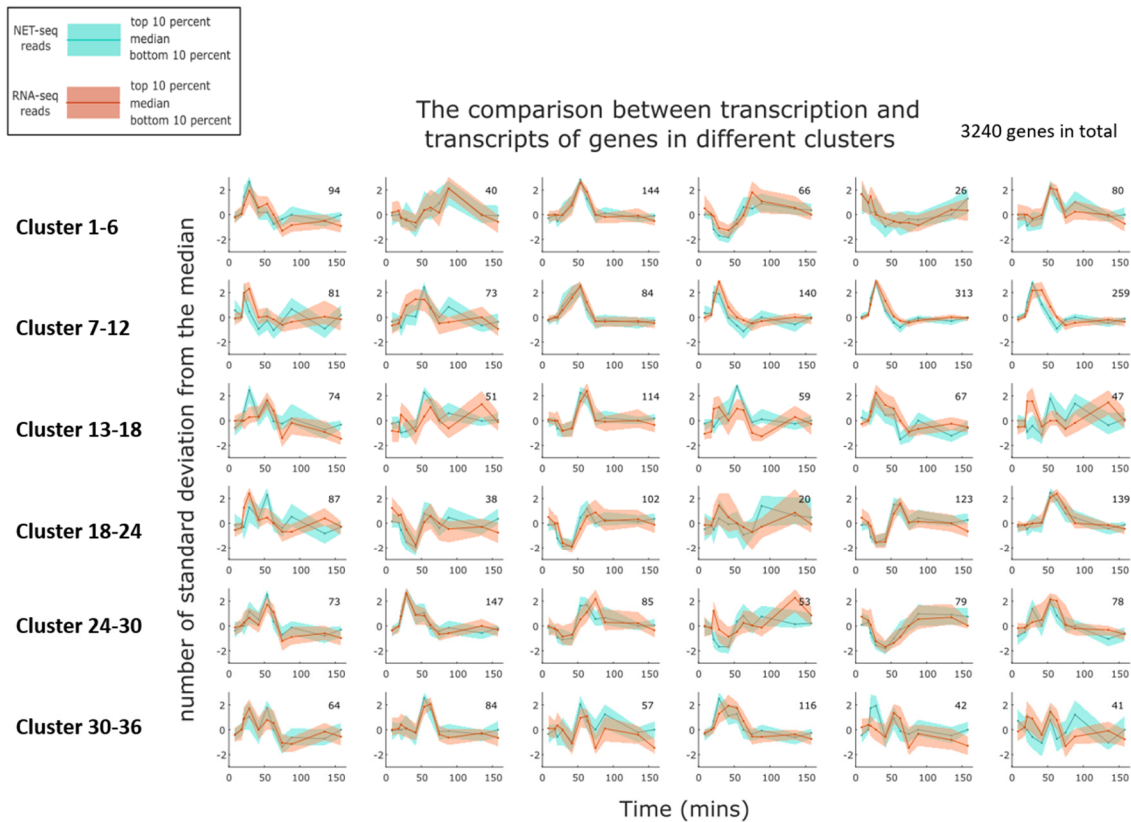


Figure S2 Comparing the temporal profiles of transcription and transcripts during the YMC. The number on the upper right corner of each small plot shows the number of genes in each cluster.

Specifically, the number of NET-seq reads in the first 400bps of each gene was used as the metric of the transcription activity for the gene. To assess the transcript levels at each time points, RNA-seq libraries were prepared from the same samples as the NET-seq libraries. Hence, both the activity of transcription and the transcript level could be measured for each gene at each of the 11 sampling points. In other words, each gene had a temporal profile for both transcription activity and transcripts levels across the YMC. Following that, the K-means clustering algorithm categorized all the 3240 genes into 36 small groups by taking into account their temporal profiles of both transcription and transcripts. Each group contained about 100 genes. In this analysis, in order to minimize the chance of missing certain patterns of

transcription and transcript, a relatively large number of groups (36 groups) was chosen to perform the k-means clustering algorithm.

The characteristic profiles for transcription activities as well as the transcript levels are plotted for each cluster in Figure S2. Generally, in most of the clusters, the changes in transcription activity correlate with the changes in transcript levels. In many cases, a peak or a trough of transcripts can be found right after the peak or trough of transcription. A good example is cluster 12. Its peak of transcription is followed by a peak of transcripts with a small time delay. It suggests that transcription plays a dominant role in regulating the levels of transcripts in most clusters. However, there are some exceptions. In cluster 8, the increase in transcript levels happens earlier than the increase in transcription. It suggests that the initial increase in the transcript levels may be caused by mechanisms other than transcription. One candidate could be the regulated reduction in the degradation of transcripts in this cluster. Conversely, in cluster 13, 19 and 35, the peak of transcription is not accompanied by a delayed peak of transcripts. One possibility is that the nascent RNA might be degraded cotranscriptionally through the NNS pathway, before turning into mRNA.

To summarize, generally, the changes in transcript levels are primarily caused by the changes in transcription. However, for certain genes, the temporal profiles of transcript levels can be modulated by other mechanisms.

S3 Genes with introns have a slightly longer time delay between transcription and transcripts than genes without introns

The previous section has shown that genes in cluster 12 have a peak of transcript lagging behind the peak of transcription. Actually, most intron containing genes of *S.cerevisiae* are found in this cluster (Figure S3a). If genes with introns are separated from genes without intron, cluster 12 can be further divided into two subclusters, cluster 12A (with introns) and 12B (without intron). The temporal profiles of transcription and transcripts are plotted for

both subclusters (Figure S3b). The time delay between transcription and transcripts for genes in cluster 12A seems to be slightly longer than that for genes in cluster 12B.

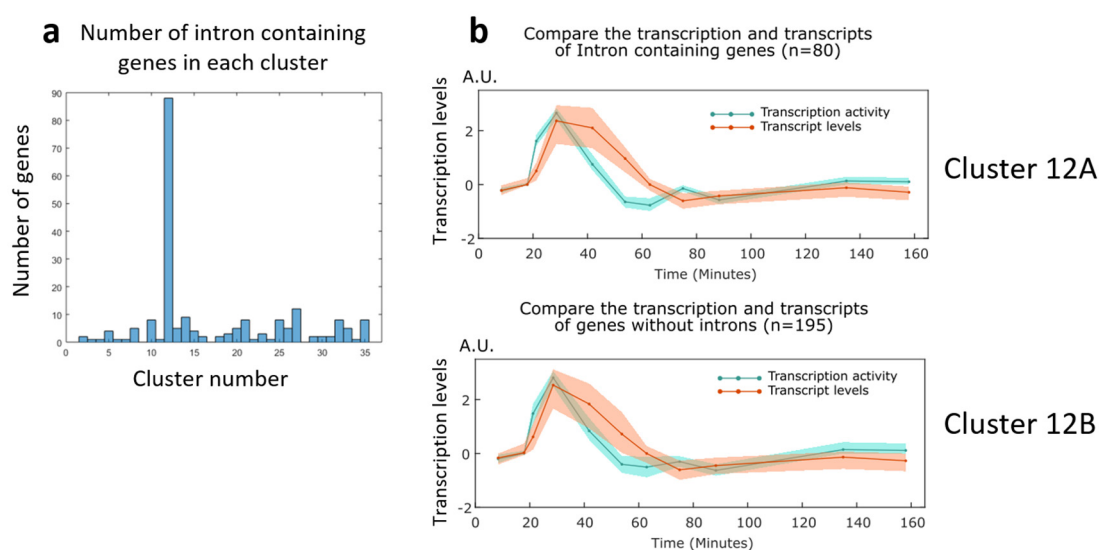


Figure S3 a. The number of intron-containing genes in each of the 36 clusters of figure S2. The x-axis is the cluster number from 1 to 36. The y-axis is the number of intron-containing genes. Most intron containing genes are in cluster 12. **b. Comparing the temporal profiles of transcription and transcripts in cluster 12A and 12B.** Genes in cluster 12A contain introns, while genes in cluster 12B do not have introns.

To examine the time delay in a slightly more quantitative way, the peaking time points of transcription and transcripts were estimated from the discrete data points by using smoothing spline algorithm. The workflow is illustrated in Figure S4. Specifically, for each gene in the two subclusters, the 11 discrete data points of both NET-seq and RNA-seq were fitted with smooth spline curves respectively, which produced a smooth curve for transcription activity and transcript levels respectively. The peak calling algorithm was then applied to find the local maximal point for each of the smooth curves. Finally, the exact time delay between transcription and transcripts can be estimated for each gene.

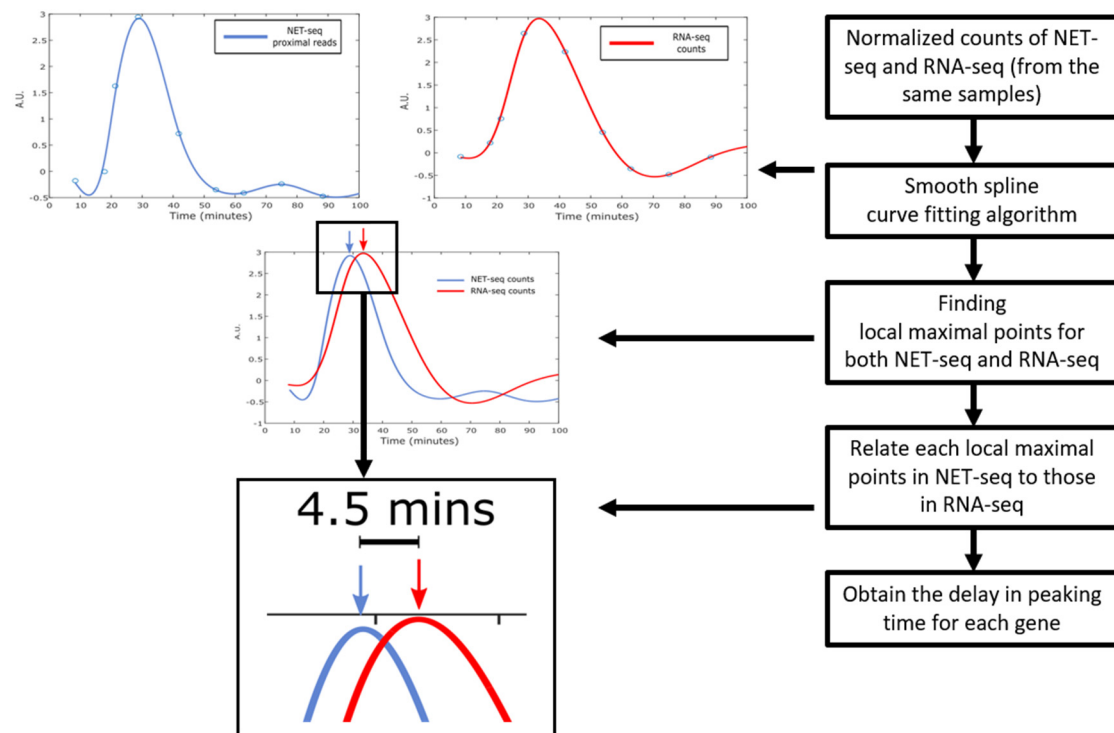
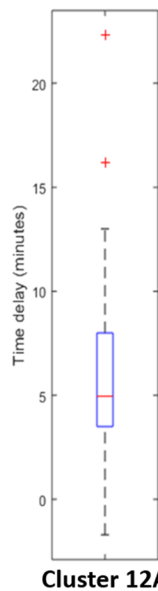


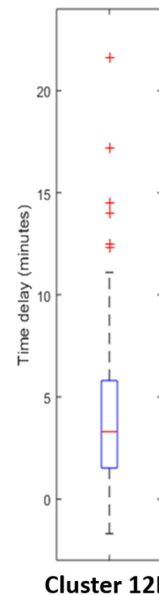
Figure S4 The workflow for estimating the time delay between the peak in the transcription profile and transcripts profile for each gene.

Distribution of time delay (Cluster 12A, n=86)



Median delay and standard deviation
 4.95 ± 0.53 mins
 (my method)

Distribution of time delay (Cluster 12B, n=195)



Median delay and standard deviation
 3.30 ± 0.32 mins
 (my method)

Figure S5 The distributions of the time delay between transcription and transcripts for cluster 12A and cluster 12B.

The median time delay for cluster 12A and 12B are 4.95 ± 0.53 mins and 3.30 ± 0.32 mins respectively (Figure S5). Intron-containing genes have a longer time delay between transcription and transcripts than genes without introns. Notably, the time delay for cluster 12A agrees with the measurement in the literature. Instead of using NET-seq, Kuang et al used RT-qPCR to measure the time delay between unspliced and spliced transcripts. They have reported that the time delay for ribosomal protein genes (which contain introns) is 5mins (Kuang et al. 2014).

A kinetic model in the literature has shown that the differential time delay can be caused by two factors (Honkela et al. 2015). The first is the processing time of intron splicing. If the genes do not have introns, the nascent RNA of these genes may take less time to become mRNA. The second factor is the RNA degradation rate. If the genes have higher RNA turnover rates, the peak of the transcript will be closer to the peak of transcription activity. Currently, it is hard to tell which factor causes the differential time delay between genes in cluster 12A and 12B.

Figure S6 The temporal changes in the levels of transcription, transcripts and histone acetylation of central metabolic genes during the YMC. The levels of transcription for the wild-type (blue curve) were assessed by NET-seq. The amounts of transcripts for the wild-type (blue curve) and the *rsc2Δ* mutant (red curve) were measured by RNA-seq. The transcript levels for the wild-type and the *rsc2Δ* mutant are comparable, because of the normalization by RNA spike-in. The signals of NET-seq and RNA-seq are expressed as the number of reads (the y-axis on the left). The ChIP-seq signals for histone acetylation (Kuang et al. 2014) are expressed as the number of standard deviations from the median (the y-axis on the right).

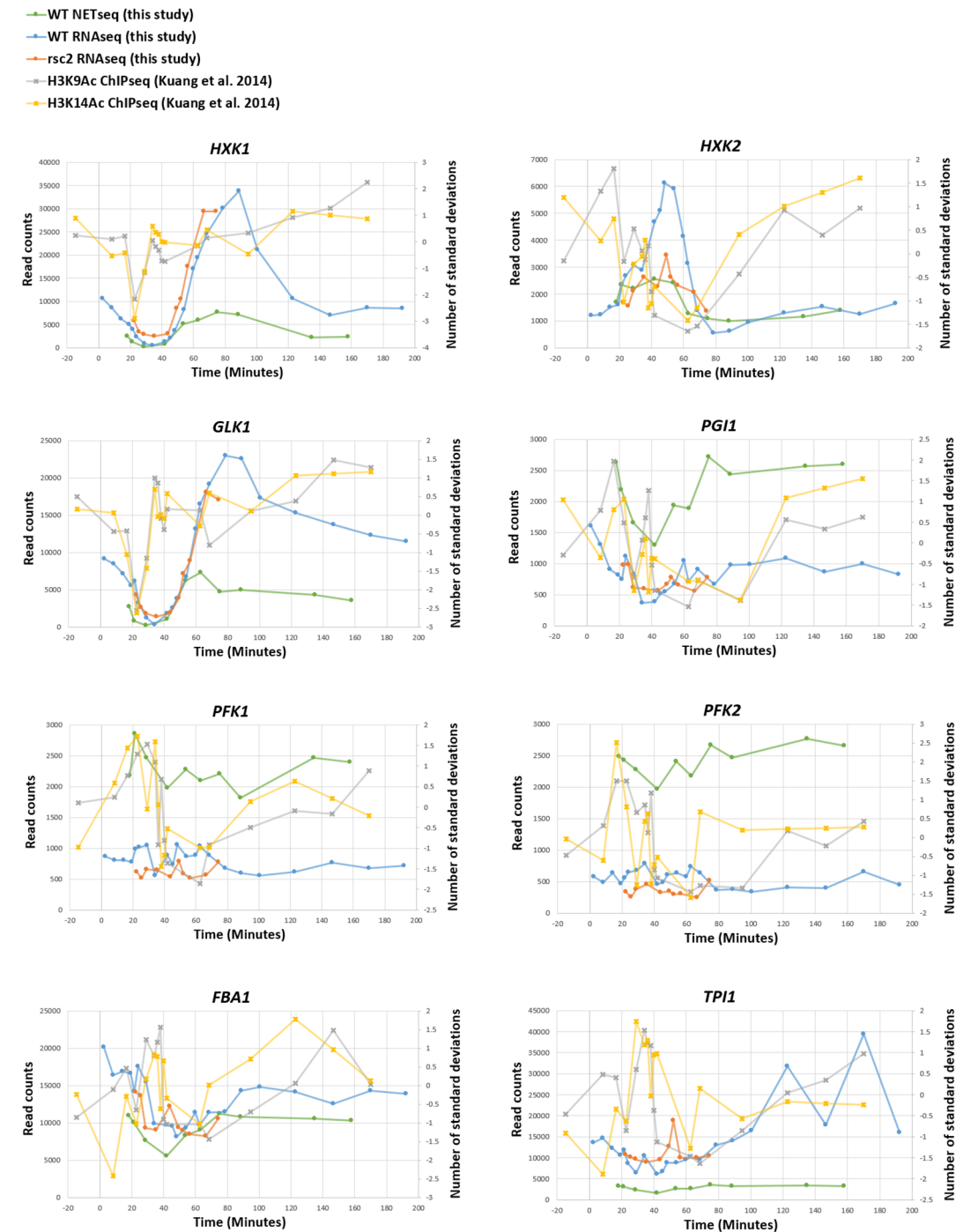


Figure S6 continued

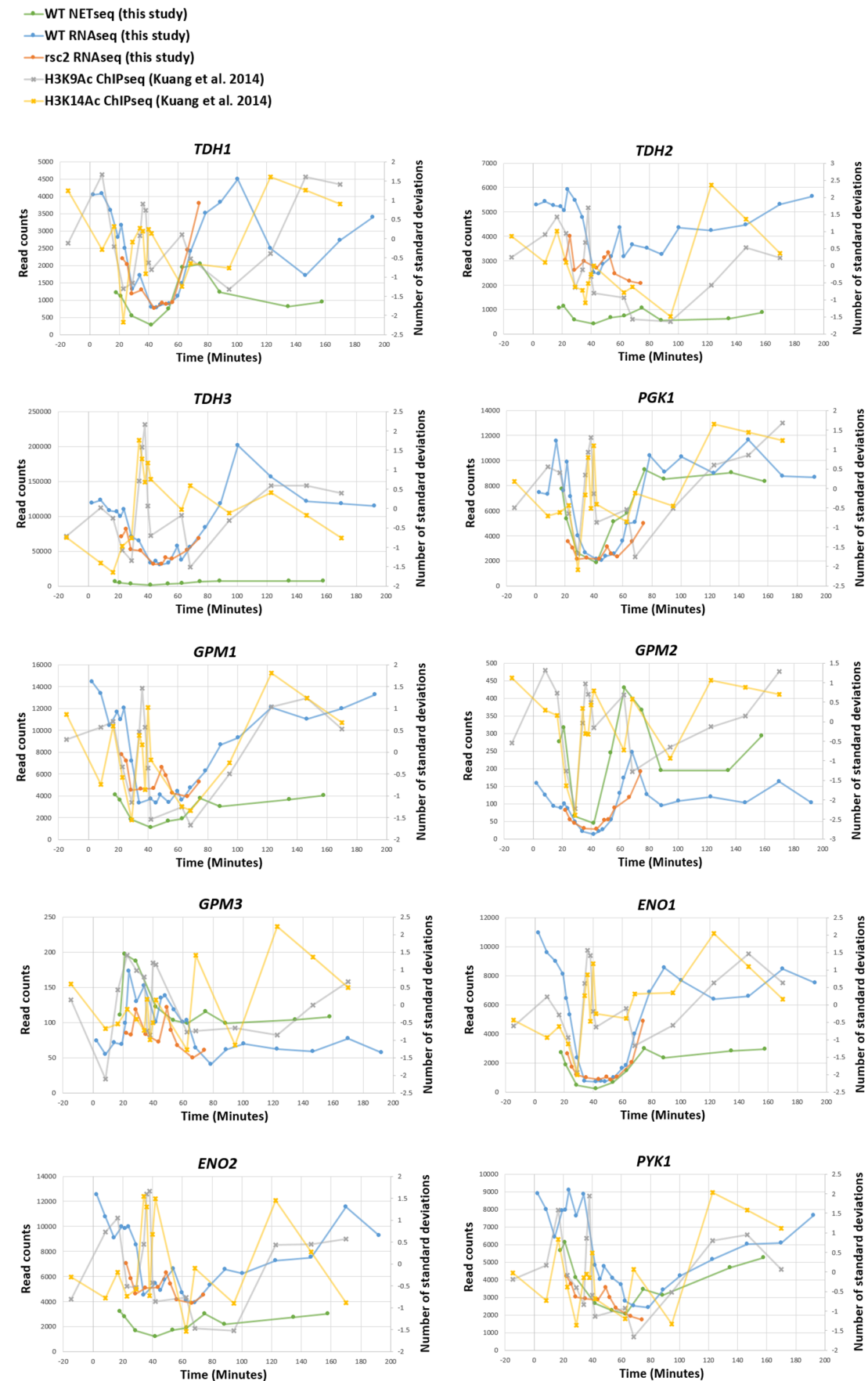


Figure S6 continued

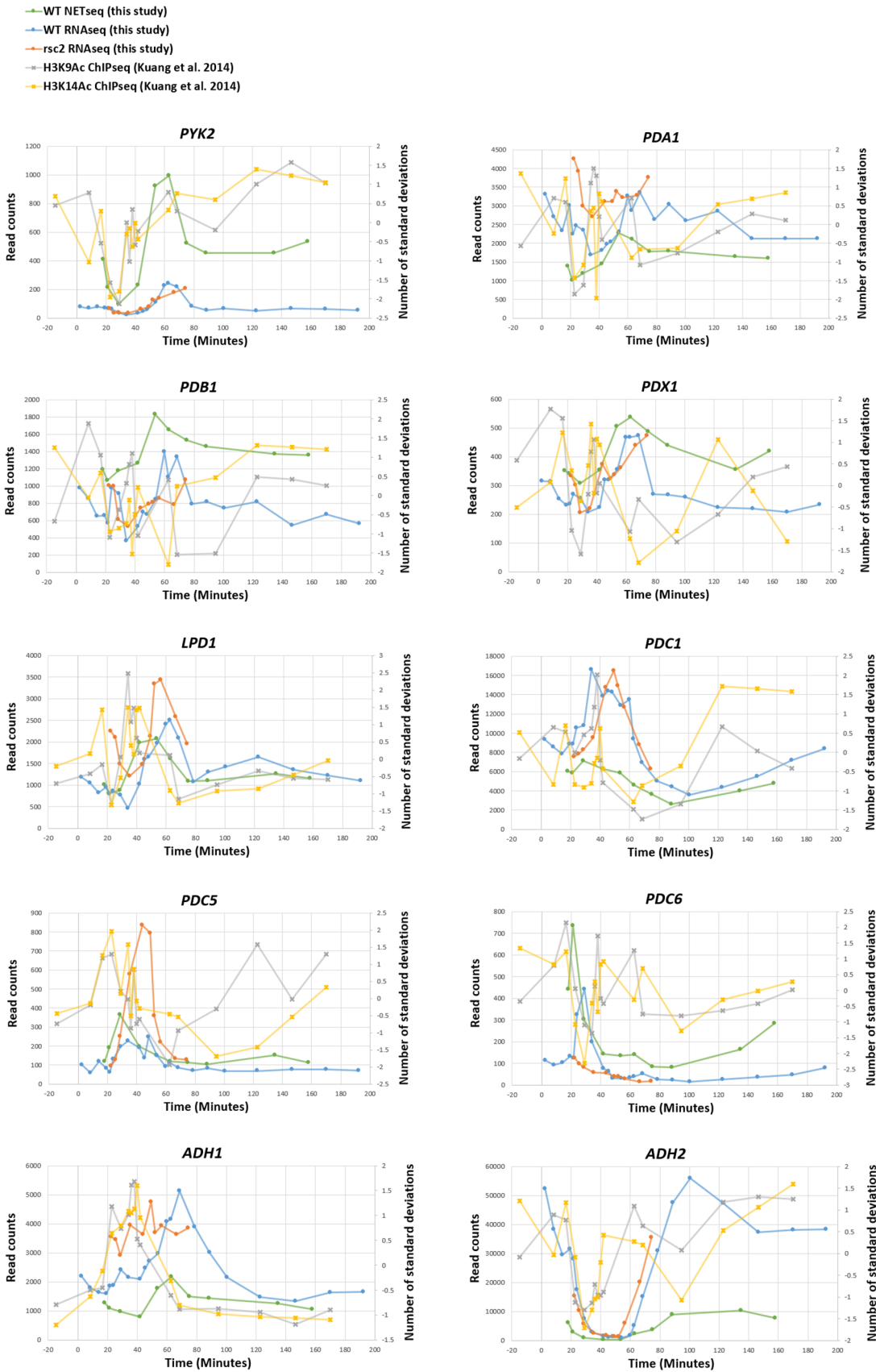


Figure S6 continued

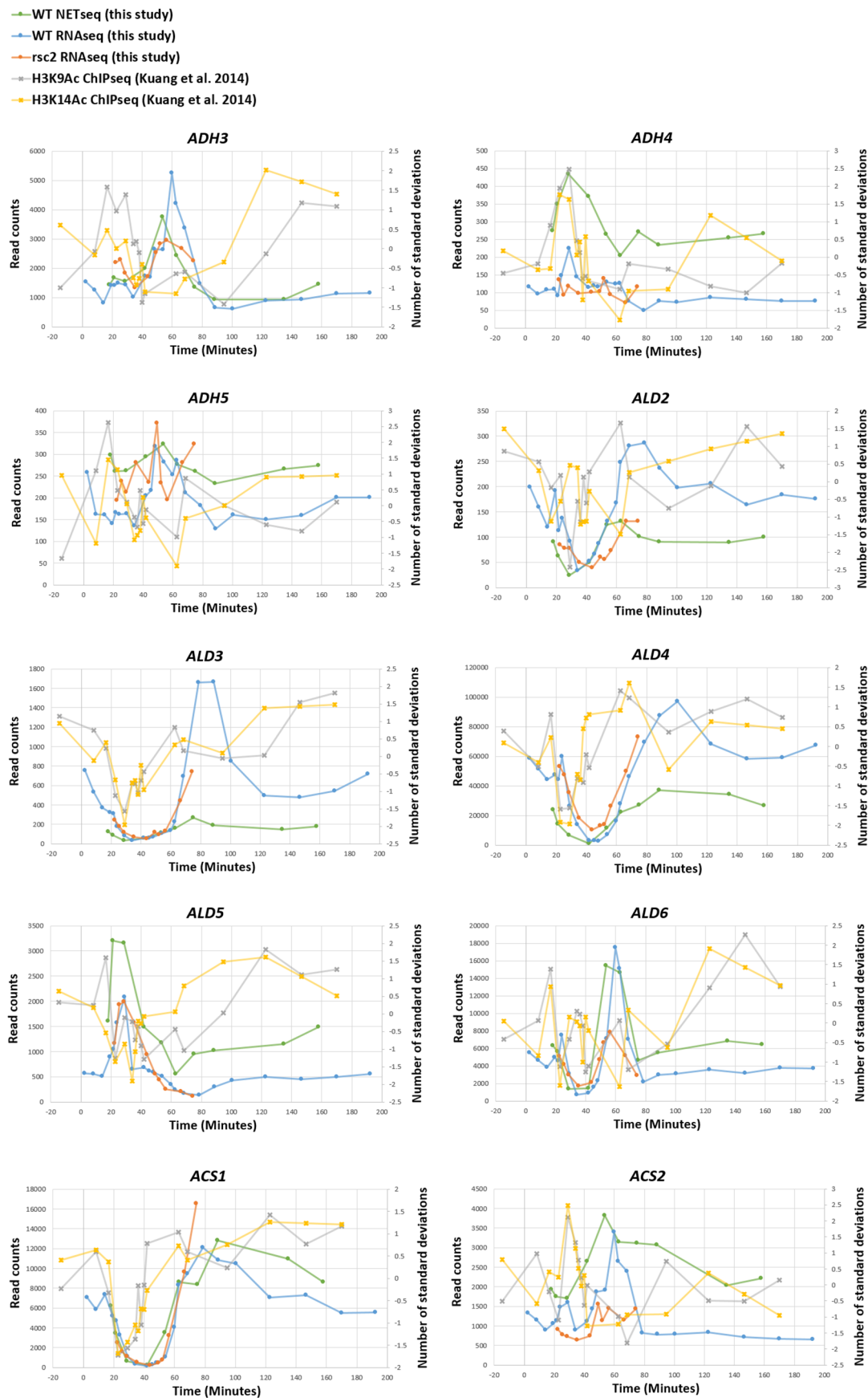


Figure S6 continued

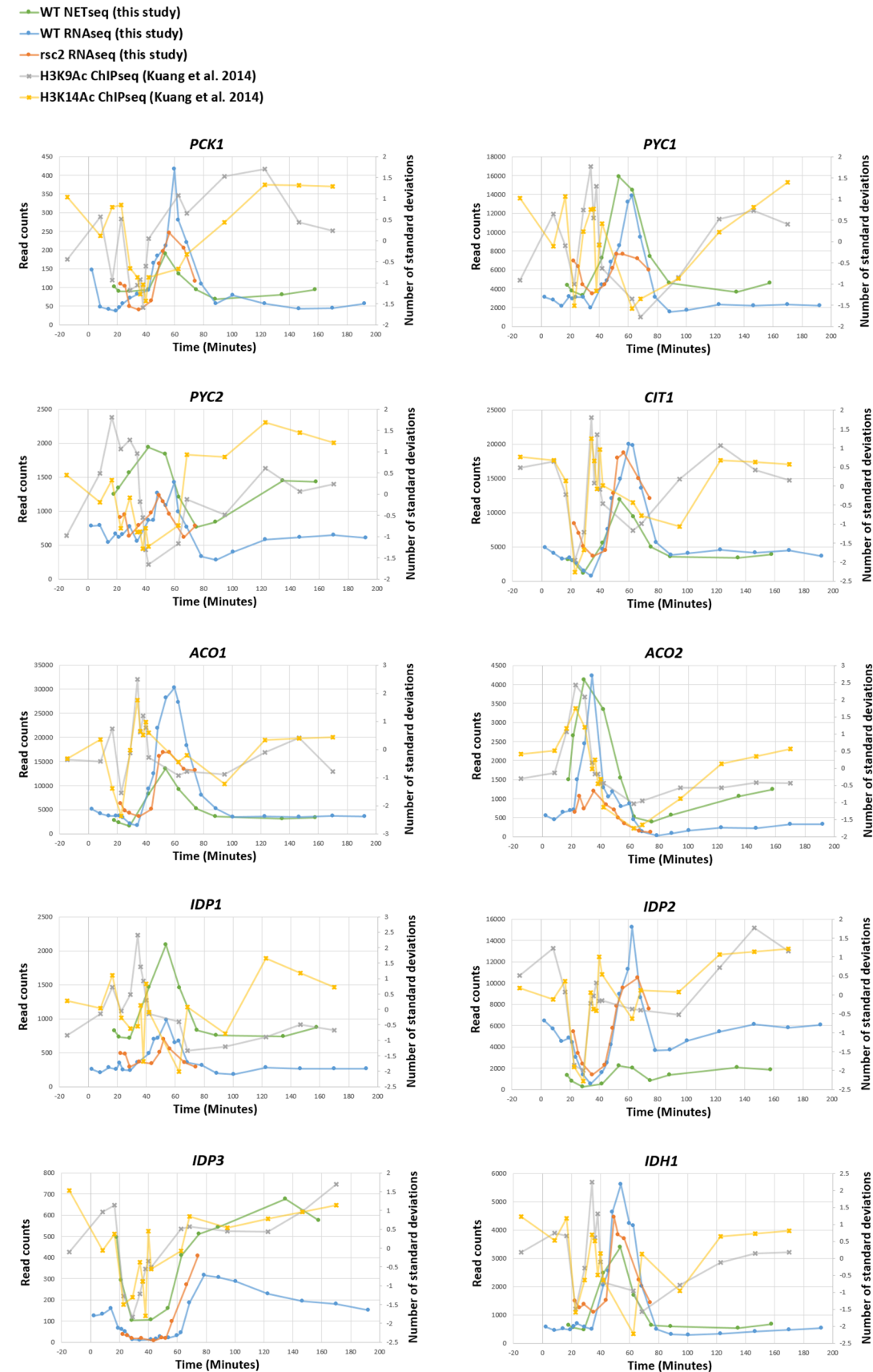


Figure S6 continued

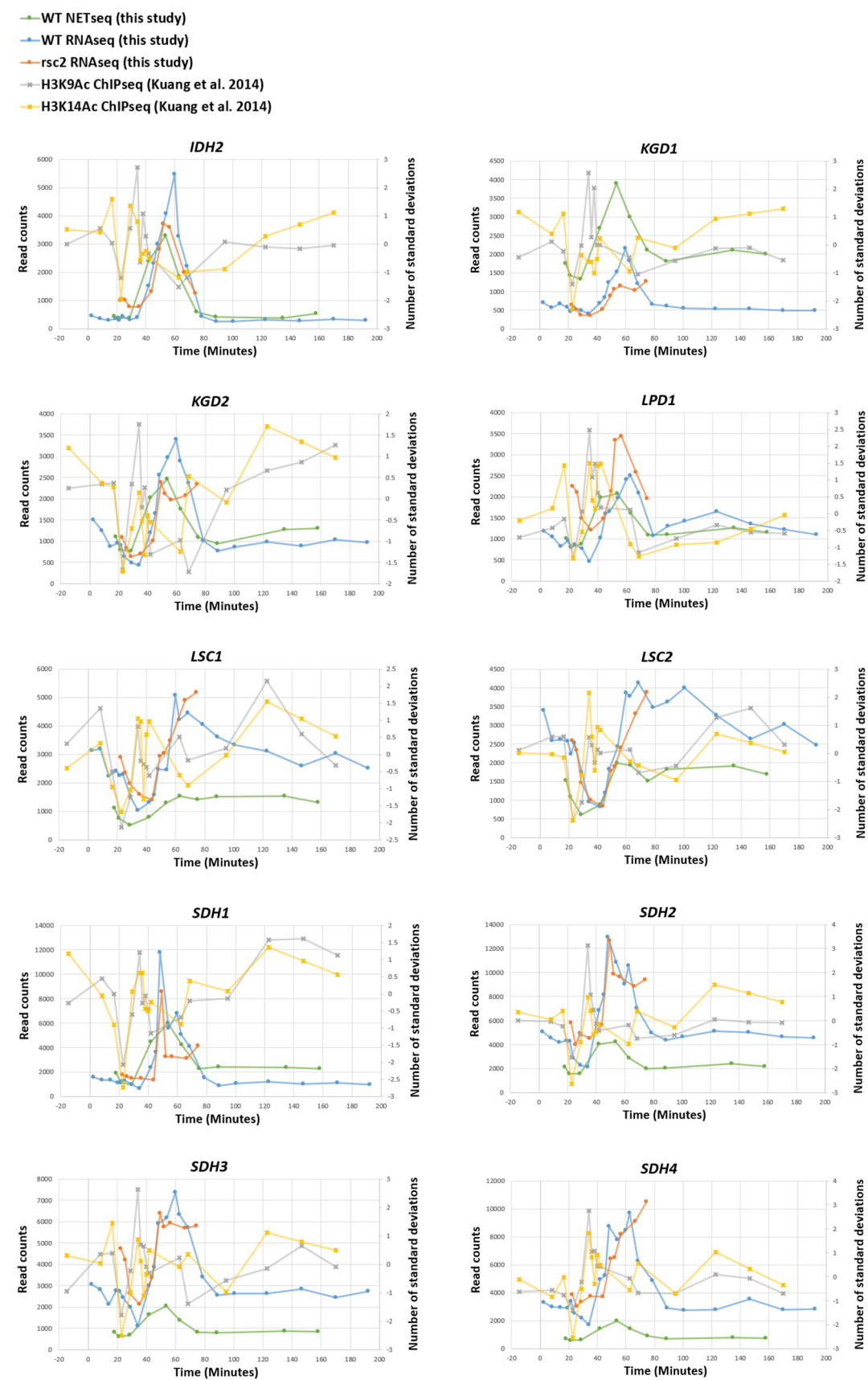
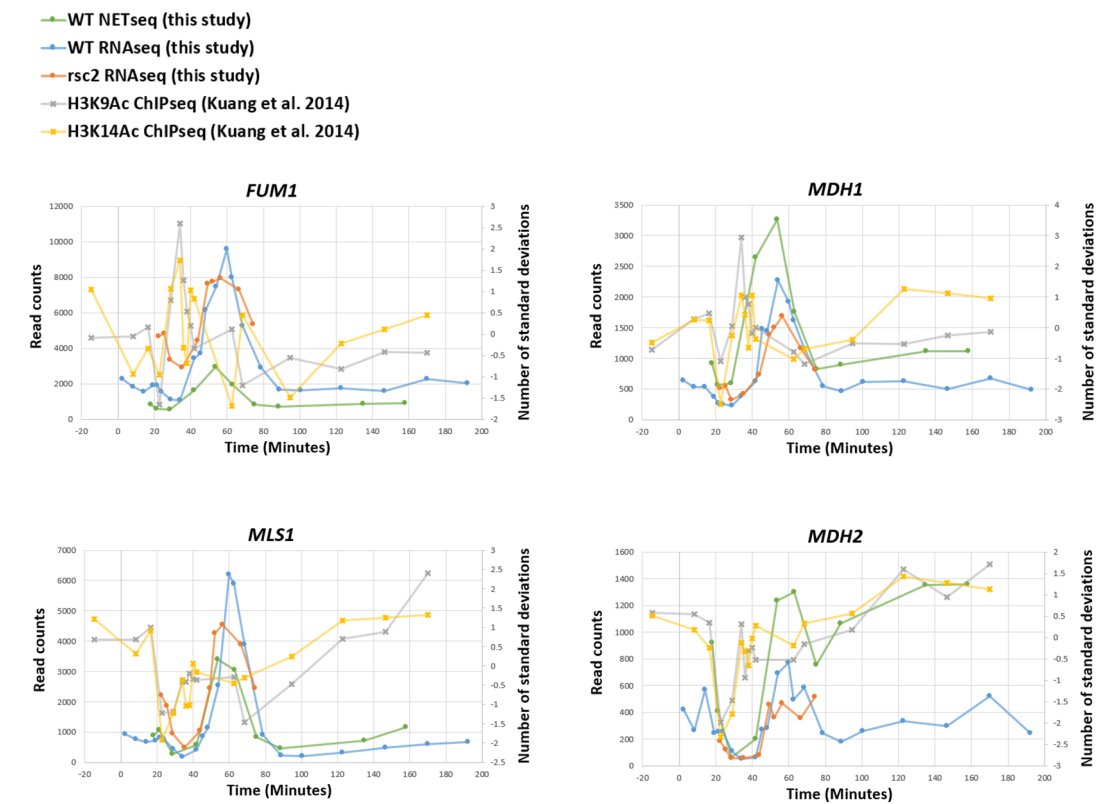


Figure S6 continued



References

- Adam, M. et al., 2001. H2A.Z is required for global chromatin integrity and for recruitment of RNA polymerase II under specific conditions. *Molecular and cellular biology*, 21(18), pp.6270–9. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/11509669> [Accessed March 20, 2019].
- Adiconis, X. et al., 2018. Comprehensive comparative analysis of 5'-end RNA-sequencing methods. *Nature Methods*, 15(7), pp.505–511. Available at: <http://www.nature.com/articles/s41592-018-0014-2> [Accessed March 22, 2019].
- Alén, C. et al., 2002. A role for chromatin remodeling in transcriptional termination by RNA polymerase II. *Molecular cell*, 10(6), pp.1441–52. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/12504018> [Accessed March 20, 2019].
- Allard, S. et al., 1999. NuA4, an essential transcription adaptor/histone H4 acetyltransferase complex containing Esa1p and the ATM-related cofactor Tra1p. *The EMBO Journal*, 18(18), pp.5108–5119. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/10487762> [Accessed March 19, 2019].
- Allfrey, V.G., Faulkner, R. & Mirsky, A.E., 1964. Acetylation and methylation of histones and their possible role in the regulation of RNA synthesis. *Proceedings of the National Academy of Sciences of the United States of America*, 51, pp.786–94. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/14172992> [Accessed March 18, 2019].
- Armstrong, J.A. et al., 2002. The Drosophila BRM complex facilitates global transcription by RNA polymerase II. *The EMBO journal*, 21(19), pp.5245–54. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/12356740> [Accessed March 20, 2019].
- Arnesen, T. et al., 2009. Proteomics analyses reveal the evolutionary conservation and divergence of N-terminal acetyltransferases from yeast and humans. *Proceedings of the National Academy of Sciences*, 106(20), pp.8157–8162. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/19420222> [Accessed March 18, 2019].
- Arribere, J.A. & Gilbert, W. V., 2013. Roles for transcript leaders in translation and mRNA decay revealed by transcript leader sequencing. *Genome Research*, 23(6), pp.977–987. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/23580730> [Accessed March 22, 2019].
- Avalos, J.L., Bever, K.M. & Wolberger, C., 2005. Mechanism of sirtuin inhibition by nicotinamide: altering the NAD(+) cosubstrate specificity of a Sir2 enzyme. *Molecular cell*, 17(6), pp.855–68. Available at: <https://linkinghub.elsevier.com/retrieve/pii/S1097276505011214> [Accessed March 18, 2019].
- Awad, S. & Hassan, A.H., 2008a. The Swi2/Snf2 bromodomain is important for the full binding and remodeling activity of the SWI/SNF complex on H3- and H4-acetylated nucleosomes. *Annals of the New York Academy of Sciences*, 1138, pp.366–375.
- Awad, S. & Hassan, A.H., 2008b. The Swi2/Snf2 Bromodomain Is Important for the Full Binding and Remodeling Activity of the SWI/SNF Complex on H3- and H4-acetylated Nucleosomes. *Annals of the New York Academy of Sciences*, 1138(1), pp.366–375. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/18837912> [Accessed March 20, 2019].
- Bai, L., Ondracka, A. & Cross, F.R., 2011. Multiple sequence-specific factors generate the

- nucleosome-depleted region on CLN2 promoter. *Molecular cell*, 42(4), pp.465–76. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/21596311> [Accessed March 22, 2019].
- Bailey, T.L. et al., 2009. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Research*, 37(Web Server), pp.W202–W208. Available at: <https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkp335> [Accessed March 22, 2019].
- Balasubramanian, R. et al., 2002. Role of the Ada2 and Ada3 Transcriptional Coactivators in Histone Acetylation. *Journal of Biological Chemistry*, 277(10), pp.7989–7995. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/11773077> [Accessed March 19, 2019].
- Barasch, D. et al., 2015. Glycolysis-Mediated Changes in Acetyl-CoA and Histone Acetylation Control the Early Differentiation of Embryonic Stem Cells. *Cell Metabolism*, 21(3), pp.392–402.
- Barber, M.F. et al., 2012. SIRT7 links H3K18 deacetylation to maintenance of oncogenic transformation. *Nature*, 487(7405), pp.114–118. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/22722849> [Accessed March 19, 2019].
- van den Berg, M.A. et al., 1996. The two acetyl-coenzyme A synthetases of *Saccharomyces cerevisiae* differ with respect to kinetic properties and transcriptional regulation. *The Journal of biological chemistry*, 271(46), pp.28953–9. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/8910545> [Accessed March 18, 2019].
- Berndsen, C.E. & Denu, J.M., 2008. Catalysis and substrate selection by histone/protein lysine acetyltransferases. *Current Opinion in Structural Biology*, 18(6), pp.682–689. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/19056256> [Accessed March 19, 2019].
- Blank, H.M. et al., 2009. Sulfur metabolism actively promotes initiation of cell division in yeast. *PLoS one*, 4(11), p.e8018. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/19956742> [Accessed March 20, 2019].
- Booth, G.T. et al., 2016. Divergence of a conserved elongation factor and transcription regulation in budding and fission yeast. , pp.1–13.
- Bosio, M.C. et al., 2017. Abf1 and other general regulatory factors control ribosome biogenesis gene expression in budding yeast. *Nucleic Acids Research*, 45(8), pp.4493–4506. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/28158860> [Accessed March 22, 2019].
- Boulton, C. & Quain, D., 2001. *Brewing yeast and fermentation*, Blackwell Science. Available at: <https://books.google.co.uk/books?id=L3SU4ahY-m4C&pg=PA85&dq=TCA+cycle+overflow+ethanol&hl=en&sa=X&ved=0ahUKEwjQidK2g5bhAhVCrHEKHZUyBvEQ6AEISTAF#v=onepage&q=TCA cycle overflow ethanol&f=false> [Accessed March 22, 2019].
- Boyer, L.A., Latek, R.R. & Peterson, C.L., 2004. The SANT domain: a unique histone-tail-binding module? *Nature Reviews Molecular Cell Biology*, 5(2), pp.158–163. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/15040448> [Accessed March 20, 2019].
- Bradley, P.H. et al., 2019. Minor Isozymes Tailor Yeast Metabolism to Carbon Availability L. A. McCue, ed. *mSystems*, 4(1), pp.e00170-18. Available at: <http://msystems.asm.org/lookup/doi/10.1128/mSystems.00170-18> [Accessed March

21, 2019].

- Bravo-San Pedro, J.M. et al., 2015. Acetyl Coenzyme A: A Central Metabolite and Second Messenger. *Cell Metabolism*, 21(6), pp.805–821.
- Brown, C.E. et al., 2001. Recruitment of HAT Complexes by Direct Activator Interactions with the ATM-Related Tra1 Subunit. *Science*, 292(5525), pp.2333–2337. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/11423663> [Accessed March 19, 2019].
- Brownlee, P.M. et al., 2012. Cancer and the bromodomains of BAF180. *Biochemical Society Transactions*, 40(2), pp.364–369.
- Bryant, G.O. et al., 2008. Activator Control of Nucleosome Occupancy in Activation and Repression of Transcription O. J. Rando, ed. *PLoS Biology*, 6(12), p.e317. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/19108605> [Accessed March 20, 2019].
- Bungard, D., Reed, M. & Winter, E., 2004. RSC1 and RSC2 Are Required for Expression of Mid-Late Sporulation-Specific Genes in *Saccharomyces cerevisiae*. *Eukaryotic Cell*, 3(4), pp.910–918.
- Burgio, G. et al., 2008. Genetic Identification of a Network of Factors that Functionally Interact with the Nucleosome Remodeling ATPase ISWI A. Akhtar, ed. *PLoS Genetics*, 4(6), p.e1000089. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/18535655> [Accessed March 20, 2019].
- Burnetti, A.J., Aydin, M. & Buchler, N.E., 2016. Cell cycle Start is coupled to entry into the yeast metabolic cycle across diverse strains and growth rates. *Molecular biology of the cell*, 27(1), pp.64–74. Available at: <http://www.molbiolcell.org/content/27/1/64.full>.
- Cai, L. et al., 2011. Acetyl-CoA Induces Cell Growth and Proliferation by Promoting the Acetylation of Histones at Growth Genes. *Molecular Cell*, 42(4), pp.426–437. Available at: <http://linkinghub.elsevier.com/retrieve/pii/S1097276511003327> [Accessed February 20, 2017].
- Cairns, B.R. et al., 1994. A multisubunit complex containing the SWI1/ADR6, SWI2/SNF2, SWI3, SNF5, and SNF6 gene products isolated from yeast. *Proceedings of the National Academy of Sciences of the United States of America*, 91(5), pp.1950–4. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/8127913> [Accessed March 20, 2019].
- Cairns, B.R., 1998. Chromatin remodeling machines: similar motors, ulterior motives. *Trends in biochemical sciences*, 23(1), pp.20–5. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/9478131> [Accessed March 20, 2019].
- Cairns, B.R. et al., 1996. RSC, an essential, abundant chromatin-remodeling complex. *Cell*, 87(7), pp.1249–60. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/8980231> [Accessed March 20, 2019].
- Cairns, B.R., Schlichter, A., et al., 1999. Two functionally distinct forms of the RSC nucleosome-remodeling complex, containing essential AT hook, BAH, and bromodomains. *Molecular cell*, 4(5), pp.715–23. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/10619019> [Accessed March 21, 2017].
- Cairns, B.R., Schlichter, A., et al., 1999. Two Functionally Distinct Forms of the RSC Nucleosome-Remodeling Complex, Containing Essential AT Hook, BAH, and Bromodomains. *Molecular Cell*, 4, pp.715–723.
- Cambronne, X.A. et al., 2016. Biosensor reveals multiple sources for mitochondrial NAD⁺.

- Science*, 352(6292), pp.1474–1477. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/27313049> [Accessed March 19, 2019].
- Candelli, T. et al., 2018. High-resolution transcription maps reveal the widespread impact of roadblock termination in yeast. *The EMBO Journal*, 37(4), p.e97490. Available at: <http://emboj.embopress.org/lookup/doi/10.15252/emboj.201797490> [Accessed March 22, 2019].
- Cantó, C., Menzies, K.J. & Auwerx, J., 2015. NAD⁺ Metabolism and the Control of Energy Homeostasis: A Balancing Act between Mitochondria and the Nucleus. *Cell Metabolism*, 22(1), pp.31–53. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/26118927> [Accessed March 18, 2019].
- Carey, M., Li, B. & Workman, J.L., 2006. RSC exploits histone acetylation to abrogate the nucleosomal block to RNA polymerase II elongation. *Molecular cell*, 24(3), pp.481–7. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/17081996> [Accessed March 20, 2019].
- Carmen, A.A. et al., 2002. HDA2 and HDA3 are related proteins that interact with and are essential for the activity of the yeast histone deacetylase HDA1. *Proceedings of the National Academy of Sciences*, 98(8), pp.4391–4396.
- Carr, A.-J.F. & Whitmore, D., 2005. Imaging of single light-responsive clock cells reveals fluctuating free-running periods. *Nature Cell Biology*, 7(3), pp.319–321. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/15738976> [Accessed March 22, 2019].
- Carrozza, M.J., Li, B., et al., 2005. Histone H3 Methylation by Set2 Directs Deacetylation of Coding Regions by Rpd3S to Suppress Spurious Intragenic Transcription. *Cell*, 123(4), pp.581–592. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/16286007> [Accessed March 19, 2019].
- Carrozza, M.J., Florens, L., et al., 2005. Stable incorporation of sequence specific repressors Ash1 and Ume6 into the Rpd3L complex. *Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression*, 1731(2), pp.77–87. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/16314178> [Accessed March 19, 2019].
- Cerevisiae Graphical, S. et al., 2016. Cyclin-Dependent Kinase Co-Ordinates Carbohydrate Metabolism and Cell Cycle in *S. cerevisiae*. *Molecular Cell*, 62, pp.546–557. Available at: <http://dx.doi.org/10.1016/j.molcel.2016.04.026> [Accessed March 22, 2019].
- Chakravarti, D. et al., 1996. Role of CBP/P300 in nuclear receptor signalling. *Nature*, 383(6595), pp.99–103. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/8779723> [Accessed March 19, 2019].
- Chambers, A.L. et al., 2013. The BAH domain of Rsc2 is a histone H3 binding domain. *Nucleic Acids Research*, 41(19), pp.9168–9182.
- Chambers, A.L. et al., 2012. The two different isoforms of the RSC chromatin remodeling complex play distinct roles in DNA damage responses. *PLoS ONE*, 7(2), pp.1–11.
- CHANCE, B., ESTABROOK, R.W. & GHOSH, A., 1964. DAMPED SINUSOIDAL OSCILLATIONS OF CYTOPLASMIC REDUCED PYRIDINE NUCLEOTIDE IN YEAST CELLS. *Proceedings of the National Academy of Sciences of the United States of America*, 51(6), pp.1244–51. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/14215650> [Accessed March 20, 2019].
- Chatterjee, N. et al., 2011. Histone H3 tail acetylation modulates ATP-dependent remodeling

- through multiple mechanisms. *Nucleic Acids Research*, 39(19), pp.8378–8391.
- Chen, Z. et al., 2007. Restriction of DNA Replication to the Reductive Phase of the Metabolic Cycle Protects Genome Integrity. *Science*, 316(5833), pp.1916–1919. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/17600220> [Accessed March 20, 2019].
- Chisolm, D.A. & Weinmann, A.S., 2018. Connections Between Metabolism and Epigenetics in Programming Cellular Differentiation. *Annual Review of Immunology*, 36(1), pp.221–246.
- Choudhary, C. et al., 2009. Lysine Acetylation Targets Protein Complexes and Co-Regulates Major Cellular Functions. *Science*, 325(5942), pp.834–840. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/19608861> [Accessed March 18, 2019].
- Churchman, L.S. & Weissman, J.S., 2011. Nascent transcript sequencing visualizes transcription at nucleotide resolution. *Nature*, 469(7330), pp.368–373. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/21248844> [Accessed March 20, 2019].
- Churchman, L.S. & Weissman, J.S., 2012. Native elongating transcript sequencing (NET-seq). *Current Protocols in Molecular Biology*, 1(SUPPL.98), pp.1–17.
- Clapier, C.R. et al., 2017. Mechanisms of action and regulation of ATP-dependent chromatin-remodelling complexes. *Nature Reviews Molecular Cell Biology*, 18(7), pp.407–422. Available at: <http://www.nature.com/doi/10.1038/nrm.2017.26> [Accessed March 20, 2019].
- Clapier, C.R. & Cairns, B.R., 2014. Chromatin Remodeling Complexes. In *Fundamentals of Chromatin*. New York, NY: Springer New York, pp. 69–146. Available at: http://link.springer.com/10.1007/978-1-4614-8624-4_3 [Accessed March 20, 2019].
- Clapier, C.R. & Cairns, B.R., 2012. Regulation of ISWI involves inhibitory modules antagonized by nucleosomal epitopes. *Nature*, 492(7428), pp.280–284. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/23143334> [Accessed March 20, 2019].
- Colin, J. et al., 2014. Article Roadblock Termination by Reb1p Restricts Cryptic and Readthrough Transcription. *Molecular Cell*, 56, pp.667–680. Available at: <http://dx.doi.org/10.1016/j.molcel.2014.10.026> [Accessed March 22, 2019].
- Core, L.J. et al., 2014. Analysis of nascent RNA identifies a unified architecture of initiation regions at mammalian promoters and enhancers. *Nature genetics*, 46(12), pp.1311–20. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/25383968> [Accessed March 22, 2019].
- Core, L.J., Waterfall, J.J. & Lis, J.T., 2008. Nascent RNA Sequencing Reveals Widespread Pausing and Divergent Initiation at Human Promoters. *Science*, 322(5909), pp.1845–1848. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/19056941> [Accessed March 20, 2019].
- Cosma, M.P., Tanaka, T. & Nasmyth, K., 1999. Ordered recruitment of transcription and chromatin remodeling factors to a cell cycle- and developmentally regulated promoter. *Cell*, 97(3), pp.299–311. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/10319811> [Accessed March 20, 2019].
- Côté, J. et al., 1994. Stimulation of GAL4 derivative binding to nucleosomal DNA by the yeast SWI/SNF complex. *Science (New York, N.Y.)*, 265(5168), pp.53–60. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/8016655> [Accessed March 20, 2019].

- Cvetesic, N. et al., 2018. SLIC-CAGE: high-resolution transcription start site mapping using nanogram-levels of total RNA. *Genome research*, 28(12), pp.1943–1956. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/30404778> [Accessed March 22, 2019].
- Dang, L. et al., 2009. Cancer-associated IDH1 mutations produce 2-hydroxyglutarate. *Nature*, 462(7274), pp.739–744. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/19935646> [Accessed March 18, 2019].
- Dang, W., 2014. The controversial world of sirtuins. *Drug Discovery Today: Technologies*, 12, pp.e9–e17. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/25027380> [Accessed March 19, 2019].
- Dang, W. & Bartholomew, B., 2007. Domain Architecture of the Catalytic Subunit in the ISW2-Nucleosome Complex. *Molecular and Cellular Biology*, 27(23), pp.8306–8317. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/17908792> [Accessed March 20, 2019].
- De Deken, R.H., 1966. *The Crabtree Effect: A Regulatory System in Yeast*, Available at: www.microbiologyresearch.org [Accessed March 22, 2019].
- DeLaForest, A. et al., 2011. HNF4A is essential for specification of hepatic progenitors from human pluripotent stem cells. *Development (Cambridge, England)*, 138(19), pp.4143–53. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/21852396> [Accessed March 18, 2019].
- Dhalluin, C. et al., 1999. Structure and ligand of a histone acetyltransferase bromodomain. *Nature*, 399(6735), pp.491–496. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/10365964> [Accessed March 20, 2019].
- Dhillon, S., 2018. Ivosidenib: First Global Approval. *Drugs*, 78(14), pp.1509–1516. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/30209701> [Accessed March 18, 2019].
- van Dijk, E.L. et al., 2011. XUTs are a class of Xrn1-sensitive antisense regulatory non-coding RNA in yeast. *Nature*, 475(7354), pp.114–117. Available at: <http://www.nature.com/articles/nature10118> [Accessed March 22, 2019].
- Dimitrova, E., Turberfield, A.H. & Klose, R.J., 2015. Histone demethylases in chromatin biology and beyond. *EMBO Reports*, 16(12), pp.1620–1639. Available at: <http://embor.embopress.org/content/16/12/1620> [Accessed March 18, 2019].
- Dolatabadi, S. et al., 2017. Cell Cycle and Cell Size Dependent Gene Expression Reveals Distinct Subpopulations at Single-Cell Level. *Frontiers in Genetics*, 8, p.1. Available at: <http://journal.frontiersin.org/article/10.3389/fgene.2017.00001/full> [Accessed March 20, 2019].
- Doyon, Y. et al., 2004. Structural and functional conservation of the NuA4 histone acetyltransferase complex from yeast to humans. *Molecular and cellular biology*, 24(5), pp.1884–96. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/14966270> [Accessed March 19, 2019].
- Du, J. et al., 1998. Sth1p, a *Saccharomyces cerevisiae* Snf2p/Swi2p homolog, is an essential ATPase in RSC and differs from Snf/Swi in its interactions with histones and chromatin-associated proteins. *Genetics*, 150(3), pp.987–1005. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/9799253> [Accessed March 21, 2019].
- Edwards, A.M. et al., 1991. Two dissociable subunits of yeast RNA polymerase II stimulate the initiation of transcription at a promoter in vitro. *The Journal of biological chemistry*,

- 266(1), pp.71–5. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/1985924> [Accessed November 1, 2017].
- Esteller, M., 2007. Cancer epigenomics: DNA methylomes and histone-modification maps. *Nature Reviews Genetics*, 8(4), pp.286–298. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/17339880> [Accessed March 18, 2019].
- Etchegaray, J.P. & Mostoslavsky, R., 2016. Interplay between Metabolism and Epigenetics: A Nuclear Adaptation to Environmental Changes. *Molecular Cell*, 62(5), pp.695–711. Available at: <http://dx.doi.org/10.1016/j.molcel.2016.05.029>.
- Eustermann, S. et al., 2018. Structural basis for ATP-dependent chromatin remodelling by the INO80 complex. *Nature*, 556(7701), pp.386–390. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/29643509> [Accessed March 20, 2019].
- Ewald, J.C. et al., 2016. The Yeast Cyclin-Dependent Kinase Routes Carbon Fluxes to Fuel Cell Cycle Progression. *Molecular cell*, 62(4), pp.532–45. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/27203178> [Accessed March 22, 2019].
- Fan, J. et al., 2015. Metabolic regulation of histone post-translational modifications. *ACS Chemical Biology*, 10(1), pp.95–108.
- Feldman, J.L., Baeza, J. & Denu, J.M., 2013. Activation of the protein deacetylase SIRT6 by long-chain fatty acids and widespread deacylation by mammalian sirtuins. *The Journal of biological chemistry*, 288(43), pp.31350–6. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/24052263> [Accessed March 19, 2019].
- Ferreira, H., Flaus, A. & Owen-Hughes, T., 2007. Histone Modifications Influence the Action of Snf2 Family Remodelling Enzymes by Different Mechanisms. *Journal of Molecular Biology*, 374(3), pp.563–579.
- Fischl, H. et al., 2017. Paf1 Has Distinct Roles in Transcription Elongation and Differential Transcript Fate. *Molecular cell*, 65(4), p.685–698.e8. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/28190769> [Accessed March 22, 2019].
- Fishburn, J., Mohibullah, N. & Hahn, S., 2005. Function of a Eukaryotic Transcription Activator during the Transcription Cycle. *Molecular Cell*, 18(3), pp.369–378. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/15866178> [Accessed March 19, 2019].
- Flanagan, J.F. et al., 2005. Double chromodomains cooperate to recognize the methylated histone H3 tail. *Nature*, 438(7071), pp.1181–1185. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/16372014> [Accessed March 20, 2019].
- Franken, J. et al., 2008. Carnitine and carnitine acetyltransferases in the yeast *Saccharomyces cerevisiae*: a role for carnitine in stress protection. *Current Genetics*, 53(6), pp.347–360. Available at: <http://link.springer.com/10.1007/s00294-008-0191-0> [Accessed March 18, 2019].
- Frauwirth, K.A. et al., 2002. The CD28 signaling pathway regulates glucose metabolism. *Immunity*, 16(6), pp.769–77. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/12121659> [Accessed March 18, 2019].
- Fu, J. et al., 2013. Interplay between *BDF1* and *BDF2* and their roles in regulating the yeast salt stress response. *FEBS Journal*, 280(9), pp.1991–2001. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/23452060> [Accessed March 22, 2019].
- Fujisawa, T. & Filippakopoulos, P., 2017. Functions of bromodomain-containing proteins and

- their roles in homeostasis and cancer. *Nature Reviews Molecular Cell Biology*, 18(4), pp.246–262. Available at: <http://dx.doi.org/10.1038/nrm.2016.143>.
- Futcher, B., 2006. Metabolic cycle, cell cycle, and the finishing kick to start. *Genome Biology*, 7(4).
- Galdieri, L. et al., 2014. Protein acetylation and acetyl coenzyme a metabolism in Budding Yeast. *Eukaryotic Cell*, 13(12), pp.1472–1483.
- Galdieri, L. et al., 2010. Transcriptional Regulation in Yeast during Diauxic Shift and Stationary Phase. *OMICS: A Journal of Integrative Biology*, 14(6), pp.629–638. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/20863251> [Accessed March 19, 2019].
- Gancedo, J.M., 2008. The early steps of glucose signalling in yeast. *FEMS Microbiology Reviews*, 32(4), pp.673–704. Available at: <https://academic.oup.com/femsre/article-lookup/doi/10.1111/j.1574-6976.2008.00117.x> [Accessed March 22, 2019].
- Giaever, G. et al., 2002. Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature*, 418(6896), pp.387–391. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/12140549> [Accessed March 21, 2019].
- Glozak, M.A. & Seto, E., 2007. Histone deacetylases and cancer. *Oncogene*, 26(37), pp.5420–5432. Available at: <http://www.nature.com/articles/1210610> [Accessed March 18, 2019].
- Golombek, D.A. & Rosenstein, R.E., 2010. Physiology of Circadian Entrainment. *Physiological Reviews*, 90(3), pp.1063–1102. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/20664079> [Accessed March 22, 2019].
- Gombert, A.K. et al., 2001. Network Identification and Flux Quantification in the Central Metabolism of *Saccharomyces cerevisiae* under Different Conditions of Glucose Repression. *Journal of Bacteriology*, 183(4), pp.1441–1451. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/11157958> [Accessed March 18, 2019].
- Gomes, A.P. et al., 2013. Declining NAD⁺ Induces a Pseudohypoxic State Disrupting Nuclear-Mitochondrial Communication during Aging. *Cell*, 155(7), pp.1624–1638. Available at: <https://www.sciencedirect.com/science/article/pii/S0092867413015213> [Accessed March 19, 2019].
- Grant, P.A. et al., 1997. Yeast Gcn5 functions in two multisubunit complexes to acetylate nucleosomal histones: characterization of an Ada complex and the SAGA (Spt/Ada) complex. *Genes & development*, 11(13), pp.1640–50. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/9224714> [Accessed March 19, 2019].
- Grüne, T. et al., 2003. Crystal structure and functional analysis of a nucleosome recognition module of the remodeling factor ISWI. *Molecular cell*, 12(2), pp.449–60. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/14536084> [Accessed March 20, 2019].
- Gu, W. et al., 2012. CapSeq and CIP-TAP Identify Pol II Start Sites and Reveal Capped Small RNAs as *C. elegans* piRNA Precursors. *Cell*, 151(7), pp.1488–1500. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/23260138> [Accessed March 22, 2019].
- Guillemette, B. et al., 2005. Variant Histone H2A.Z Is Globally Localized to the Promoters of Inactive Yeast Genes and Regulates Nucleosome Positioning M. Groudine, ed. *PLoS Biology*, 3(12), p.e384. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/16248679> [Accessed March 20, 2019].

- Hagman, A., Säll, T. & Piškur, J., 2014. Analysis of the yeast short-term Crabtree effect and its origin. *The FEBS journal*, 281(21), pp.4805–14. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/25161062> [Accessed March 22, 2019].
- Hallows, W.C., Lee, S. & Denu, J.M., 2006. Sirtuins deacetylate and activate mammalian acetyl-CoA synthetases. *Proceedings of the National Academy of Sciences*, 103(27), pp.10230–10235.
- Harbers, M. & Kahl, G., 2011. *Tag-based next generation sequencing*, John Wiley & Sons.
- Harris, R.A. et al., 2005. Overview of the Molecular and Biochemical Basis of Branched-Chain Amino Acid Catabolism. *The Journal of Nutrition*, 135(6), p.1527S–1530S. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/15930464> [Accessed March 18, 2019].
- Hartley, P.D. & Madhani, H.D., 2009. Mechanisms that Specify Promoter Nucleosome Location and Identity. *Cell*, 137(3), pp.445–458. Available at: <https://www.sciencedirect.com/science/article/pii/S0092867409002578> [Accessed March 22, 2019].
- Haruki, H., Nishikawa, J. & Laemmli, U.K., 2008. The Anchor-Away Technique: Rapid, Conditional Establishment of Yeast Mutant Phenotypes. *Molecular Cell*, 31(6), pp.925–932. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/18922474> [Accessed March 20, 2019].
- Hashimoto, S. et al., 2004. 5'-end SAGE for the analysis of transcriptional start sites. *Nature Biotechnology*, 22(9), pp.1146–1149. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/15300261> [Accessed March 22, 2019].
- Hassan, A.H. et al., 2002a. Function and selectivity of bromodomains in anchoring chromatin-modifying complexes to promoter nucleosomes. *Cell*, 111(3), pp.369–79. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/12419247> [Accessed March 20, 2019].
- Hassan, A.H. et al., 2002b. Function and selectivity of bromodomains in anchoring chromatin-modifying complexes to promoter nucleosomes. *Cell*, 111(3), pp.369–79. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/12419247> [Accessed March 22, 2019].
- Hassan, A.H., Awad, S. & Prochasson, P., 2006. The Swi2/Snf2 bromodomain is required for the displacement of SAGA and the octamer transfer of SAGA+acetylated nucleosomes. *Journal of Biological Chemistry*, 281(26), pp.18126–18134.
- Haynes, S.R. et al., 1992. The bromodomain: a conserved sequence found in human, Drosophila and yeast proteins. *Nucleic acids research*, 20(10), p.2603. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/1350857> [Accessed March 20, 2019].
- Henry, K.A. et al., 2010. The Unfolded Protein Response Is Not Necessary for the G1/S Transition, but It Is Required for Chromosome Maintenance in *Saccharomyces cerevisiae* M. Kaeberlein, ed. *PLoS ONE*, 5(9), p.e12732. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/20856872> [Accessed March 20, 2019].
- Herzig, S. et al., 2012. Identification and Functional Expression of the Mitochondrial Pyruvate Carrier. *Science*, 337(6090), pp.93–96. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/22628554> [Accessed March 18, 2019].
- Hider, R.C. & Hoffbrand, A.V., 2018. The Role of Deferiprone in Iron Chelation D. L. Longo, ed. *New England Journal of Medicine*, 379(22), pp.2140–2150. Available at:

- <http://www.nejm.org/doi/10.1056/NEJMra1800219> [Accessed March 21, 2019].
- Hino, S. et al., 2012. FAD-dependent lysine-specific demethylase-1 regulates cellular energy expenditure. *Nature Communications*, 3, pp.712–758. Available at: <http://dx.doi.org/10.1038/ncomms1755>.
- Honkela, A. et al., 2015. Genome-wide modeling of transcription kinetics reveals patterns of RNA production delays. *Proceedings of the National Academy of Sciences of the United States of America*, 112(42), pp.13115–20. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/26438844> [Accessed March 23, 2019].
- Houten, S.M. et al., 2016. The Biochemistry and Physiology of Mitochondrial Fatty Acid β -Oxidation and Its Genetic Disorders. *The Annual Review of Physiology is online at*, 78, pp.23–44. Available at: www.annualreviews.org [Accessed March 18, 2019].
- Hovland, P. et al., 1989. Galactose as a gratuitous inducer of GAL gene expression in yeasts growing on glucose. *Gene*, 83(1), pp.57–64. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/2512199> [Accessed March 22, 2019].
- Huh, W.-K. et al., 2003. Global analysis of protein localization in budding yeast. *Nature*, 425(6959), pp.686–691. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/14562095> [Accessed March 18, 2019].
- Huppertz, I. et al., 2014. iCLIP: Protein–RNA interactions at nucleotide resolution. *Methods*, 65(3), pp.274–287. Available at: <https://www.sciencedirect.com/science/article/pii/S1046202313004143> [Accessed March 22, 2019].
- Husse, J., Eichele, G. & Oster, H., 2015. Synchronization of the mammalian circadian timing system: Light can control peripheral clocks independently of the SCN clock: alternate routes of entrainment optimize the alignment of the body's circadian clock network with external time. *BioEssays : news and reviews in molecular, cellular and developmental biology*, 37(10), pp.1119–28. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/26252253> [Accessed March 21, 2019].
- Ito, A. et al., 2001. p300/CBP-mediated p53 acetylation is commonly induced by p53-activating agents and inhibited by MDM2. *The EMBO journal*, 20(6), pp.1331–40. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/11250899> [Accessed March 19, 2019].
- Jacobson, R.H. et al., 2000. Structure and function of a human TAFII250 double bromodomain module. *Science (New York, N.Y.)*, 288(5470), pp.1422–5. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/10827952> [Accessed March 19, 2019].
- Jaehning, J.A., 2010. The Paf1 complex: platform or player in RNA polymerase II transcription? *Biochimica et biophysica acta*, 1799(5–6), pp.379–88. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/20060942> [Accessed March 22, 2019].
- Janke, R., Dodson, A.E. & Rine, J., 2015. Metabolism and Epigenetics. *Annual Review of Cell and Developmental Biology*, 31(1), pp.473–496.
- Jiang, W. et al., 2011. Acetylation Regulates Gluconeogenesis by Promoting PEPCK1 Degradation via Recruiting the UBR5 Ubiquitin Ligase. *Molecular Cell*, 43(1), pp.33–44. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/21726808> [Accessed March 18, 2019].
- Jónsson, Z.O. et al., 2004. Rvb1p/Rvb2p Recruit Arp5p and Assemble a Functional Ino80

- Chromatin Remodeling Complex. *Molecular Cell*, 16(3), pp.465–477. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/15525518> [Accessed March 20, 2019].
- Joshi, A.A. & Struhl, K., 2005. Eaf3 Chromodomain Interaction with Methylated H3-K36 Links Histone Deacetylation to Pol II Elongation. *Molecular Cell*, 20(6), pp.971–978. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/16364921> [Accessed March 19, 2019].
- Josling, G.A. et al., 2012. The role of bromodomain proteins in regulating gene expression. *Genes*, 3(2), pp.320–343.
- Kaelin, W.G. & McKnight, S.L., 2013. Influence of metabolism on epigenetics and disease. *Cell*, 153(1), pp.56–69. Available at: <http://dx.doi.org/10.1016/j.cell.2013.03.004>.
- Kals, M. et al., 2005. YPL.db2: the yeast protein localization database, version 2.0. *Yeast*, 22(3), pp.213–218. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/15704222> [Accessed March 18, 2019].
- Kapoor, P. et al., 2013. Evidence for monomeric actin function in INO80 chromatin remodeling. *Nature structural & molecular biology*, 20(4), pp.426–32. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/23524535> [Accessed March 20, 2019].
- Kasten, M. et al., 2004. Tandem bromodomains in the chromatin remodeler RSC recognize acetylated histone H3 Lys14. *EMBO Journal*, 23(6), pp.1348–1359.
- Katoh, Y. et al., 2011. Methionine Adenosyltransferase II Serves as a Transcriptional Corepressor of Maf Oncoprotein. *Molecular Cell*, 41(5), pp.554–566. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/21362551> [Accessed March 18, 2019].
- Kemmeren, P. et al., 2014. Large-Scale Genetic Perturbations Reveal Regulatory Networks and an Abundance of Gene-Specific Repressors. *Cell*, 157(3), pp.740–752. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/24766815> [Accessed March 20, 2019].
- Kennedy, M.C. et al., 1992. Purification and characterization of cytosolic aconitase from beef liver and its relationship to the iron-responsive element binding protein. *Proceedings of the National Academy of Sciences of the United States of America*, 89(24), pp.11730–4. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/1334546> [Accessed March 21, 2019].
- Keogh, M.-C. et al., 2005. Cotranscriptional Set2 Methylation of Histone H3 Lysine 36 Recruits a Repressive Rpd3 Complex. *Cell*, 123(4), pp.593–605. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/16286008> [Accessed March 19, 2019].
- Khodor, Y.L. et al., 2011. Nascent-seq indicates widespread cotranscriptional pre-mRNA splicing in Drosophila. *Genes & Development*, 25(23), pp.2502–2512. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/22156210> [Accessed March 20, 2019].
- Khorasanizadeh, S., 2004. The nucleosome: from genomic organization to genomic regulation. *Cell*, 116(2), pp.259–72. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/14744436> [Accessed March 18, 2019].
- Kim, B. et al., 2007. The transcription elongation factor TFIIIS is a component of RNA polymerase II preinitiation complexes. *Proceedings of the National Academy of Sciences of the United States of America*, 104(41), pp.16068–73. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/17913884> [Accessed March 22, 2019].
- Kimura, M., Ishiguro, A. & Ishihama, A., RNA Polymerase II Subunits 2, 3, and 11 Form a Core Subassembly with DNA Binding Activity*. Available at:

- <http://www.jbc.org/content/272/41/25851.full.pdf> [Accessed April 24, 2017].
- Kinkelin, K. et al., 2013. Structures of RNA polymerase II complexes with Bye1, a chromatin-binding PHF3/DIDO homologue. *Proceedings of the National Academy of Sciences*, 110(38), pp.15277–15282. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/24003114> [Accessed March 22, 2019].
- Kirk G. Tanner, Michael R. Langer, and & Denu*, J.M., 2000. Kinetic Mechanism of Human Histone Acetyltransferase P/CAF[†]. Available at: <https://pubs.acs.org/doi/abs/10.1021/bi001272h> [Accessed March 19, 2019].
- Kleff, S. et al., 1995. Identification of a gene encoding a yeast histone H4 acetyltransferase. *The Journal of biological chemistry*, 270(42), pp.24674–7. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/7559580> [Accessed March 19, 2019].
- Knutson, B.A. & Hahn, S., 2011. Domains of Tra1 Important for Activator Recruitment and Transcription Coactivator Functions of SAGA and NuA4 Complexes. *Molecular and Cellular Biology*, 31(4), pp.818–831. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/21149579> [Accessed March 19, 2019].
- Kobor, M.S. et al., 2004. A Protein Complex Containing the Conserved Swi2/Snf2-Related ATPase Swr1p Deposits Histone Variant H2A.Z into Euchromatin Peter Becker, ed. *PLoS Biology*, 2(5), p.e131. Available at: <http://dx.plos.org/10.1371/journal.pbio.0020131> [Accessed March 22, 2019].
- Kouzarides, T., 2007. Chromatin Modifications and Their Function. *Cell*, 128(4), pp.693–705. Available at: <https://www.sciencedirect.com/science/article/pii/S0092867407001845> [Accessed March 18, 2019].
- Kretschmer, M., Wang, J. & Kronstad, J.W., 2012. Peroxisomal and Mitochondrial β -Oxidation Pathways Influence the Virulence of the Pathogenic Fungus *Cryptococcus neoformans*. *Eukaryotic Cell*, 11(8), p.1042. Available at: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3416069/> [Accessed May 29, 2019].
- Krietenstein, N. et al., 2016. Genomic Nucleosome Organization Reconstituted with Pure Proteins. *Cell*, 167(3), p.709–721.e12. Available at: <https://www.sciencedirect.com/science/article/pii/S0092867416313307> [Accessed March 22, 2019].
- Krogan, N.J. et al., 2003. A Snf2 Family ATPase Complex Required for Recruitment of the Histone H2A Variant Htz1. *Molecular Cell*, 12(6), pp.1565–1576. Available at: <https://www.sciencedirect.com/science/article/pii/S1097276503004970> [Accessed March 22, 2019].
- Kuang, Z. et al., 2014. High-temporal-resolution view of transcription and chromatin states across distinct metabolic states in budding yeast. *Nature structural & molecular biology*, 21(10), pp.854–63. Available at: <http://dx.doi.org/10.1038/nsmb.2881>.
- Kuo, M.H. et al., 2000. Gcn4 activator targets Gcn5 histone acetyltransferase to specific promoters independently of transcription. *Molecular cell*, 6(6), pp.1309–20. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/11163205> [Accessed March 19, 2019].
- Kwak, H. et al., 2013. Precise maps of RNA polymerase reveal how promoters direct initiation and pausing. *Science (New York, N.Y.)*, 339(6122), pp.950–3. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/23430654> [Accessed March 22, 2019].
- Kwak, H. et al., 2013. Precise Maps of RNA Polymerase Reveal How Promoters Direct

- Initiation and Pausing. *Science*, 339(6122), pp.950–953. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/23430654> [Accessed March 20, 2019].
- Ladurner, A.G. et al., 2003. Bromodomains Mediate an Acetyl-Histone Encoded Antisilencing Function at Heterochromatin Boundaries. *Molecular Cell*, 11(2), pp.365–376. Available at: <https://www.sciencedirect.com/science/article/pii/S1097276503000352> [Accessed March 20, 2019].
- Lahtvee, P.-J. et al., 2017. Absolute Quantification of Protein and mRNA Abundances Demonstrate Variability in Gene-Specific Translation Efficiency in Yeast. *Cell systems*, 4(5), p.495–504.e5. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/28365149> [Accessed June 21, 2018].
- Lam, M.T.Y. et al., 2013. Rev-Erbs repress macrophage gene expression by inhibiting enhancer-directed transcription. *Nature*, 498(7455), pp.511–515. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/23728303> [Accessed March 22, 2019].
- Lamonica, J.M. et al., 2011. Bromodomain protein Brd3 associates with acetylated GATA1 to promote its chromatin occupancy at erythroid target genes. *Proceedings of the National Academy of Sciences*, 108(22), pp.E159–E168. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/21536911> [Accessed March 21, 2019].
- Lange, M. et al., 2008. Regulation of muscle development by DPF3, a novel histone acetylation and methylation reader of the BAF chromatin remodeling complex. *Genes & Development*, 22(17), pp.2370–2384. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/18765789> [Accessed March 19, 2019].
- Langer, M.R. et al., 2002. Modulating Acetyl-CoA Binding in the GCN5 Family of Histone Acetyltransferases. *Journal of Biological Chemistry*, 277(30), pp.27337–27344. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/11994311> [Accessed March 19, 2019].
- Lawrence, M., Daujat, S. & Schneider, R., 2016. Lateral Thinking: How Histone Modifications Regulate Gene Expression. *Trends in Genetics*, 32(1), pp.42–56. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/26704082> [Accessed March 18, 2019].
- Lee, F.J., Lin, L.W. & Smith, J.A., 1996. Acetyl-CoA hydrolase involved in acetate utilization in *Saccharomyces cerevisiae*. *Biochimica et biophysica acta*, 1297(1), pp.105–9. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/8841387> [Accessed March 18, 2019].
- Lee, J.V., Carrer, A., et al., 2014. Akt-Dependent Metabolic Reprogramming Regulates Tumor Cell Histone Acetylation. *Cell Metabolism*, 20(2), pp.306–319. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/24998913> [Accessed March 19, 2019].
- Lee, J.V., Haas, N.B., et al., 2014. Akt-Dependent Metabolic Reprogramming Regulates Tumor Cell Histone Acetylation. *Cell Metabolism*, 20(2), pp.306–319. Available at: <http://dx.doi.org/10.1016/j.cmet.2014.06.004>.
- Li, E., Bestor, T.H. & Jaenisch, R., 1992. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell*, 69(6), pp.915–26. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/1606615> [Accessed March 18, 2019].
- Li, Q. et al., 2008. Acetylation of histone H3 lysine 56 regulates replication-coupled nucleosome assembly. *Cell*, 134(2), pp.244–55. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/18662540> [Accessed March 19, 2019].
- Li, S. & Shogren-Knaak, M.A., 2009. The Gcn5 bromodomain of the SAGA complex facilitates cooperative and cross-tail acetylation of nucleosomes. *The Journal of biological*

- chemistry*, 284(14), pp.9411–7. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/19218239> [Accessed March 22, 2019].
- Liang, H. & Gaber, R.F., 1996. A novel signal transduction pathway in *Saccharomyces cerevisiae* defined by Snf3-regulated expression of HXT6. *Molecular biology of the cell*, 7(12), pp.1953–66. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/8970157> [Accessed March 22, 2019].
- Liu, X. et al., 2008. The structural basis of protein acetylation by the p300/CBP transcriptional coactivator. *Nature*, 451(7180), pp.846–850. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/18273021> [Accessed March 19, 2019].
- Lloret-Llinares, M. et al., 2018. The RNA exosome contributes to gene expression regulation during stem cell differentiation. *Nucleic Acids Research*, 46(21), pp.11502–11513. Available at: <https://academic.oup.com/nar/article/46/21/11502/5096074> [Accessed March 23, 2019].
- Lobet, Y., Lhoest, J. & Colson, C., 1989. Partial purification and characterization of the specific protein-lysine N-methyltransferase of YL32, a yeast ribosomal protein. *Biochimica et biophysica acta*, 997(3), pp.224–31. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/2504290> [Accessed March 18, 2019].
- Lombardi, P.M. et al., 2011. Structure, mechanism, and inhibition of histone deacetylases and related metalloenzymes. *Current opinion in structural biology*, 21(6), pp.735–43. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/21872466> [Accessed March 19, 2019].
- Longtine, M.S. et al., 1998. Additional modules for versatile and economical PCR-based gene deletion and modification in *Saccharomyces cerevisiae*. *Yeast*, 14(10), pp.953–961. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/9717241> [Accessed March 22, 2019].
- Lonze, B.E. & Ginty, D.D., 2002. Function and Regulation of CREB Family Transcription Factors in the Nervous System. *Neuron*, 35(4), pp.605–623. Available at: <https://www.sciencedirect.com/science/article/pii/S0896627302008280> [Accessed March 18, 2019].
- Losman, J.-A. & Kaelin, W.G., 2013. What a difference a hydroxyl makes: mutant IDH, (R)-2-hydroxyglutarate, and cancer. *Genes & Development*, 27(8), pp.836–852. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/23630074> [Accessed March 18, 2019].
- Lu, C. et al., 2012. IDH mutation impairs histone demethylation and results in a block to cell differentiation. *Nature*, 483(7390), pp.474–478. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/22343901> [Accessed March 18, 2019].
- Lu, Q. et al., 2003. Acetylation of cAMP-responsive Element-binding Protein (CREB) by CREB-binding Protein Enhances CREB-dependent Transcription. *Journal of Biological Chemistry*, 278(18), pp.15727–15734. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/12595525> [Accessed March 18, 2019].
- Mahat, D.B. et al., 2016. Base-pair-resolution genome-wide mapping of active RNA polymerases using precision nuclear run-on (PRO-seq). *Nature Protocols*, 11(8), pp.1455–1476. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/27442863> [Accessed March 20, 2019].
- Mahmood, S. et al., 2016. Lack of mitochondria-generated acetyl-CoA by pyruvate

- dehydrogenase complex downregulates gene expression in the hepatic de novo lipogenic pathway. *American journal of physiology. Endocrinology and metabolism*, 311(1), pp.E117-27. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/27166281> [Accessed March 18, 2019].
- Malik, P. & Cashen, A.F., 2014. Decitabine in the treatment of acute myeloid leukemia in elderly patients. *Cancer management and research*, 6, pp.53–61. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/24520204> [Accessed March 18, 2019].
- Marguerat, S. et al., 2014. Contributions of transcription and mRNA decay to gene expression dynamics of fission yeast in response to oxidative stress. *RNA biology*, 11(6), pp.702–14. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/25007214> [Accessed March 20, 2019].
- Mariño, G. et al., 2014. Regulation of Autophagy by Cytosolic Acetyl-Coenzyme A. *Molecular Cell*, 53(5), pp.710–725. Available at: <https://www.sciencedirect.com/science/article/pii/S109727651400077X> [Accessed March 19, 2019].
- Marmorstein, R. & Zhou, M.-M., 2014. Writers and readers of histone acetylation: structure, mechanism, and inhibition. *Cold Spring Harbor perspectives in biology*, 6(7), p.a018762. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/24984779> [Accessed March 19, 2019].
- Martinez-Rucobo, F.W. et al., 2011. Architecture of the RNA polymerase-Spt4/5 complex and basis of universal transcription processivity. *The EMBO journal*, 30(7), pp.1302–10. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/21386817> [Accessed March 22, 2019].
- Maruyama, K. & Sugano, S., 1994. Oligo-capping: a simple method to replace the cap structure of eukaryotic mRNAs with oligoribonucleotides. *Gene*, 138(1–2), pp.171–4. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/8125298> [Accessed March 22, 2019].
- Mason, P.B. & Struhl, K., 2003. The FACT complex travels with elongating RNA polymerase II and is important for the fidelity of transcriptional initiation in vivo. *Molecular and cellular biology*, 23(22), pp.8323–33. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/14585989> [Accessed March 22, 2019].
- Masumoto, H. et al., 2005. A role for cell-cycle-regulated histone H3 lysine 56 acetylation in the DNA damage response. *Nature*, 436(7048), pp.294–298. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/16015338> [Accessed March 19, 2019].
- Matangkasombut, O. & Buratowski, S., 2003. Different sensitivities of bromodomain factors 1 and 2 to histone H4 acetylation. *Molecular cell*, 11(2), pp.353–63. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/12620224> [Accessed March 22, 2019].
- Mato, J.M. et al., 2017. One-Carbon Metabolism in Liver Health and Disease. *Liver Pathophysiology: Therapies and Antioxidants*, 25(1), pp.761–765. Available at: <http://dx.doi.org/10.1016/j.cmet.2016.08.009>.
- Matsuda, S. et al., 2016. Nuclear pyruvate kinase M2 complex serves as a transcriptional coactivator of arylhydrocarbon receptor. *Nucleic Acids Research*, 44(2), pp.636–647. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/26405201> [Accessed March 19, 2019].

- Mellor, J., 2016. The molecular basis of metabolic cycles and their relationship to circadian rhythms. *Nature Structural & Molecular Biology*, 23(12), pp.1035–1044. Available at: <http://www.nature.com/articles/nsmb.3311> [Accessed March 22, 2019].
- Mellor, J., Woloszczuk, R. & Howe, F.S., 2016. The Interleaved Genome. *Trends in Genetics*, 32(1), pp.57–71. Available at: <https://www.sciencedirect.com/science/article/pii/S0168952515001924> [Accessed March 22, 2019].
- Melnik, S. et al., 2000. Measurement of plasma and intracellular S-adenosylmethionine and S-adenosylhomocysteine utilizing coulometric electrochemical detection: alterations with plasma homocysteine and pyridoxal 5'-phosphate concentrations. *Clinical chemistry*, 46(2), pp.265–72. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/10657384> [Accessed March 18, 2019].
- Menzies, K.J. et al., 2016. Protein acetylation in metabolism-metabolites and cofactors. *Nature Reviews Endocrinology*, 12(1), pp.43–60. Available at: <http://dx.doi.org/10.1038/nrendo.2015.181>.
- Michishita, E. et al., 2008. SIRT6 is a histone H3 lysine 9 deacetylase that modulates telomeric chromatin. *Nature*, 452(7186), pp.492–496. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/18337721> [Accessed March 19, 2019].
- Miller, T.J. et al., 2002. A Real-Time Ratiometric Method for the Determination of Molecular Oxygen Inside Living Cells Using Sol–Gel-Based Spherical Optical Nanosensors with Applications to Rat C6 Glioma. *Analytical Chemistry*, 73(17), pp.4124–4133.
- Mizuguchi, G. et al., 2004a. ATP-Driven Exchange of Histone H2AZ Variant Catalyzed by SWR1 Chromatin Remodeling Complex. *Science*, 303(5656), pp.343–348. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/14645854> [Accessed March 20, 2019].
- Mizuguchi, G. et al., 2004b. ATP-Driven Exchange of Histone H2AZ Variant Catalyzed by SWR1 Chromatin Remodeling Complex. *Science*, 303(5656), pp.343–348. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/14645854> [Accessed March 22, 2019].
- Mizzen, C.A. et al., 1996. The TAF(II)250 subunit of TFIID has histone acetyltransferase activity. *Cell*, 87(7), pp.1261–1270. Available at: [http://dx.doi.org/10.1016/S0092-8674\(00\)81821-8](http://dx.doi.org/10.1016/S0092-8674(00)81821-8).
- Montgomery, D.C. et al., 2015. Metabolic Regulation of Histone Acetyltransferases by Endogenous Acyl-CoA Cofactors. *Chemistry and Biology*, 22(8), pp.1030–1039. Available at: <http://dx.doi.org/10.1016/j.chembiol.2015.06.015>.
- Moreau, J.-L. et al., 2003. Regulated displacement of TBP from the PHO8 promoter in vivo requires Cbf1 and the Isw1 chromatin remodeling complex. *Molecular cell*, 11(6), pp.1609–20. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/12820973> [Accessed March 20, 2019].
- Moussaieff, A., Kogan, N.M. & Aberdam, D., 2015. Concise Review: Energy Metabolites: Key Mediators of the Epigenetic State of Pluripotency. *STEM CELLS*, 33(8), pp.2374–2380. Available at: <http://doi.wiley.com/10.1002/stem.2041> [Accessed March 18, 2019].
- Mueller-Planitz, F. et al., 2013. The ATPase domain of ISWI is an autonomous nucleosome remodeling machine. *Nature Structural & Molecular Biology*, 20(1), pp.82–89. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/23202585> [Accessed March 20, 2019].
- Müller, D. et al., 2003. Cyclic AMP mediates the cell cycle dynamics of energy metabolism in

- Saccharomyces cerevisiae. *Yeast*, 20, pp.351–367. Available at: www.interscience.wiley.com [Accessed March 22, 2019].
- Murata, M. et al., 2014. Detecting Expressed Genes Using CAGE. In *Methods in molecular biology (Clifton, N.J.)*. pp. 67–85. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/24927836> [Accessed March 22, 2019].
- Murray, D.B., Klevecz, R.R. & Lloyd, D., 2003. Generation and maintenance of synchrony in *Saccharomyces cerevisiae* continuous culture. *Experimental cell research*, 287(1), pp.10–5. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/12799177> [Accessed March 20, 2019].
- Nagoshi, E. et al., 2004. Circadian Gene Expression in Individual Fibroblasts. *Cell*, 119(5), pp.693–705. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/15550250> [Accessed March 22, 2019].
- Narahari, J. et al., 2000. The aconitase function of iron regulatory protein 1. Genetic studies in yeast implicate its role in iron-mediated redox regulation. *The Journal of biological chemistry*, 275(21), pp.16227–34. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/10748225> [Accessed March 21, 2019].
- Natsume, T. & Kanemaki, M.T., 2017. Conditional Degrons for Controlling Protein Expression at the Protein Level. *Annual Review of Genetics*, 51(1), pp.83–102. Available at: <http://www.annualreviews.org/doi/10.1146/annurev-genet-120116-024656> [Accessed March 26, 2019].
- Nechaev, S. et al., 2010. Global Analysis of Short RNAs Reveals Widespread Promoter-Proximal Stalling and Arrest of Pol II in *Drosophila*. *Science*, 327(5963), pp.335–338. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/20007866> [Accessed March 22, 2019].
- Neuwald, A.F. & Landsman, D., 1997. GCN5-related histone N-acetyltransferases belong to a diverse superfamily that includes the yeast SPT10 protein. *Trends in biochemical sciences*, 22(5), pp.154–5. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/9175471> [Accessed March 19, 2019].
- Ng, H.H. et al., 2002. Genome-wide location and regulated recruitment of the RSC nucleosome-remodeling complex. *Genes and Development*, 16(7), pp.806–819.
- Nguyen, T. et al., 2014. Transcription mediated insulation and interference direct gene cluster expression switches. *eLife*, 3, p.e03635. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/25407679> [Accessed March 22, 2019].
- Nishimura, K. et al., 2009. An auxin-based degron system for the rapid depletion of proteins in nonplant cells. *Nature Methods*, 6(12), pp.917–922. Available at: <http://www.nature.com/articles/nmeth.1401> [Accessed March 20, 2019].
- Nocetti, N. & Whitehouse, I., 2016. Nucleosome repositioning underlies dynamic gene expression. *Genes & development*, 30(6), pp.660–72. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/26966245> [Accessed March 20, 2019].
- Nojima, T. et al., 2015. Mammalian NET-Seq Reveals Genome-wide Nascent Transcription Coupled to RNA Processing. *Cell*, 161(3), pp.526–540. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/25910207> [Accessed March 20, 2019].
- Nojima, T. et al., 2018. RNA Polymerase II Phosphorylated on CTD Serine 5 Interacts with the Spliceosome during Co-transcriptional Splicing. *Molecular Cell*, 72(2), p.369–379.e4.

Available at: <https://www.sciencedirect.com/science/article/pii/S1097276518307512> [Accessed March 22, 2019].

- Noushmehr, H. et al., 2010. Identification of a CpG Island Methylator Phenotype that Defines a Distinct Subgroup of Glioma. *Cancer Cell*, 17(5), pp.510–522. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/20399149> [Accessed March 18, 2019].
- Oden, K.L. & Clarke, S., 1983. S-adenosyl-L-methionine synthetase from human erythrocytes: role in the regulation of cellular S-adenosylmethionine levels. *Biochemistry*, 22(12), pp.2978–86. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/6871179> [Accessed March 18, 2019].
- Okano, M. et al., 1999. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell*, 99(3), pp.247–57. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/10555141> [Accessed March 18, 2019].
- Othman, S. et al., 2007. Selective recognition of acetylated histones by bromodomains in transcriptional co-activators. *Biochemical Journal*, 402(1), pp.125–133.
- Paik, W.K. et al., 1970. Nonenzymatic acetylation of histones with acetyl-CoA. *Biochimica et biophysica acta*, 213(2), pp.513–22. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/5534125> [Accessed March 18, 2019].
- Papagiannakis, A. et al., 2017. Autonomous Metabolic Oscillations Robustly Gate the Early and Late Cell Cycle. *Molecular Cell*, 65(2), pp.285–295.
- Parthun, M.R., Widom, J. & Gottschling, D.E., 1996. The major cytoplasmic histone acetyltransferase in yeast: links to chromatin replication and histone metabolism. *Cell*, 87(1), pp.85–94. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/8858151> [Accessed March 19, 2019].
- Patel, M.S. et al., 2014. The pyruvate dehydrogenase complexes: structure-based function and regulation. *The Journal of biological chemistry*, 289(24), pp.16615–23. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/24798336> [Accessed March 18, 2019].
- Pelechano, V., Wei, W. & Steinmetz, L.M., 2013. Extensive transcriptional heterogeneity revealed by isoform profiling. *Nature*, 497(7447), pp.127–131. Available at: <http://www.nature.com/doi/10.1038/nature12121> [Accessed March 22, 2019].
- Peng, M. et al., 2016. Aerobic glycolysis promotes T helper 1 cell differentiation through an epigenetic mechanism. *Science*, 354(6311), pp.481–484. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/27708054> [Accessed March 18, 2019].
- Peng, W. et al., 2008. Regulators of Cellular Levels of Histone Acetylation in *Saccharomyces cerevisiae*. *Genetics*, 179(1), pp.277–289. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/18493053> [Accessed March 21, 2019].
- Porro, D. et al., 1988. Oscillations in Continuous Cultures of Budding Yeast: A Segregated Parameter Analysis. Available at: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/bit.260320402> [Accessed March 22, 2019].
- Pronk, J.T., Yde Steensma, H. & Van Dijken, J.P., 1996. Pyruvate metabolism in *Saccharomyces cerevisiae*. *Yeast (Chichester, England)*, 12(16), pp.1607–33. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/9123965> [Accessed March 18, 2019].
- Qiu, Y. et al., 2006. HDAC1 Acetylation Is Linked to Progressive Modulation of Steroid

- Receptor-Induced Gene Transcription. *Molecular Cell*, 22(5), pp.669–679. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/16762839> [Accessed March 19, 2019].
- Rabani, M. et al., 2014. High-resolution sequencing and modeling identifies distinct dynamic RNA regulatory strategies. *Cell*, 159(7), pp.1698–710. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/25497548> [Accessed March 23, 2019].
- Rabani, M. et al., 2011. Metabolic labeling of RNA uncovers principles of RNA production and degradation dynamics in mammalian cells. *Nature Biotechnology*, 29(5), pp.436–442. Available at: <http://www.nature.com/articles/nbt.1861> [Accessed March 20, 2019].
- Ramsköld, D. et al., 2012. Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. *Nature Biotechnology*, 30(8), pp.777–782. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/22820318> [Accessed March 22, 2019].
- Rao, A.R. & Pellegrini, M., 2011. Regulation of the yeast metabolic cycle by transcription factors with periodic activities. *BMC Systems Biology*, 5(1), p.160. Available at: <http://bmcsystbiol.biomedcentral.com/articles/10.1186/1752-0509-5-160> [Accessed March 21, 2019].
- Rathert, P. et al., 2008. Analysis of the Substrate Specificity of the Dim-5 Histone Lysine Methyltransferase Using Peptide Arrays. *Chemistry & Biology*, 15(1), pp.5–11. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/18215768> [Accessed March 18, 2019].
- Reid, M.A., Dai, Z. & Locasale, J.W., 2017. The impact of cellular metabolism on chromatin dynamics and epigenetics. *Nature Cell Biology*, 19(11), pp.1298–1306. Available at: <http://dx.doi.org/10.1038/ncb3629>.
- Reinke, H., Gregory, P.D. & Hörz, W., 2001. A transient histone hyperacetylation signal marks nucleosomes for remodeling at the PHO8 promoter in vivo. *Molecular Cell*, 7(3), pp.529–538.
- Robert, F. et al., 2004. Global Position and Recruitment of HATs and HDACs in the Yeast Genome. *Molecular Cell*, 16(2), pp.199–209. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/15494307> [Accessed March 19, 2019].
- Robertson, J.B. et al., 2008. Real-time luminescence monitoring of cell-cycle and respiratory oscillations in yeast. *Proceedings of the National Academy of Sciences*, 105(46), pp.17988–17993. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/19004762> [Accessed March 20, 2019].
- Robertson, L.S. & Fink, G.R., 1998. The three yeast A kinases have specific signaling functions in pseudohyphal growth. *Proceedings of the National Academy of Sciences of the United States of America*, 95(23), pp.13783–7. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/9811878> [Accessed March 22, 2019].
- Robyr, D. et al., 2002. Microarray Deacetylation Maps Determine Genome-Wide Functions for Yeast Histone Deacetylases. *Cell*, 109(4), pp.437–446. Available at: <https://www.sciencedirect.com/science/article/pii/S0092867402007468> [Accessed March 19, 2019].
- van Roermund, C.W.T. et al., 2003. Fatty acid metabolism in *Saccharomyces cerevisiae*. *Cellular and Molecular Life Sciences (CMLS)*, 60(9), pp.1838–1851. Available at: <http://link.springer.com/10.1007/s00018-003-3076-x> [Accessed May 29, 2019].
- van Roermund, C.W.T. et al., 1999. Molecular characterization of carnitine-dependent transport of acetyl-CoA from peroxisomes to mitochondria in *Saccharomyces cerevisiae*

- and identification of a plasma membrane carnitine transporter, Agp2p. *The EMBO Journal*, 18(21), pp.5843–5852. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/10545096> [Accessed March 18, 2019].
- Ropero, S. & Esteller, M., 2007. The role of histone deacetylases (HDACs) in human cancer. *Molecular Oncology*, 1(1), pp.19–25. Available at: <https://www.sciencedirect.com/science/article/pii/S1574789107000026> [Accessed March 18, 2019].
- Ryan, D.P. et al., 2011. The DNA-binding domain of the Chd1 chromatin-remodelling enzyme contains SANT and SLIDE domains. *The EMBO Journal*, 30(13), pp.2596–2609. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/21623345> [Accessed March 20, 2019].
- Salah Ud-Din, A., Tikhomirova, A. & Roujeinikova, A., 2016. Structure and Functional Diversity of GCN5-Related N-Acetyltransferases (GNAT). *International Journal of Molecular Sciences*, 17(7), p.1018. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/27367672> [Accessed March 19, 2019].
- Salimullah, M. et al., 2011. NanoCAGE: a high-resolution technique to discover and interrogate cell transcriptomes. *Cold Spring Harbor protocols*, 2011(1), p.pdb.prot5559. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/21205859> [Accessed March 22, 2019].
- Sánchez-Gaya, V. et al., 2018. Elucidating the Role of Chromatin State and Transcription Factors on the Regulation of the Yeast Metabolic Cycle: A Multi-Omic Integrative Approach. *Frontiers in Genetics*, 9, p.578. Available at: <https://www.frontiersin.org/article/10.3389/fgene.2018.00578/full> [Accessed March 21, 2019].
- Santisteban, M.S., Kalashnikova, T. & Smith, M.M., 2000. Histone H2A.Z regulates transcription and is partially redundant with nucleosome remodeling complexes. *Cell*, 103(3), pp.411–22. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/11081628> [Accessed March 20, 2019].
- Saravanan, M. et al., 2012. Interactions between the nucleosome histone core and Arp8 in the INO80 chromatin remodeling complex. *Proceedings of the National Academy of Sciences*, 109(51), pp.20883–20888. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/23213201> [Accessed March 20, 2019].
- Satroutdinov, A.D., Kuriyama, H. & Kobayashi, H., 1992. Oscillatory metabolism of *Saccharomyces cerevisiae* in continuous culture. *FEMS Microbiology Letters*, 98(1–3), pp.261–267. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/1334018> [Accessed March 20, 2019].
- Schulz, D. et al., 2013a. Transcriptome Surveillance by Selective Termination of Noncoding RNA Synthesis. *Cell*, 155(5), pp.1075–1087. Available at: <https://www.sciencedirect.com/science/article/pii/S0092867413013007> [Accessed March 22, 2019].
- Schulz, D. et al., 2013b. Transcriptome Surveillance by Selective Termination of Noncoding RNA Synthesis. *Cell*, 155(5), pp.1075–1087. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/24210918> [Accessed March 22, 2019].
- Schwalb, B. et al., 2016. TT-seq maps the human transient transcriptome. *Science*, 352(6290).

- Shahbazian, M.D. & Grunstein, M., 2007. Functions of Site-Specific Histone Acetylation and Deacetylation. *Annual Review of Biochemistry*, 76(1), pp.75–100.
- Shen, X. et al., 2003. Involvement of actin-related proteins in ATP-dependent chromatin remodeling. *Molecular cell*, 12(1), pp.147–55. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/12887900> [Accessed March 20, 2019].
- Shi, L. et al., 2010. Trehalose is a key determinant of the quiescent metabolic state that fuels cell cycle progression upon return to growth. *Molecular biology of the cell*, 21(12), pp.1982–90. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/20427572> [Accessed March 22, 2019].
- Shi, L. & Tu, B.P., 2013. Acetyl-CoA induces transcription of the key G1 cyclin CLN3 to promote entry into the cell division cycle in *Saccharomyces cerevisiae*. *Proceedings of the National Academy of Sciences*, 110(18), pp.7318–7323.
- Shi, Y. et al., 2004. Histone Demethylation Mediated by the Nuclear Amine Oxidase Homolog LSD1. *Cell*, 119(7), pp.941–953. Available at: <https://www.sciencedirect.com/science/article/pii/S0092867404011997> [Accessed March 18, 2019].
- Shi, Y. & Whetstone, J.R., 2007. Dynamic Regulation of Histone Lysine Methylation by Demethylases. *Molecular Cell*, 25(1), pp.1–14. Available at: <https://www.sciencedirect.com/science/article/pii/S1097276506008689> [Accessed March 18, 2019].
- Shimazu, T. et al., 2013. Suppression of Oxidative Stress by γ -Hydroxybutyrate, an Endogenous Histone Deacetylase Inhibitor. *Science*, 339(6116), pp.211–214. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/23223453> [Accessed March 19, 2019].
- Shyh-Chang, N. & Daley, G.Q., 2015. Metabolic Switches Linked to Pluripotency and Embryonic Stem Cell Differentiation. *Cell Metabolism*, 21(3), pp.349–350. Available at: <https://www.sciencedirect.com/science/article/pii/S1550413115000637> [Accessed March 18, 2019].
- Siegfried, Z. et al., 1999. DNA methylation represses transcription in vivo. *Nature Genetics*, 22(2), pp.203–206. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/10369268> [Accessed March 18, 2019].
- Skiniotis, G., Moazed, D. & Walz, T., 2007. Acetylated histone tail peptides induce structural rearrangements in the RSC chromatin remodeling complex. *Journal of Biological Chemistry*, 282(29), pp.20804–20808.
- Slavov, N. & Botstein, D., 2011. Coupling among growth rate response, metabolic cycle, and cell division cycle in yeast. *Molecular Biology of the Cell*, 22(12), pp.1997–2009. Available at: <http://www.molbiolcell.org/cgi/doi/10.1091/mbc.E11-02-0132> [Accessed February 20, 2017].
- Smolle, M., Workman, J.L. & Venkatesh, S., 2013. reSETting chromatin during transcription elongation. *Epigenetics*, 8(1), pp.10–15. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/23257840> [Accessed March 19, 2019].
- Sobel, R.E. et al., 1995. Conservation of deposition-related acetylation sites in newly synthesized histones H3 and H4. *Proceedings of the National Academy of Sciences of the United States of America*, 92(4), pp.1237–41. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/7862667> [Accessed March 19, 2019].

- Soularue, P. et al., 2006. Rsc4 Connects the Chromatin Remodeler RSC to RNA Polymerases. *Molecular and Cellular Biology*, 26(13), pp.4920–4933.
- Soutoglou, E., Katrakili, N. & Talianidis, I., 2000. Acetylation regulates transcription factor activity at multiple levels. *Molecular Cell*, 5(4), pp.745–751.
- Spain, M.M. et al., 2014. The RSC Complex Localizes to Coding Sequences to Regulate Pol II and Histone Occupancy. *Molecular Cell*, 56(5), pp.653–666. Available at: <http://linkinghub.elsevier.com/retrieve/pii/S1097276514007862> [Accessed March 21, 2017].
- Starheim, K.K., Gevaert, K. & Arnesen, T., 2012. Protein N-terminal acetyltransferases: when the start matters. *Trends in Biochemical Sciences*, 37(4), pp.152–161. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/22405572> [Accessed March 18, 2019].
- Steunou, A.-L., Rossetto, D. & Côté, J., 2014. Regulating Chromatin by Histone Acetylation. In *Fundamentals of Chromatin*. New York, NY: Springer New York, pp. 147–212. Available at: http://link.springer.com/10.1007/978-1-4614-8624-4_4 [Accessed March 19, 2019].
- Su, D. et al., 2012. Structural basis for recognition of H3K56-acetylated histone H3–H4 by the chaperone Rtt106. *Nature*, 483(7387), pp.104–107. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/22307274> [Accessed March 19, 2019].
- Suganuma, T. & Workman, J.L., 2018. Chromatin and Metabolism. *Annual Review of Biochemistry*, 87(1), pp.27–49.
- Sugiyama, T. et al., 2007. SHREC, an Effector Complex for Heterochromatic Transcriptional Silencing. *Cell*, 128(3), pp.491–504. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/17289569> [Accessed March 20, 2019].
- Szerlong, H. et al., 2008. The HSA domain binds nuclear actin-related proteins to regulate chromatin-remodeling ATPases. *Nature Structural & Molecular Biology*, 15(5), pp.469–476. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/18408732> [Accessed March 20, 2019].
- Szerlong, H., Saha, A. & Cairns, B.R., 2003. The nuclear actin-related proteins Arp7 and Arp9: a dimeric module that cooperates with architectural proteins for chromatin remodeling. *The EMBO journal*, 22(12), pp.3175–87. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/12805231> [Accessed March 20, 2019].
- Takahashi, H. et al., 2006. Nucleocytoplasmic Acetyl-Coenzyme A Synthetase Is Required for Histone Acetylation and Global Transcription. *Molecular Cell*, 23(2), pp.207–217. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/16857587> [Accessed March 18, 2019].
- Tamaki, H., 2007. Glucose-stimulated cAMP-protein kinase a pathway in yeast *Saccharomyces cerevisiae*. *Journal of Bioscience and Bioengineering*, 104(4), pp.245–250. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/18023794> [Accessed March 22, 2019].
- Tang, D.T.P. et al., 2013. Suppression of artifacts and barcode bias in high-throughput transcriptome analyses utilizing template switching. *Nucleic Acids Research*, 41(3), pp.e44–e44. Available at: <https://academic.oup.com/nar/article/41/3/e44/2902349> [Accessed March 22, 2019].
- Tang, L., Nogales, E. & Ciferri, C., 2010. Structure and function of SWI/SNF chromatin remodeling complexes and mechanistic implications for transcription. *Progress in*

- biophysics and molecular biology*, 102(2–3), pp.122–8. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/20493208> [Accessed March 20, 2019].
- Tang, Y. et al., 2018. Overexpression of PCK1 Gene Antagonizes Hepatocellular Carcinoma Through the Activation of Gluconeogenesis and Suppression of Glycolysis Pathways. *Cellular Physiology and Biochemistry*, 47(1), pp.344–355. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/29768256> [Accessed March 18, 2019].
- Tanner, K.G. et al., 2000. Kinetic Mechanism of the Histone Acetyltransferase GCN5 from Yeast. *Journal of Biological Chemistry*, 275(29), pp.22048–22055. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/10811654> [Accessed March 19, 2019].
- Teixeira, M.C. et al., 2018. YEASTRACT: an upgraded database for the analysis of transcription regulatory networks in *Saccharomyces cerevisiae*. *Nucleic Acids Research*, 46(D1), pp.D348–D353. Available at: <http://academic.oup.com/nar/article/46/D1/D348/4191333> [Accessed March 22, 2019].
- Thompson, A. et al., 2003. Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. *Analytical chemistry*, 75(8), pp.1895–904. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/12713048> [Accessed March 21, 2019].
- Tigano, M. et al., 2015. Elongator-dependent modification of cytoplasmic tRNA^{Lys}_{UUU} is required for mitochondrial function under stress conditions. *Nucleic Acids Research*, 43(17), pp.8368–8380. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/26240381> [Accessed March 22, 2019].
- Tome, J.M., Tippens, N.D. & Lis, J.T., 2018. Single-molecule nascent RNA sequencing identifies regulatory domain architecture at promoters and enhancers. *Nature Genetics*, 50(11), pp.1533–1541. Available at: <http://www.nature.com/articles/s41588-018-0234-5> [Accessed March 22, 2019].
- Tropberger, P. et al., 2013. Regulation of Transcription through Acetylation of H3K122 on the Lateral Surface of the Histone Octamer. *Cell*, 152(4), pp.859–872. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/23415232> [Accessed March 19, 2019].
- Tsuchihara, K. et al., 2009. Massive transcriptional start site analysis of human genes in hypoxia cells. *Nucleic Acids Research*, 37(7), pp.2249–2263. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/19237398> [Accessed March 22, 2019].
- Tsuchiya, E. et al., 2015. RSC Chromatin-Remodeling Complex Is Important for Mitochondrial Function in *Saccharomyces cerevisiae*. *Plos One*, 10(6), p.e0130397.
- Tu, B.P. et al., 2007. Cyclic changes in metabolic state during the life of a yeast cell. *Proceedings of the National Academy of Sciences of the United States of America*, 104(43), pp.16886–91. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/17940006> [Accessed October 3, 2016].
- Tu, B.P. et al., 2005. Logic of the yeast metabolic cycle: temporal compartmentalization of cellular processes. *Science (New York, NY)*, 310(5751), pp.1152–1158. Available at: <http://www.sciencemag.org/cgi/doi/10.1126/science.1120499>.
- Tu, B.P. & McKnight, S.L., 2009. Evidence of carbon monoxide-mediated phase advancement of the yeast metabolic cycle. *Proceedings of the National Academy of Sciences of the United States of America*, 106(34), pp.14293–6. Available at:

- <http://www.ncbi.nlm.nih.gov/pubmed/19706514> [Accessed October 3, 2016].
- Tuck, M.T., Farooqui, J.Z. & Paik, W.K., 1985. Two histone H1-specific protein-lysine N-methyltransferases from *Euglena gracilis*. Purification and characterization. *The Journal of biological chemistry*, 260(11), pp.7114–21. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/3922984> [Accessed March 18, 2019].
- Utle, R.T. et al., 1998. Transcriptional activators direct histone acetyltransferase complexes to nucleosomes. *Nature*, 394(6692), pp.498–502. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/9697775> [Accessed March 19, 2019].
- VanDemark, A.P., Kasten, M.M., Ferris, E., Heroux, A., Hill, C.P., Cairns, B.R., et al., 2007. Autoregulation of the Rsc4 Tandem Bromodomain by Gcn5 Acetylation. *Molecular Cell*, 27(5), pp.817–828. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/17803945> [Accessed May 2, 2017].
- VanDemark, A.P., Kasten, M.M., Ferris, E., Heroux, A., Hill, C.P. & Cairns, B.R., 2007. Autoregulation of the Rsc4 Tandem Bromodomain by Gcn5 Acetylation. *Molecular Cell*, 27(5), pp.817–828. Available at: <http://www.sciencedirect.com/science/article/pii/S1097276507005618> [Accessed June 22, 2017].
- Vaquero, A. et al., 2006. SirT2 is a histone deacetylase with preference for histone H4 Lys 16 during mitosis. *Genes & development*, 20(10), pp.1256–61. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/16648462> [Accessed March 19, 2019].
- Villeneuve, L.M. & Natarajan, R., 2010. The role of epigenetics in the pathology of diabetic complications. *American journal of physiology. Renal physiology*, 299(1), pp.F14-25. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/20462972> [Accessed March 18, 2019].
- Wagner, G.R. & Payne, R.M., 2013. Widespread and Enzyme-independent N^{ϵ} -Acetylation and N^{ϵ} -Succinylation of Proteins in the Chemical Conditions of the Mitochondrial Matrix. *Journal of Biological Chemistry*, 288(40), pp.29036–29045. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/23946487> [Accessed March 18, 2019].
- Wanders, R.J. et al., 2001. Peroxisomal fatty acid alpha- and beta-oxidation in humans: enzymology, peroxisomal metabolite transporters and peroxisomal diseases. *Biochemical Society transactions*, 29(Pt 2), pp.250–67. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/11356164> [Accessed May 29, 2019].
- Wanders, R.J.A. et al., 2010. Peroxisomes, lipid metabolism and lipotoxicity. *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids*, 1801(3), pp.272–280. Available at: <https://linkinghub.elsevier.com/retrieve/pii/S1388198110000028> [Accessed May 29, 2019].
- Wang, W., 2003. The SWI/SNF family of ATP-dependent chromatin remodelers: similar mechanisms for diverse functions. *Current topics in microbiology and immunology*, 274, pp.143–69. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/12596907> [Accessed March 20, 2019].
- Weber, A. et al., 2011. Regulation of NF- κ B activity by competition between RelA acetylation and ubiquitination. *Oncogene*, 31(5), pp.611–623.
- Weinert, B.T. et al., 2014. Acetylation dynamics and stoichiometry in *Saccharomyces cerevisiae*. *Molecular Systems Biology*, 10(1), p.716. Available at:

- <http://www.ncbi.nlm.nih.gov/pubmed/24489116> [Accessed March 18, 2019].
- Wellen, K.E. et al., 2009. ATP-Citrate Lyase Links Cellular Metabolism to Histone Acetylation. *Science*, 324(5930), pp.1076–1080.
- Welsh, D.K. et al., 2004. Bioluminescence Imaging of Individual Fibroblasts Reveals Persistent, Independently Phased Circadian Rhythms of Clock Gene Expression. *Current Biology*, 14(24), pp.2289–2295. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/15620658> [Accessed March 22, 2019].
- Wen, A.Y., Sakamoto, K.M. & Miller, L.S., 2010. The Role of the Transcription Factor CREB in Immune Function. *The Journal of Immunology*, 185(11), pp.6413–6419. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/21084670> [Accessed March 18, 2019].
- Wery, M. et al., 2016. Nonsense-Mediated Decay Restricts LncRNA Levels in Yeast Unless Blocked by Double-Stranded RNA Structure. *Molecular cell*, 61(3), pp.379–392. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/26805575> [Accessed March 22, 2019].
- Whitehouse, I. & Tsukiyama, T., 2006. Antagonistic forces that position nucleosomes in vivo. *Nature Structural & Molecular Biology*, 13(7), pp.633–640. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/16819518> [Accessed March 20, 2019].
- Wijnen, H., Landman, A. & Futcher, B., 2002. The G(1) cyclin Cln3 promotes cell cycle entry via the transcription factor Swi6. *Molecular and cellular biology*, 22(12), pp.4402–18. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/12024050> [Accessed March 22, 2019].
- Winston, F. & Carlson, M., 1992. Yeast SNF/SWI transcriptional activators and the SPT/SIN chromatin connection. *Trends in genetics : TIG*, 8(11), pp.387–91. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/1332230> [Accessed March 20, 2019].
- Wong, C.C., Qian, Y. & Yu, J., 2017. Interplay between epigenetics and metabolism in oncogenesis: Mechanisms and therapeutic approaches. *Oncogene*, 36(24), pp.3359–3374.
- Wu, W.-H. et al., 2005. Swc2 is a widely conserved H2AZ-binding module essential for ATP-dependent histone exchange. *Nature Structural & Molecular Biology*, 12(12), pp.1064–1071. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/16299513> [Accessed March 20, 2019].
- Wyers, F. et al., 2005. Cryptic Pol II Transcripts Are Degraded by a Nuclear Quality Control Pathway Involving a New Poly(A) Polymerase. *Cell*, 121(5), pp.725–737. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/15935759> [Accessed March 22, 2019].
- Xi, Q. et al., 2011. A Poised Chromatin Platform for TGF- β Access to Master Regulators. *Cell*, 147(7), pp.1511–1524. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/22196728> [Accessed March 19, 2019].
- Xiao, M. et al., 2012. Inhibition of α -KG-dependent histone and DNA demethylases by fumarate and succinate that are accumulated in mutations of FH and SDH tumor suppressors. *Genes and Development*, 26(12), pp.1326–1338.
- Xu, W. et al., 2011. Oncometabolite 2-Hydroxyglutarate Is a Competitive Inhibitor of α -Ketoglutarate-Dependent Dioxygenases. *Cancer Cell*, 19(1), pp.17–30. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/21251613> [Accessed March 18, 2019].

- Xu, Z. et al., 2009. Bidirectional promoters generate pervasive transcription in yeast. *Nature*, 457(7232), pp.1033–1037. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/19169243> [Accessed March 22, 2019].
- Yadon, A.N. et al., 2010. Chromatin Remodeling around Nucleosome-Free Regions Leads to Repression of Noncoding RNA Transcription. *Molecular and Cellular Biology*, 30(21), pp.5110–5122. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/20805356> [Accessed March 20, 2019].
- Yang, H. et al., 2012. IDH1 and IDH2 mutations in tumorigenesis: Mechanistic insights and clinical perspectives. *Clinical Cancer Research*, 18(20), pp.5562–5571.
- Yang, X.-J., 2004. Lysine acetylation and the bromodomain: a new partnership for signaling. *BioEssays*, 26(10), pp.1076–1087. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/15382140> [Accessed March 20, 2019].
- Yoon, M.J. et al., 2015. SIRT1-Mediated eNAMPT Secretion from Adipose Tissue Regulates Hypothalamic NAD⁺ and Function in Mice. *Cell Metabolism*, 21(5), pp.706–717. Available at: <https://www.sciencedirect.com/science/article/pii/S1550413115001576> [Accessed March 19, 2019].
- Yoon, S. et al., 2003. Recruitment of SWI/SNF by Gcn4p does not require Snf2p or Gcn5p but depends strongly on SWI/SNF integrity, SRB mediator, and SAGA. *Molecular and cellular biology*, 23(23), pp.8829–45. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/14612422> [Accessed March 20, 2019].
- Yoshino, J. et al., 2011. Nicotinamide Mononucleotide, a Key NAD⁺ Intermediate, Treats the Pathophysiology of Diet- and Age-Induced Diabetes in Mice. *Cell Metabolism*, 14(4), pp.528–536. Available at: <https://www.sciencedirect.com/science/article/pii/S1550413111003469> [Accessed March 19, 2019].
- Yu, F. et al., 2015. The yeast chromatin remodeler Rsc1-RSC complex is required for transcriptional activation of autophagy-related genes and inhibition of the TORC1 pathway in response to nitrogen starvation. *Biochemical and Biophysical Research Communications*, 464(4), pp.1248–1253. Available at: <http://dx.doi.org/10.1016/j.bbrc.2015.07.114>.
- Yu, H. et al., 2012. Acetylation of sphingosine kinase 1 regulates cell growth and cell-cycle progression. *Biochemical and Biophysical Research Communications*, 417(4), pp.1242–1247. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/22227192> [Accessed March 18, 2019].
- Zaidi, N., Swinnen, J. V. & Smans, K., 2012. ATP-Citrate Lyase: A Key Player in Cancer Metabolism. *Cancer Research*, 72(15), pp.3709–3714. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/22787121> [Accessed March 18, 2019].
- Zajac, P. et al., 2013. Base Preferences in Non-Templated Nucleotide Incorporation by MMLV-Derived Reverse Transcriptases L. Menéndez-Arias, ed. *PLoS ONE*, 8(12), p.e85270. Available at: <https://dx.plos.org/10.1371/journal.pone.0085270> [Accessed March 22, 2019].
- Zhang, C. & Darnell, R.B., 2011. Mapping in vivo protein-RNA interactions at single-nucleotide resolution from HITS-CLIP data. *Nature Biotechnology*, 29(7), pp.607–614. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/21633356> [Accessed March 22, 2019].

- Zhang, H., Roberts, D.N. & Cairns, B.R., 2005a. Genome-Wide Dynamics of Htz1, a Histone H2A Variant that Poises Repressed/Basal Promoters for Activation through Histone Loss. *Cell*, 123(2), pp.219–231. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/16239141> [Accessed March 20, 2019].
- Zhang, H., Roberts, D.N. & Cairns, B.R., 2005b. Genome-Wide Dynamics of Htz1, a Histone H2A Variant that Poises Repressed/Basal Promoters for Activation through Histone Loss. *Cell*, 123(2), pp.219–231. Available at: <https://www.sciencedirect.com/science/article/pii/S0092867405008731> [Accessed March 22, 2019].
- Zhang, Q. et al., 2010. Biochemical profiling of histone binding selectivity of the yeast bromodomain family. *PLoS ONE*, 5(1), pp.1–10.
- Zhang, T. et al., 2009. Enzymes in the NAD⁺ Salvage Pathway Regulate SIRT1 Activity at Target Gene Promoters. *Journal of Biological Chemistry*, 284(30), pp.20408–20417. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/19478080> [Accessed March 18, 2019].
- Zhang, X. et al., 2018. Structure and functional interactions of INO80 actin/Arp module. *Journal of Molecular Cell Biology*. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/30388237> [Accessed March 20, 2019].
- Zhao, G. et al., 2016. Cyclin-Dependent Kinase Co-Ordinates Carbohydrate Metabolism and Cell Cycle in *S. cerevisiae*. *Molecular Cell*, 62(4), pp.546–557. Available at: <http://dx.doi.org/10.1016/j.molcel.2016.04.026>.
- Zhao, Y. & Garcia, B.A., 2015. Comprehensive Catalog of Currently Documented Histone Modifications. *Cold Spring Harbor perspectives in biology*, 7(9), p.a025064. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/26330523> [Accessed March 18, 2019].