Mathematical Modelling of the Yeast Metabolic Cycle



Alberto Merchante González

Corpus Christi College

University of Oxford

Thesis submitted at the University of Oxford in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Trinity Term, 2018

Mathematical Modelling of the Yeast Metabolic Cycle

Thesis submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at the University of Oxford

Alberto Merchante González, Corpus Christi College, Trinity Term, 2018

Abstract

The Yeast Metabolic Cycle (YMC) is an ultradian rhythm whose precise regulatory mechanism remains elusive. The YMC is a unique system that combines respiratory bursts with oscillations in both transcript levels and hundreds of metabolites in a tightly-controlled manner, creating an effective single-cell behavior that can be studied using sequencing techniques that require large quantities of cells. In this DPhil thesis I present a quantitative mathematical model of the YMC that is able to both describe the yeast synchronized metabolic activity and predict the outcome of several validating experiments.

Histone Modifications (HMs) are intimately related to gene activation and epigenetic changes regulated through chromatin organization. mRNA degradation is related to some HMs accumulation levels, and both are actively involved in the YMC regulatory mechanism. By comparing the RNA and HMs dynamics across the YMC, I have explored the functions of some HMs like H3K4me3 or H3 early-tail lysine acetylations in the yeast metabolism, discovering that H3K18ac abundance is incompatible with mRNA stability.

By clustering the yeast genome according to its expression dynamics during the YMC, I have revealed new insights into the way metabolic processes are compartmentalized at the different phases of the cycle and elucidated potential transcription factors like Met4p, Adr1p, or Thi2p, that may act as regulators of the different clusters, constituting a first attempt of unveiling the YMC regulatory mechanism.

I have employed these clusters as the cornerstone of a mathematical model describing RNA and protein dynamics at an isolated yeast cell using ordinary differential equations, concluding that a transcriptional oscillator guiding the yeast activity is more plausible that a fully metabolic oscillator.

Finally, I have combined a collection of three-component, robust, transcriptional oscillators entrained by a mean-field Hamiltonian to explain the coordination mechanisms underlying the origin and maintenance of the YMC. The resulting mathematical model is able to both replicate the YMC features described by the literature and predict the system behaviour when experimental conditions are changed or perturbations are introduced. In particular, our model explains the shape of the dissolved oxygen trace observed in the chemostat, replicates the spontaneous yeast synchronization obtained after a glucose starvation phase, and predicts the effect that carbon-source perturbations will have on the YMC.

Omnia munda mundis
A. Manzoni. I promessi sposi, 1827

All models are wrong, but some are useful
George Box. Science and Statistics, 1976

To my parents, my sister, and my granny.

Acknowledgements

I would like to thank my supervisors, Dr Andrew Angel and Prof Jane Mellor, for their inputs, feedback, helpful suggestions, and support during this three years of graduate study. Jack Feltham and Shidong Xi have performed the wet lab experiments associated with this thesis. I would also like to thank the rest of the Mellor and Angel labs for providing a friendly atmosphere in the department, continuous support, and valuable scientific discussions.

I would like to thank Prof Sonia Contera for her incredible support during this thesis, providing me with very helpful guidance and inputs, as well as Prof Bela Novak and Prof Elspeth Garman, for the time and effort the three of them spent in guiding this doctoral work. Corpus Christi College has provided me excellent support during my stay at Oxford, being an excellent social and cultural hub that has always been helpful with my studies. Finally, the Royal Society has made this project possible thanks to its generous studentship.

Contents

1 Introduction **1**

1.1 Oscillations in biology: a ubiquitous phenomenon 2
1.2 Periodic metabolic rhythms: origin and regulation 2
1.3 Synchronized metabolic activity across different species 4
1.4 The yeast metabolic cycle 5
1.4.1 Acetate and other metabolites coordinate the YMC 8
1.4.2 Environment composition alters the YMC 11
1.4.3 The YMC is divided into three phases 11
1.4.4 The YMC and the CDC 16
1.4.5 Histone modifications and the YMC 17
1.4.6 Single cells in unsynchronized cultures have an analogous cycle 17
1.4.7 YMC and yeast mutants 19
1.4.8 Mathematical models of yeast 19
1.4.9 A unique system? 20
1.5 Yeast respiration during the YMC 20
1.5.1 Is oxygen coordinating the YMC? 22
1.5.2 Cellular oxygen consumption calculation 23
1.6 Thesis aims 31
1.6.1 Interplay of histone modifications, transcription, and transcript degradation in the regulation of the cycle 31
1.6.2 Investigate methods of clustering the data to identify classes of genes that could be co-regulated 32
1.6.3 Is a transcriptional oscillator the basic control mechanism for the cycling of individual yeast cells? 32
1.6.4 Investigate how the cycles of individual cells are linked to form the observed synchronised cycling behaviour 32
1.6.5 Overview 34

2 Role of Histone Modifications and mRNA Degradation in the YMC 35
2.1 Are HMs involved in YMC regulation? . . . 37
2.1.1 Phenelzine and the YMC . . . 38
2.2 RNA and HMs are correlated across the YMC . . . 41
2.2.1 ChIP-seq and RNA-seq data . . . 41
2.2.2 Methods . . . 42
2.2.3 Results . . . 43
2.2.4 Conclusion . . . 44
2.3 mRNA degradation: are genes regulated at the point of synthesis or at the point of turnover? . . . 46
2.4 mRNA degradation calculation . . . 48
2.4.1 Methods . . . 49
2.4.2 Results . . . 51
2.4.3 Conclusion . . . 57
2.5 mRNA degradation is related to some HMs accumulation levels . . . 57
2.6 Conclusions . . . 61
2.6.1 H3K4me$_3$ cycles weakly and accumulates after the RNA . . . 62
2.6.2 H3 tail lysine acetylations mirror RNA cycling . . . 62
2.6.3 H3K18ac abundance is incompatible with mRNA stability . . . 63

3 Gene Clustering Algorithms 64
3.1 Data to be clustered . . . 65
3.1.1 RNA expression . . . 66
3.1.2 HM accumulation . . . 67
3.1.3 RNA polymerase density . . . 67
3.2 Clustering algorithms . . . 68
3.2.1 Expression peak clustering . . . 69
3.2.2 Oscillatory pattern clustering . . . 70
3.2.3 k-means clustering . . . 77
3.2.4 Other clustering methods . . . 77
3.3 Optimal number of clusters . . . 77
3.3.1 Bayesian Information Criterion . . . 78
3.3.2 Akaike Information Criterion . . . 79
3.3.3 Silhouette . . . 79
3.3.4 ELBOW . . . 81
3.3.5 Comparison . . . 81
3.4 Measures of clustering similarity . . . 82
3.4.1 Cluster comparison criteria . . . 83

3.4.2 A novel cluster dissimilarity measure . . . 84
3.4.3 Index application to genomic data . . . 86
3.5 Clustering including multiple datasets . . . 88
3.5.1 Multiple dataset clustering technique . . . 88
3.5.2 Exhaustive clustering results . . . 91
3.5.3 Non-exhaustive clustering results . . . 101
3.6 Conclusion . . . 106

4 A Single-cell Transcriptional Oscillator 109
4.1 RNA and protein dynamics define the YMC . . . 110
4.2 Ordinary differential equations form the basis of the model . . . 112
4.3 Regulation mechanisms and feedback loops . . . 114
4.3.1 Positive regulation mechanisms . . . 115
4.3.2 Negative regulation mechanisms . . . 115
4.3.3 Different functional forms for the regulation mechanisms . . . 116
4.3.4 Combination of interactions to form loops . . . 117
4.3.5 Computational algorithms for numerical solutions . . . 119
4.4 RNA and protein equations . . . 122
4.4.1 RNA synthesis . . . 122
4.4.2 RNA degradation . . . 124
4.4.3 Protein synthesis . . . 125
4.4.4 Protein degradation . . . 125
4.5 Biochemical estimation of constants . . . 127
4.5.1 Synthesis rates . . . 127
4.5.2 mRNA degradation rates . . . 129
4.6 Numerical simulations to obtain the remaining constants . . . 129
4.6.1 Matching phase lengths with their experimental values . . . 130
4.6.2 Metagenes shape adjustment . . . 132
4.6.3 Other constants . . . 133
4.7 Parameter elasticity and range . . . 134
4.7.1 Parameter elasticity . . . 135
4.7.2 Numerical values of the constants in the transcriptional oscillator 136
4.8 A five-component oscillator . . . 140
4.8.1 Comparison with the model described in Section 4.5 . . . 142
4.9 Conclusion: The YMC can be reproduced by a robust three-phase transcriptional oscillator . . . 142

5 Intercellular Synchronisation Model **145**
5.1 Cells are synchronized during the YMC . . . 146
5.1.1 k_α measurement . . . 148
5.1.2 Multiple synchronization points during the YMC? . . . 149
5.2 Initial perturbation . . . 150
5.3 Noise . . . 151
5.4 Cells are driven into OX via a carbon source . . . 155
5.4.1 A threshold for acetate spiking . . . 157
5.4.2 Cells are not driven into OX phase merely by a decrease in the media O_2 concentration . . . 160
5.5 O_2 consumption is identified with OX protein concentration . . . 161
5.6 Different period strains will not coordinate their oscillations . . . 164
5.6.1 dO_2 trace prediction . . . 166
5.6.2 Discussion and perspectives . . . 170
5.7 Conclusion . . . 171

6 Conclusion **173**
6.1 Proposed future work . . . 175

A Appendix **178**
A.1 Data description . . . 178
A.1.1 RNA-seq and Microarray . . . 179
A.1.2 ChIP-seq . . . 181
A.1.3 NET-seq . . . 182
A.2 Data quality and possible error sources . . . 182
A.3 Information condensation . . . 183
A.3.1 From base pairs to genes: appropriate size of genomic windows 184
A.3.2 Comparison with Boeke 2014 windows . . . 185
A.4 Metagene construction . . . 186
A.5 Proof of theorems in Subsection 3.4.2 . . . 189
A.6 Computational details . . . 190
A.6.1 Code in Subsection 4.3.5 . . . 190
A.6.2 Details of the simulated annealing algorithm in Subsection 4.6.1 191
A.7 Proteomics . . . 192
A.8 Iron-Sulfur proteins, Fe^{2+} ion, and $xrn1\Delta$ . . . 197
A.8.1 Iron proteins . . . 197
A.8.2 Why Fe^{2+} and not SO_4^{2-}? . . . 197
A.8.3 Iron genes analysis . . . 198

List of Figures

1.1 Yeast respiratory oscillations 6
1.2 Yeast cultures synchronized respiration is matched by tightly-controlled oscillations in transcript levels for over 60% of their genome 7
1.3 Experimental setup scheme of the chemostat 9
1.4 Hundreds of yeast metabolites oscillate in the intercellular media during the YMC 10
1.5 Glucose concentration alters the YMC period 12
1.6 Graphical representation of the YMC and its main cellular processes . 13
1.7 Identification of dissolved oxygen concentration with the three YMC phases 14
1.8 Chemostat dissolved oxygen traces of wild type (WT) yeast and histone point mutants undergoing the YMC 18
1.9 Transient administration of argon has no apparent effect on metabolic cycling 23
1.10 Dissolved oxygen concentration in the chemostat & Regular chemostat oscillations 25
1.11 Enlarged section of six respiratory oscillations originally shown in Fig. 1.10.b & Hamiltonian plot of the oscillations 28
1.12 Dissolved oxygen and cellular oxygen consumption YMC curves . . . 30

2.1 RNA and H3K18ac dynamics during the YMC 39
2.2 H3K4me$_3$, H3K14ac & H3K9ac accumulation dynamics during the YMC 40
2.3 Histograms showing the correlation between RNA and H3K4me$_3$, H3K9ac, & H3K56ac 43
2.4 Correlations distribution between RNA and H3K4me$_3$ & H3K18ac dynamics for the 5770 yeast genes and three different time shifts . . . 45
2.5 Histogram of mRNA half-lives distribution in the Young 1998 experiment 47
2.6 YEASTRACT TF influence on gene expression for Gcr1p & Hap2p . 52

2.7 Regulatory tree for the 300 genes with smallest c_V in the yeast genome 53
2.8 Scatter plot of H3K18ac accumulation peak compared with mRNA degradation time . . . 59
2.9 Distribution of HM time delays with respect to degradation rates . . 60

3.1 The sixteen most frequent RNA expression patterns . . . 72
3.2 Four main clusters for RNA expression & H3K18ac accumulation dynamics . . . 73
3.3 Fourier transform waves of Clusters 1 & 3 in Figure 3.2 metagenes . . 75
3.4 AIC and BIC values for the Li *et. al.* dataset under the k-means algorithm . . . 80
3.5 k-means algorithm: Inter-cluster distance evolution with k . . . 87
3.6 Eight clusters of genes obtained by combined k-means algorithm across time in mRNA, nascent transcription, and H3K9ac accumulation levels during the YMC . . . 93
3.7 Raw sequencing data for gene YLN122W ($DBP2$) . . . 95
3.8 Raw sequencing data for gene YLR409C ($UTP21$) . . . 96
3.9 Enlarged view of the transitory signals in H3K9ac accumulation observed for genes $DBP2$ and $UTP21$ in the two previous Figures. . . . 97
3.10 GO tree stemming from the genes included in Cluster 3 of Fig. 3.6 . . 98
3.11 Metagenes associated with different HM data for genes belonging to Cluster 3 in Figure 3.6 . . . 99
3.12 Multiple-omics non-exhaustive clustering results . . . 103
3.13 GO tree derived from Cluster 1 (OX) in the oscillation-shape clustering algorithm . . . 104

4.1 Transition from experimentally measured metagenes to YMC ODEs models . . . 112
4.2 A two-component negative feedback forming regulation model involving two elements of the control circuit . . . 118
4.3 Different simulations of the two components model . . . 119
4.4 Negative and positive ODE regulation mechanisms . . . 121
4.5 A three-component negative feedback loop . . . 123
4.6 Gaußian Martingale walk algorithm simulation results . . . 132
4.7 Transcriptional oscillator changes in period and amplitude when two parameters are changed simultaneously . . . 137
4.8 SimBiology parameter swap robustness simulation . . . 138
4.9 A five-component YMC oscillator . . . 141
4.10 Final form of the YMC transcriptional oscillator . . . 143

5.1 Chemostat dO_2 oscillations & Start of the YMC 147
5.2 Overall system phase distributions at different YMC stages 151
5.3 Multicellular system simulations including stochastic gene expression and acetate spiking 154
5.4 Acetate spiking during OX phase does not induce a new OX phase & two consecutive acetate spikes induce two consecutive OX phases . . 156
5.5 Predicted dO_2 trace when different acetate concentrations are poured into the YMC Reductive phase 158
5.6 Predicted RNA_{OX} dynamics when different sustained acetate concentrations are induced in the media 159
5.7 Dissolved oxygen trace and protein concentration model with OX protein levels matching cellular oxygen consumption 163
5.8 Single yeast cell system model & Prediction of a co-culture system . . 167
5.9 dO_2 trace in a yeast co-culture - Prediction & Corresponding Fourier transformation via FFT 169

A.1 Sampling timepoints of the Boeke 2014 sequencing data 179
A.2 Sampling timepoints of the Li 2006 sequencing data 180
A.3 12 NET-seq sampling points 182
A.4 H3K14ac, $H3K36me_3$, and H3K9ac metagene profiles 187
A.5 RNA, $H3K4me_3$, and H3K18ac metagene profiles 188
A.6 $TEF1$ & Tef1p dynamics during the YMC 194
A.7 RNA and protein peaks distribution in time in the YMC 196
A.8 WT & $xrn1\Delta$ YMC dynamics for genes associated with the iron ions Fe^{2+} and Fe^{3+} 199

List of Tables

1.1 Enriched GO terms associated with each of the three YMC phases . . 15

2.1 Histone modifications deposition enzymes and position on the genes . 37
2.2 mRNA degradation data: comparison of Young *et al.* and Farina *et al.* experiments 48
2.3 Most significant GO terms associated with more stable mRNA genes 54
2.4 Average mRNA half-lives for the different YMC clusters 55
2.5 Average deposition time delays for different HMs 61

3.1 Different YMC experimental conditions comparison 68
3.2 Optimal k-value for the k-means algorithm implementation according to the BIC, AIC, Silhouette, and ELBOW number of clusters criteria 82
3.3 TFs and GO terms associated with OX multilevel clusters 100
3.4 Gene Ontology terms derived from the oscillation-shape clustering algorithm 105
3.5 Transcription factors derived from oscillation-shape clustering algorithm 105

4.1 Numerical values of the YMC model constants 139

5.1 Differences between CEN.PK and IFO 0233 strains YMC period, media composition, and culture parameters 171

A.1 Genes whose phenotype is associated with iron ions Fe^{2+} or Fe^{3+} . . . 199

Glossary

Below are listed the most common abbreviations used in this thesis.

AIC Akaike Information Criterion

BIC Bayesian Information Criterion

bp Base Pair

CDC Cell Division Cycle

ChIP-seq Chromatin immunoprecipitation (ChIP) combined with massively parallel DNA sequencing to identify the binding sites of DNA-associated proteins

CUR Circadian and Ultradian Rhythms

DNA DeoxyriboNucleic Acid

$\mathbf{dO_2}$ Dissolved Oxygen (O_2 molecules) in the media

DoFs Degrees of Freedom

FDP Fructose 1,6-DiPhosphate

FDR False Discovery Rate

FISH Fluorescent In Situ Hybridization

GO Gene Ontology

HM Histone Modification

H$\boldsymbol{n_1}$K$\boldsymbol{n_2}$ac Covalent modification consisting on the addition of an acetyl group (-$COCH_3$) on Histone Hn_1 at the histone tail lysine (K) n_2 position

$\mathbf{Hn_1Kn_2me_3}$ $Hn_1Kn_2me_3$: Idem, but adding three methyl groups (-CH_3) instead

HOC High Oxygen Consumption YMC phase

kb KiloBase (1000 bp)

LHS Left Hand Side of an equation

LOC Low Oxygen Consumption YMC phase

MAO MonoAmine Oxidase

mRNA Messenger RNA

NADH Nicotinamide adenine dinucleotide reduced form. NADPH: Its phosphate

NDR Nucleosome Depleted Region

NET-seq Native Elongating Transcript sequencing

ODE Ordinary Differential Equation

OX Oxidative phase of the YMC

PDE Partial Differential Equation

PCR Polymerase Chain Reaction

Pol II RNA polymerase II multiprotein complex

PZ Phenelzine

Q_{O_2} Cellular O_2 consumption

RB Reductive/Building phase of the YMC

RC Reductive/Charging phase of the YMC

RHS Right Hand Side of an equation

RNA RiboNucleic Acid

RNAP II RNA polymerase II

RNA-seq RNA sequencing

RPN Reduced Pyridine nucleotide

TF Transcription Factor

TSS Transcription Start Site

TTS Transcription Termination Site, sometimes referred to in the literature as Transcription End Site (TES)

WT Wild Type

YMC Yeast Metabolic Cycle

The word *yeast*, when used on this text by its own, refers to the budding yeast *Saccharomyces cerevisiae*. When we talk about *fission yeast* we refer to the *Schizosaccharomyces pombe*.

Genes naming convention: we follow the *S. cerevisiae* species-specific convention recommended by SGD[1], on which:

CDC28 refers to the gene

Cdc28p refers to the protein coded by that gene

cdc28Δ refers to the mutant strain on which the gene has been deleted

This lab used the CEN.PK113-7D *S. cerevisiae* strain, as will be discussed in the main text.

[1] `https://sites.google.com/view/yeastgenome-help/community-help/nomenclature-conventions`

Chapter 1

Introduction

The Yeast Metabolic Cycle (YMC) is an ultradian rhythm whose precise regulatory mechanism remains elusive, and no satisfactory mechanism has been proposed for its description. The aim of this DPhil thesis is to apply mathematical modelling methods to investigate the processes that control the cycle, in order to describe the yeast synchronized metabolic activity and make predictions that can be experimentally validated.

For reaching this aim, I use bioinformatic tools (Chapters 2 and 3) that provide a novel approach to the YMC by describing the dynamics of both mRNA and some histone modifications (Chapter 2) and clustering yeast genes according to their temporal dynamics (Chapter 3).

In the second part of this thesis (Chapters 4 and 5), I use such results to build a transcriptional oscillator that reproduces the observed RNA and protein dynamics for an isolated yeast cell (Chapter 4) and then I combine a number of these oscillators to make a full description of the YMC as a coordinated activity of a population of yeast (Chapter 5).

This introductory Chapter starts with a brief review on biological fluctuations in general (Section 1.1) and metabolic rhythms in particular (Section 1.2), discusses examples of synchronized metabolic activity in different species other than yeasts (Section 1.3) and focuses on an exhaustive description of the YMC dynamics and the state of the art knowledge about this system (Section 1.4). The Chapter continues with a preliminary mathematical analysis of yeast respiration during the YMC (Section 1.5), and finishes with a description of the precise aims and questions posed

in this doctoral work (Section 1.6).

1.1 Oscillations in biology: a ubiquitous phenomenon

Cyclic mechanisms are present at every level in the organization of living organisms [1]; it could even be said that the cellular steady-state operation is rare [2] and periodic behaviour is indeed prevalent in nature [3]. Chemical oscillators [4, 5] and entrainment to environmental conditions [6] are the two bases on which biological oscillators are set and maintained. This is also the case in some chemical reactions: the Belousov-Zhabotinsky reaction showed in 1964 that chaotic chemical oscillators with as few as four inorganic components are possible [7, 8].

Fluctuations are naturally modelled by differential equations and feedback loops [9, 10], an approach that I follow in Chapters 4 and 5. Biological time has been described as a topological space of its own [11] with toroidal characteristic [12] on which the particularities of phase-dependent metabolic oscillators are more easily described [13, 14].

Oscillatory dynamics have evident advantages, since they protect enzymes and possibly cells against toxic substances [15]; the oscillatory metabolism is known to depend on cell density [16], suggesting that organisms in the same biological system interact with each other. However, the field of biological oscillations in general is beyond the scope of this thesis, which is only concerned with yeast metabolic oscillations, and therefore they are only mentioned *en passant* in this chapter.

1.2 Periodic metabolic rhythms: origin and regulation

Circadian and ultradian rhythms (CURs) are present in virtually every living organism [17,18]. CURs are set and maintained by biochemical oscillators whose regularity acts as a time keeper [19, 20]. The most common metabolic rhythm is the circadian clock [21], an oscillation observed in some genes' expression levels of an organism with a period of approximately 24 hours which allows for internal synchronizing of day/night cycles. Other periods are also possible: for example, the marine annelid *Platynereis dumerilii* possesses an endogenous ∼28 days long circalunar clock [22], and many organisms in coastal or estuarine habitats like the crabs *Uca pugnax* or

Carcinus maenas possess an approximately 12.4 hours long circatidal clock not entrained to their circadian rhythm [23, 24]. Several CURs can coexist and regulate different metabolic aspects of the same organism [25–27].

CURs are usually entrained by periodicity in the corresponding environmental conditions [28, 29] and are thought to be originated by metabolic cycles [30]. Unlike the YMC, the period of these rhythms is independent of nutrient composition or changes in environmental variables unrelated to the entrainment factor [31]. CURs are evolutionarily conserved mechanisms linked to primitive mitochondria, suggesting a common ancestral origin across the different forms of life [32, 33] that has been proposed as a basic universal necessity for fitness by coordinating intracellular metabolism [2].

CURs' precise regulatory mechanisms remain elusive, but it has been proposed that they are regulated by protein biosynthesis [34] and post-transcriptional modifications of DNA and histone tails [35]. CURs rely on feedback loops of the underlying metabolic oscillator [36–38], a process that has been widely modelled [39,40] and that is described in Chapter 4.

Metabolic cycles are characterized by metabolite oscillations [1,41], and since the 1970s there has been a sustained interest in understanding their regulation and describing them via mathematical pathway models [42,43]. In a metabolic pathway, the fluxes and concentrations generally depend on the activities and kinetic properties of all of the involved enzymes simultaneously [44, 45]. The antioxidant peroxiredoxin proteins have been identified as conserved markers for 24-hour rhythms [46]. Circadian rhythms entail oscillations in gene products [36] and are transcriptionally regulated [47], with $PER2$ protein having been shown to be the chief *Drosophila melanogaster* circadian clock regulator [48, 49].

Yeast possesses both a circadian clock [50, 51] and an ultradian clock which is matched by its glycolytic oscillations [52] and has been modelled via phase shifts [53], resonances [54], and enzymatic activity [55]. Yeast metabolism and respiration are affected by both visible light [56] and temperature variations [57–60]. Yeast metabolic rhythms and intercellular synchrony mechanisms are discussed at wider length in Sections 1.4 and 5.1. Finally, it is important to note that rounds of new transcription are not a requirement for a clock, since red blood cells have a circadian clock [61, 62] without transcription [63, 64].

1.3 Synchronized metabolic activity across different species

Yeasts are not the only organisms to synchronize their metabolic activity in the absence of an external driving fluctuation, known as an entrainment element, such as sunlight or tidal rhythms. In 1969, it was discovered that cells from the same tissue can signal their positional information via a particular ultradian rhythm in amphibians [65]. Soon afterwards, synchronous cultures of *Schizosaccharomyces pombe* [66], *Crithidia fasciculata* [67], *Acanthamoeba castellanii* [68–70], *Dictyostelium discoideum* [71] and many other lower eukaryotes were shown to coordinate their metabolic rhythms within the cultures.

Synchronized metabolic oscillations often imply coordination in both respiration and protein production, as it is the case for example in *E. coli* [72], *Schizosaccharomyces pombe* [70], *Klebsiella aerogenes* [73], *A. castellanii* [74][1] or *Candida utilis* [75]. Similarly, these metabolic rhythms are associated with an ordered and sequential expression of most of the organism's genome, not just in yeast [76, 77] but also in more complex, multicellular, organisms, like *Arabidopsis thaliana* [78], *Drosophila melanogaster* [79, 80], or *Mus musculus* [81–84].

More recently, collective cooperative metabolic oscillations within biofilms in *Bacillus subtilis* have been discovered and modelled [85], a process regarded as a two-dimensional bacterial metabolic cycle in which the distribution of the scarce nutrients in the media, via metabolic rhythms, is beneficial for the whole community, thus leading the organisms to express their genomes in a coordinated and periodic manner.

Synchronized metabolic activity has been modelled using molecular regulatory networks [86, 87], protein-activity gradients [88–91], and in general a wide variety of computational methods [92, 93].

[1] Fry and Lloyd [74] managed to change the *A. castellanii* cell division time from 8 hours to 16 hours while maintaining the period of the respiratory oscillations, showing that both processes can be uncoupled.

1.4 The yeast metabolic cycle

The first part of this Section consists in a literature review of budding yeast synchronized metabolic oscillations, since they were first observed in the 60s to the state of the art knowledge, explaining the most relevant processes associated with the YMC and its phases. The second part of the Section contains a collection of facts about the YMC that provide some relevant background knowledge about the core principles and key features of the YMC, including the relationship between the YMC and other synchronous culture cycles: glycolytic oscillations and the Cell Division Cycle (CDC).

Under specific continuous, nutrient-limited conditions, *S. cerevisiae* colonies exhibit robust, highly periodic cycles in the form of respiratory bursts [3, 94, 95], as shown in Figure 1.1. These oscillations do not arise spontaneously: to obtain them, a batch phase on which cells are grown on a high-glucose media to high density is followed by a starvation phase on which the yeast consume all of the glucose available in the media, therefore reducing their metabolic activity (and hence oxygen consumption) to a minimum. After such a starvation phase, adding glucose initiates the respiratory cycles. Respiratory bursts are experimentally observed as oscillations in the dissolved oxygen concentration in the media (dO_2). Once the oscillations are set, if the system conditions are kept constant, the oscillations are maintained for indefinite periods of time.

In 1964, oscillations of glycolytic intermediates fructose 1,6-diphosphate (FDP) and reduced pyridine nucleotide (RPN) were first observed in yeast cell suspensions [94, 96–100], a process that has been modelled using a perturbation term [101], two-state models [102], or diffusion models [103]. A few years later, oscillations in the yeast culture media dissolved oxygen [104] and several metabolites [79, 105] concentrations were observed. Finally, respiration was shown to be related to metabolic oscillations on ethanol-grown cultures [106], and living cells were kept indefinitely in a well-defined oscillatory state when starved cells, glucose, and cyanide were pumped into the system [107].

In 2005, Tu *et. al.* [3] discovered that synchronized yeast cultures' respiration is matched by tightly-controlled oscillations in transcript levels for over 60% of the genome, as shown in Figure 1.2, coining the term Yeast Metabolic Cycle. They realized the extraordinary utility of the system to learn about genetic regulation because synchronized cultures entail an effective single-cell behaviour. This feature allowed the introduction of sequencing techniques (Microarrays in 2005, RNA-seq after 2008)

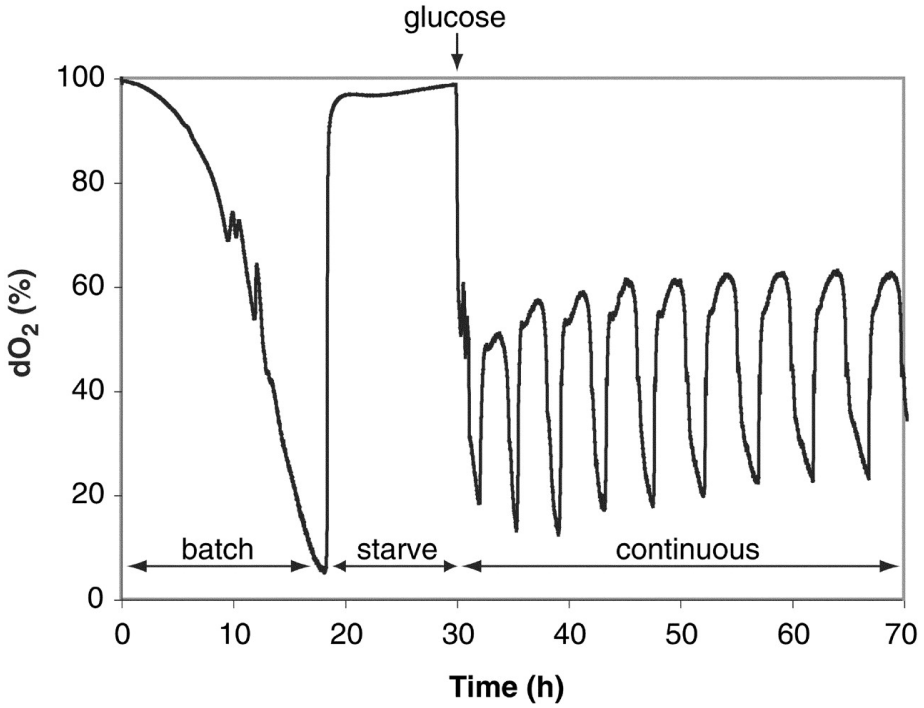


Figure 1.1: Yeast respiratory oscillations. y-axis: dissolved oxygen concentration (% saturation) in the media (dO_2). x-axis: time. The first 18 hours are the batch phase, on which cells are grown to high density on a high-glucose media. From 18 hours to 30 hours is the starvation phase, on which the yeast consume all of the glucose available in the media. After 30 hours, the continuous phase appears, in which the YMC is set and maintained. Synchronized respiratory oscillations are evidenced by the periodic oscillation of dO_2 in this region. Taken from [3].

requiring large quantities of cells to accurately infer the dynamics of individual genes.

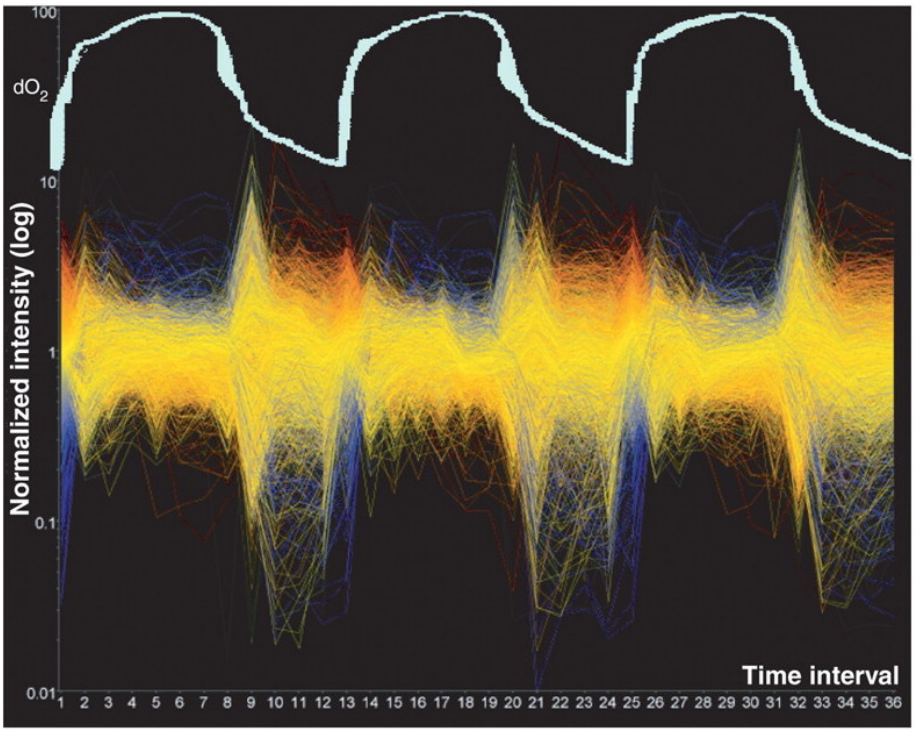


Figure 1.2: Yeast cultures synchronized respiration is matched by tightly-controlled oscillations in transcript levels for over 60% of the genome. x-axis: time; 36 time points covering three yeast respiratory oscillations with 12 samples per cycle, spaced approximately 25 minutes. y-axis: Top: dissolved oxygen present in the culture (dO_2), expressed as a percent of saturation level (pale blue curve). Bottom: Microarray expression profiles of ∼6209 unique expressed genes. Spectral colouring scheme of genes has its basis in relative expression at the start of reductive phase (red indicates higher; blue, lower). Synchronized respiratory and metabolic oscillations are evidenced by the co-evolution of dO_2 and gene expression across the cycle. Taken from [3].

A similar experiment performed by Li and Klevecz in 2006 [108] on a different yeast strain retrieved similarly shaped, faster[2] metabolic oscillations. This result expands their previous observation of oscillations in transcription during the YMC

[2]This strain exhibited oscillations with a period of 40 minutes, as opposed to the 3.5 hours in Tu *et. al.* [3].

linked to both DNA replication and the cell cycle [109]. While both experiments retrieved metabolic respiratory oscillations in yeast, they were remarkably different in their methods and techniques, a discussion of which I defer to Section 5.6 (see Table 5.1). Their analysis techniques were also diverse: while Tu was chiefly concerned with regularity in gene expression, Li and Klevecz measured the differential gene expression induced by the addition of a drug (phenelzine) into the system, an experiment that is discussed in Section 2.1.

The yeast culture takes place in a chemostat [110], an experimental device used for a wide variety of cultures into which media and nutrients are delivered at a constant rate, while maintaining the total volume unchanged. As Figure 1.3 explains, the chemostat contains a media reservoir used to feed the cells, and the media is introduced at a constant rate into the chemostat vessel via the media tube. The vessel is aerated at a constant rate (5 L/min) using an air pump and a magnetic stirrer (speed 600 rpm), which makes a turbulent fluid inside the vessel. Acidity is kept constant at a pH=3.5 using buffer and a pH probe, and dissolved oxygen (dO_2) is measured using a second probe. In the vessel, yeast cells are grown to a high density [$OD_{600} \sim 8.5$] and temperature is kept constant at T=30°C. Finally, a collecting receptacle receives the chemostat excess volume via an effluent tube, ensuring that the vessel volume is kept constant. The following subsections contain a review of the YMC literature that explains several properties of this singular system.

1.4.1 Acetate and other metabolites coordinate the YMC

During the YMC, the intracellular concentration of hundreds of metabolites fluctuates, closely replicating these respiratory bursts [112,113], as we can see in Figure 1.4. Nicotinamide adenine dinucleotide phosphate reduced (NADPH) [114] and adenosine triphosphate (ATP) [115] oscillations have been identified as key candidates to explain the YMC cellular synchronization dynamics. Strong variations in acetyl-CoA concentration [116] act as a gauge of cellular metabolic state [117], and induce cell growth by promoting histone acetylation [118]. In fact, when low doses of acetate, ethanol, acetaldehyde, or lactate are added to the system, a sudden drop in the media dissolved oxygen concentration is induced within ten minutes [118]. H_2S [119], acetaldehyde [120], and ammonia [121] have also been proposed as yeast cultures synchrony mediators able to guide the inter-cellular coordination.

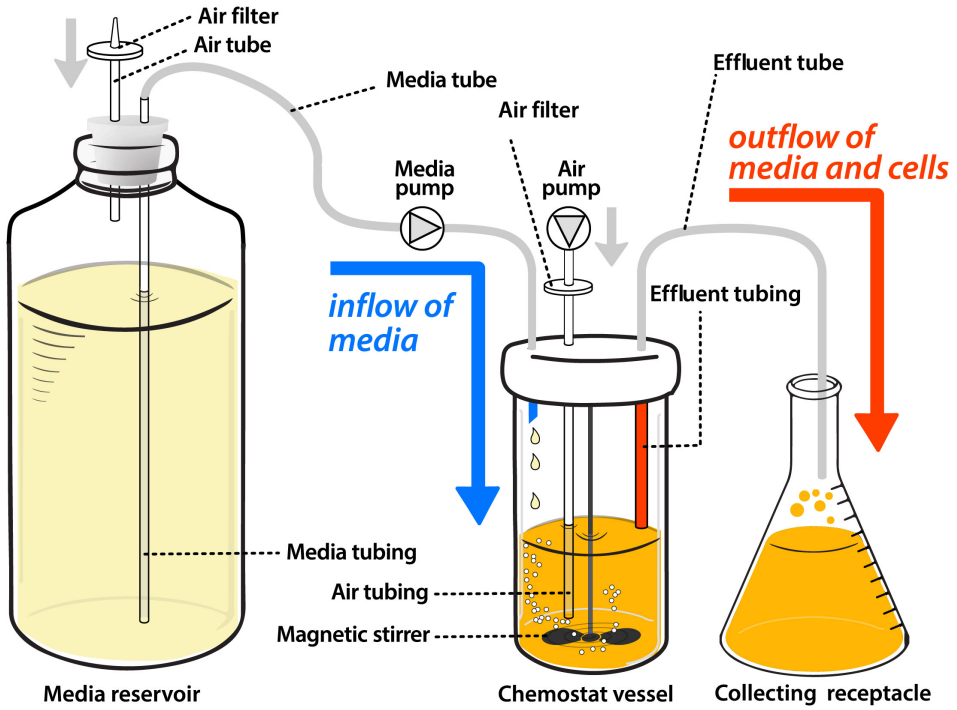


Figure 1.3: Experimental setup scheme of the chemostat. A chemostat [110] is an experimental device used for a wide variety of cultures into which media and nutrients are delivered at a constant rate, while maintaining the total volume unchanged. The media reservoir on the left contains the media used to feed the cells. The media is introduced at a constant rate into the chemostat vessel via the media tube. The vessel is aerated at a constant rate (5 L/min) using the air pump and the magnetic stirrer (speed 600 rpm), which makes a turbulent fluid inside the vessel. Acidity is kept constant at a pH=3.5 using buffer and a pH probe (not shown), and dissolved oxygen (dO_2) is measured using a second probe (not shown). In the vessel, yeast cells are grown to a high density [$OD_{600} \sim 8.5$] and temperature is kept constant at T=30°C. Finally, the collecting receptacle on the right receives the chemostat excess volume via the effluent tube, ensuring that the vessel volume is kept constant. Image taken from [111].

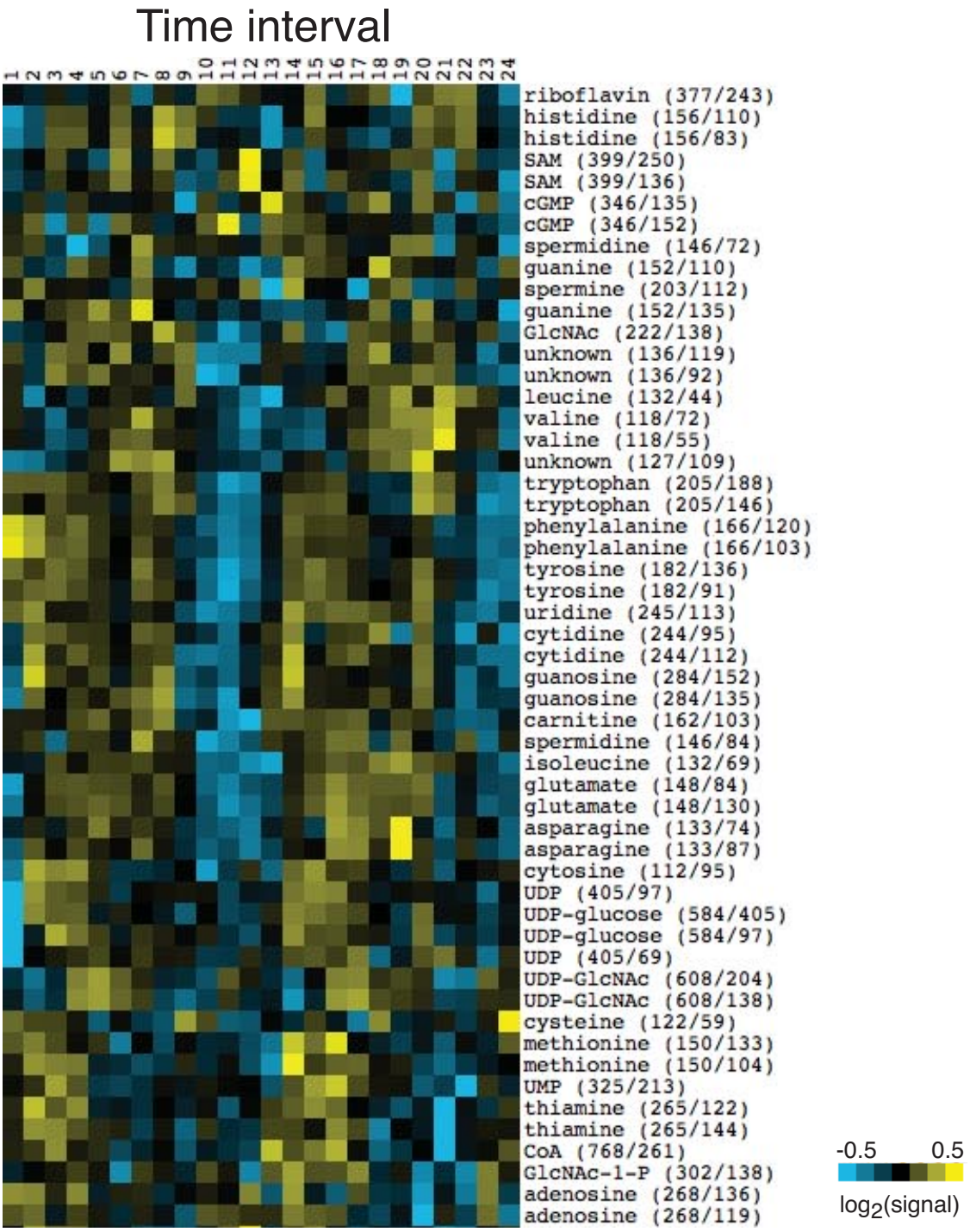


Figure 1.4: During the YMC, hundreds of yeast metabolites oscillate in the intercellular media. Rows correspond to the indicated metabolites and columns correspond to 24 time intervals over two consecutive yeast metabolic periods. Colour (see legend) represents $\log_2$ of the metabolite signal intensity change during one oscillation: blue is -0.5 ($\log_2$ change) and yellow is +0.5. Measured using liquid chromatography combined with tandem mass spectrometry (LC-MS/MS technique). Taken from [112].

1.4.2 Environment composition alters the YMC

Yeast exposure to visible light inhibits respiration, provoking an interruption of the YMC when intense, visible, light in the blue/green part of the spectrum is projected onto the culture [56].

Changes in the media glucose concentration are able to change both the YMC period (see Figure 1.5) and the number and amplitude of cycling genes [122], which implies that cyclic gene expression is an efficient strategy for optimizing metabolic cost, since it has been found [122] that otherwise the expression of such genes would be twice as energy-consuming. Figure 1.5 contains only two data sets, but a roughly proportional functional relationship between the YMC frequency and the media glucose concentration can be inferred.

The addition of antidepressants like lithium chloride [125] or phenelzine [108] causes great distortions to the shape and period of the metabolic oscillations, as we will explain in Section 2.1.

1.4.3 The YMC is divided into three phases

Tu *et. al.* [3] made a YMC phases taxonomy that has been widely accepted in the subsequent literature. They divided the YMC into three phases [3, Fig. 3D], each of them characterized by a set of genes sharply expressed at each temporal window. Aggregated oxygen consumption is inferred from the media dissolved oxygen concentration (see Subsection 1.5.2), while individual genes' temporal expression is measured using RNA-seq and other sequencing techniques. This double classification by both respiratory cycles and genome-wide oscillations of transcription [126] shows a number of metabolic processes being characteristic of each of the three phases, as depicted in Figure 1.6.

These three phases, along with their key cellular processes, are:

- Oxidative (OX) phase: Ribosome synthesis, amino acid synthesis, sulfur metabolism, RNA metabolism
- Reductive/Building (RB) phase: Cell division cycle, mitochondria biogenesis
- Reductive/Charging (RC) phase: Fatty acid oxidation, glycolysis, quiescence, DNA repair

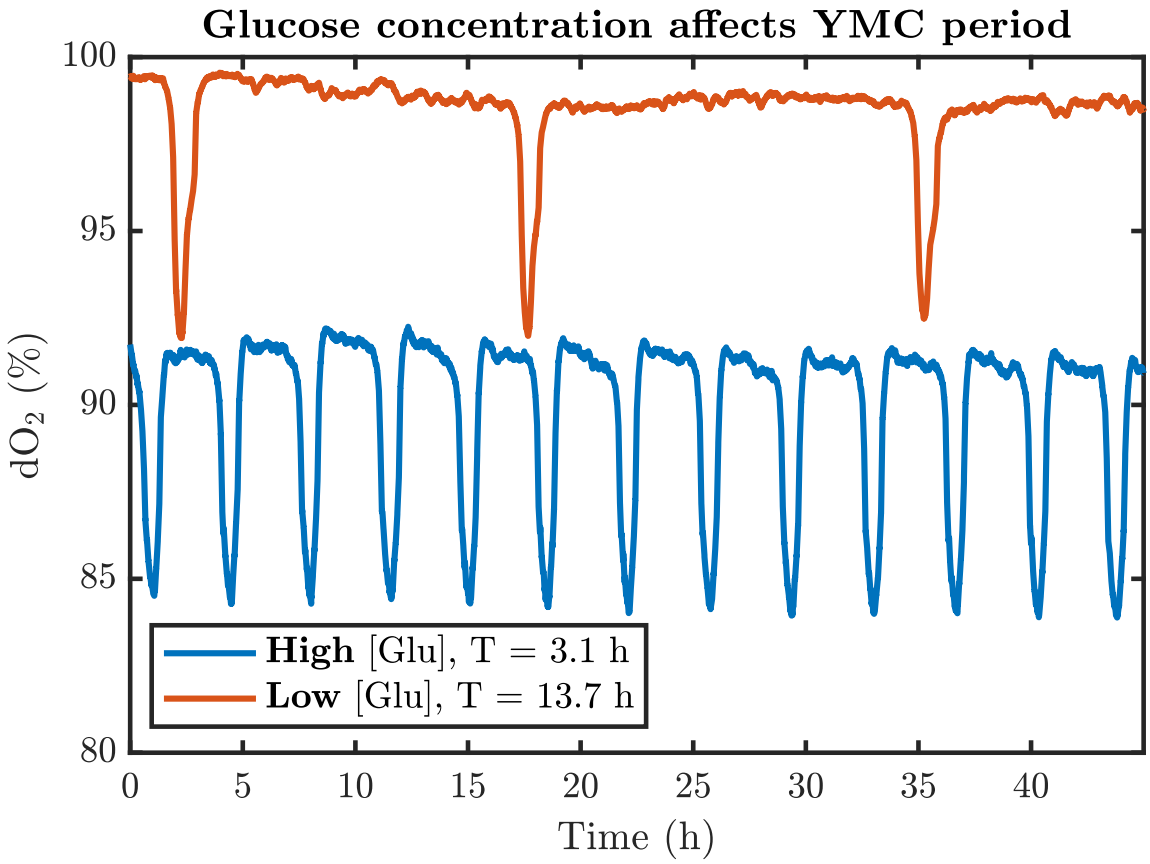


Figure 1.5: Glucose concentration alters the YMC period. The short cycle is grown at $[C_6H_{12}O_6]$ = 10g/L, resulting in a 3.1 hours period (T), while the long cycle is grown at $[C_6H_{12}O_6]$ = 2.5g/L, resulting in a 13.7 hours period. y-axis: dissolved oxygen concentration (as % of saturation) in the media (dO_2). Note that dO_2 levels are systematically lower in the high glucose concentration environment, revealing a more intense respiratory activity when cells are surrounded by more nutrients. When dO_2 levels are high, some form of 'kinks', or high-frequency oscillations, seem to be present; these low-amplitude ($A < 0.2\%$) kinks are explained by the manufacturer's user guide [123] as noise caused by air bubbles sticking to the oxygen sensor, resulting in an error range of $\pm 0.2\%$ with respect to the saturated dO_2 level. The chemostat is a turbulent fluid, and this noise is precisely an expression of such turbulent dynamics. Both data series [124] refer to the same yeast strain under analogous experimental conditions.

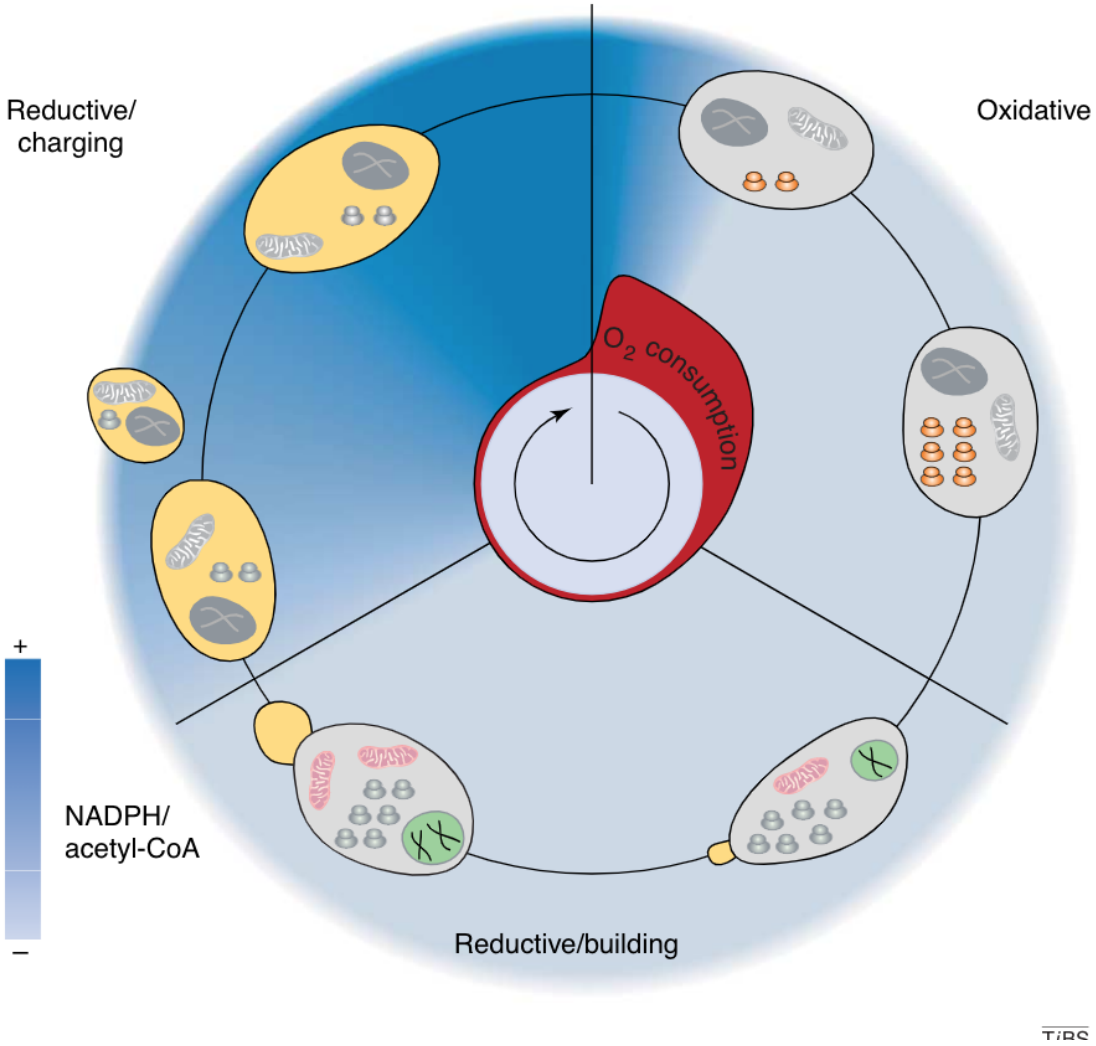


Figure 1.6: Graphical representation of the YMC and its main cellular processes. The cellular substructures undergoing the most pronounced changes in each phase are shown. These include ribosomes (orange), the nucleus (green), mitochondria (pink) and the processes of budding and daughter cell separation (yellow). The blue gradient symbolizes the intracellular concentrations of NADPH and acetyl-CoA. The red area represents the amount of oxygen consumption from the medium. Taken from [126].

Figure 1.7 represents a clear correspondence between these three YMC phases and the associated media dissolved oxygen concentration.

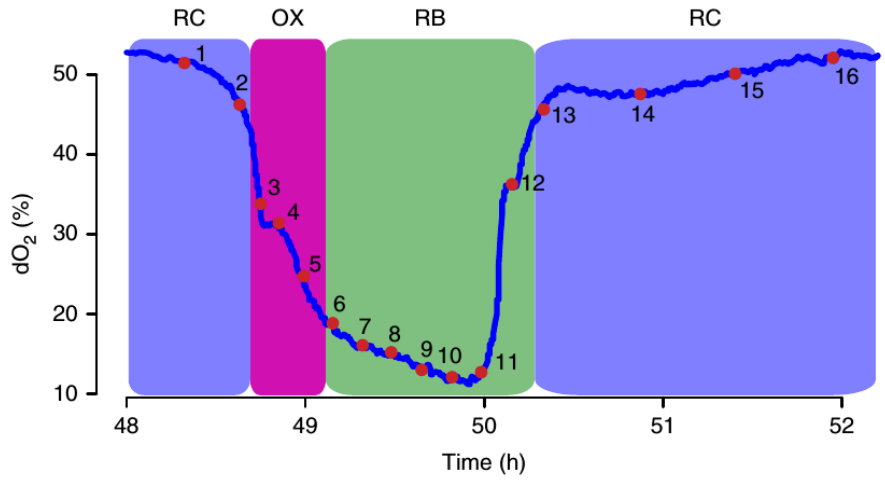


Figure 1.7: Identification of dissolved oxygen concentration with the three YMC phases: Reductive/Charging (RC, blue rectangles), Oxidative (OX, red rectangle), and Reductive/Building (RB, green rectangle); y-axis is the media dissolved oxygen concentration (dO_2), as measured by the chemostat probe, x-axis is time. Also represented (red dots) are the 16 sampling timepoints in Boeke *et. al.* [127]. Taken from [127].

Boeke *et. al.* [127] clustered the cycling genes in the yeast genome into these three phases using their temporal expression profile (RNA-seq reads) and then performed a Gene Ontology (GO) analysis on each of the clusters, identifying for each of the YMC phases the processes listed in Table. 1.1.

Table 1.1: Enriched GO terms associated with each of the three YMC phases. Cycling yeast genes are clustered into the three YMC phases according to their temporal expression profile. GO analysis is then run on each of the clusters, and these processes were identified as predominant at each phase. Adapted from [127].

OX	**RB**	**RC**
rRNA processing	Cell cycle	Mitochondrion organization
Ribosome biogenesis	DNA replication	Cytoskeleton
Methyltransferase activity	Mitochondrial translation	Mitochondrial envelope
RNA modification	DNA repair	Peroxisome vacuole
Ribosomal protein	Mitochondrion organization	Membrane invagination
Cytoplasmic translation	Mitochondrial ribosomal protein	Response to oxidative stress

1.4.4 The YMC and the CDC

The YMC is intimately coordinated with the Cell Division Cycle (CDC) [128–132], as evidenced by the cell cycle-regulated genes' periodic expression during the YMC [133]. For example, the yeast metabolic oscillator separately gates and maintains synchrony with the early and late cell cycle [134], and cells enter quiescence only during the late RC phase, when they have accumulated enough trehalose [135]. Related to this point, a question addressed in this thesis is whether the YMC implies that cells enter a quiescent state during the RC phase [136], or quiescence is a *parallel* process to metabolic cycling [137] in which, after each oscillation, a fraction of the cells undertake a new round of metabolism and the rest of them enter quiescence for one cycle, *superimposing* a metabolic cycle on the cell cycle [130].

In this synchronized system, metabolic oscillations and cellular doubling time follow a well-studied response curve [138, Fig. 3]. As we discussed in Subsection 1.4.2, the YMC oscillatory period can be altered merely by changing the carbon-source metabolite media concentration, following the functional form described by Papagiannakis *et. al.* [134, Fig. 1C], and therefore altering the CDC period. The Botstein lab (Princeton University) has performed some experiments on the form of this interplay, showing that the dynamics of the high-to-low oxygen consumption YMC transition mechanism is entirely analogous to the CDC environmental stress response (ESR) machinery. The Botstein lab performed a detailed analysis of changes in gene expression [139], concluding that growth rate response correlates with changes in the relative durations of the YMC phases; therefore, both oscillators are co-regulated [129]. Most remarkably, they also show that it is possible to have metabolic cycling without the CDC, and that metabolic synchrony does not require a carbon-source limitation [140].

These results were confirmed when the CDC Start phase was shown to be triggered by the YMC entry into OX phase [138], and similarly CDC progression to be guided by the subsequent YMC phases [134], indicating that the YMC and the CDC form a system of coupled oscillators on which quantized coupled frequencies can coexist. Several authors have identified metabolic processes whose periods are an integer fraction of the corresponding CDC period in organisms like *Cricetulus griseus* [141], *Thalassiosira Fluviatilis* [142], or *Gonyaulax polyedra* [143], therefore it would not be a surprise to observe the same feature in yeast under the YMC.

1.4.5 Histone modifications and the YMC

Histone modifications (HMs) are covalent post-translational modifications to the histone proteins, the proteins that package and help to arrange the DNA into nucleosomes. HMs are intimately related to both gene activation and epigenetic changes regulated through chromatin organization [144–146], while their function in biological processes is in general a highly debated topic [147,148]. HMs are identified using the ChIP-seq technology [149], on which chromatin immunoprecipitation is followed by sequencing.

In 2014, Boeke *et. al.* [127] combined RNA-seq and ChIP-seq measurements to unveil the YMC dynamics of RNA and several HMs: H3K4me$_3$, H3K36me$_3$, H3K9ac, H3K14ac, H3K56ac, H4K5ac, and H4K16ac (see Appendix A.1.2). They sampled sixteen timepoints across the YMC (see Fig. 1.7) aiming to elucidate some of the yeast transcriptional regulation mechanisms. They also measured the dynamic localization of three chromatin-modifying enzymes (Gcn5p, Esa1p, and Set1p) and explored the yeast cycling viability under a series of H3 mutants (see Figure 1.8.b-c). They observed that no individual lysine to arginine (K→R) mutant could hinder the YMC, but some particular combinations of amino acid substitutions could. For example, they found that the H3K(9,14,18,23)R mutant can cycle (see Figure 1.8.b), while the H3K(9,14,18,23,27)R mutant cannot (see Figure 1.8.c).

The sequencing data obtained by Boeke *et. al.* [127] were used by the same group as an input to build a transcription factors (TFs) binding sites retrieval algorithm (DynaMO) [150] that proposed Msn2p and Msn4p as the TFs controlling transition from the RC/quiescent state to the OX/growth state by regulating the intracellular level of acetyl-CoA through glycolysis [151]. DynaMO has also been employed to learn that stress response factors drive the regrowth of quiescent cells [152].

1.4.6 Single cells in unsynchronized cultures have an analogous cycle

Regular metabolic cycling has been observed in single yeast cells using fluorescent gene expression reporters [136]. This result has been further characterized by using Fluorescent In Situ Hybridization (FISH) to measure mRNA levels in single yeast cells, showing that mRNA transcription oscillates with the metabolic cycling periodicity [153].

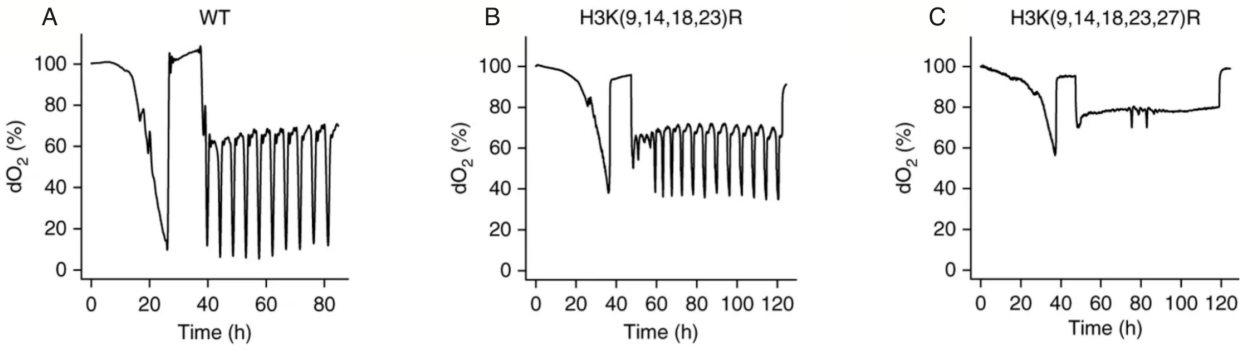


Figure 1.8: Chemostat dissolved oxygen traces of WT yeast and histone point mutants undergoing the YMC. y-axis: dissolved oxygen present in the culture (dO_2); x-axis: time from the start of the culture. (a) WT yeast is induced into the YMC following the same experimental setup as in Fig. 1.1. (b) The cycle can be maintained even if several lysine to arginine (K→R) substitutions take place. (c) If too many HMs are prevented by amino acid substitutions, dO_2 traces indicate that either cycling no longer takes place or the cells have desynchronized. We can observe how as the number of K→R mutations is increased, the amplitude of the respiratory oscillation decreases. A similar decrease is observed in the initial depth that separates the batch growth and starvation phases, occurring at $t = 26$ h in (a) and at $t = 38$ h in (b,c). Such a decrease could be explained by the loss of the cellular synchronization ability as more mutants are introduced. In panels (b) and (c), the experiment is interrupted (glucose feeding stops) after 120 hours. In panel (c), some residual activity at $t = 78$ h and $t = 81$ h seems to suggest that yeast are trying to synchronize their entrance into the YMC OX phase, but they are not succeeding. It is suggested that the proper functioning of post-translational modifications caused by such a number of mutations is preventing the normal coordinated cycling. Adapted from [127].

Metabolic cycling does take place in single yeast cells on glucose limitation from unsynchronized steady-state populations, forming an oscillation entirely analogous to the 'macro' YMC oscillation [154]. In unsynchronized cultures, there is no possible intercellular coordination term, since both metabolites and proteins concentration are kept constant in the intercellular media [154, Methods]. In Section 4.9, discussing the simulation of a single-cell transcriptional oscillator (Fig. 4.10), I do obtain mRNA oscillations remarkably similar to the ones measured by Silverman *et. al.* [154, Fig. 1]. This result is consistent with the fact that single cells ATP and NAD(P)H levels oscillate at almost the budding frequency [134, 155].

1.4.7 YMC and yeast mutants

Different aspects of the YMC regulatory machinery have been elucidated by analogous experiments to the ones discussed in Subsection 1.4.5, Fig. 1.8, but measuring respiration for several yeast mutants, an idea started by [156, Fig. 1A] that has been followed by others [112, 118]. It is observed, for example, that the $\Delta hmx1$ period is 25% longer than the wild type (WT) period [157]; that oscillations are defective in the $msn2\Delta$, $msn4\Delta$, and $msn2\Delta msn4\Delta$ mutants [151, Fig. 2A]; many other such cases have been identified. The YMC period and gene expression in the $xrn1\Delta$ mutant [158] is discussed briefly in Appendix A.8.3 (see Fig. A.8).

1.4.8 Mathematical models of yeast

Yeast metabolism and the division cycle have been widely modelled by Béla Novák by measuring mutants' growth conditions [159, 160], using ordinary differential equations [161, 162], employing network dynamics [163, 164], and all three techniques combined [165].

Other yeast metabolism models have used stoichiometry analysis of microbial growth and product formation [166] or metabolic flux analysis [167, 168]. When the yeast genome was sequenced, a genome-scale metabolic model was produced [169] and a few years later, a consensus genome-scale metabolic network of yeast was produced [170], which was afterwards rendered suitable for computational analysis [171].

Bioinformatic models of the gene expression measurements of Tu *et. al.* [3] and Li *et. al.* [108] have found a global feedback system between metabolism and chromatin [172] and a global reset point during the OX phase [173], as well as one of the first models of the yeast oscillatory dynamics regulation [174].

Yeast metabolic oscillations have been modelled using perturbation terms [175], protein turnover regulation [176], and glucose/ATP cellular concentrations [101].

1.4.9 A unique system?

Highly similar metabolic cycles have been found in many species [46, 177], from archaeas to humans, indicating that metabolic cycles may be the original, primitive, cycles from which other biological cycles, like biological timekeeping [30], transcriptional activation and repression [178], or clock-controlled chromatin remodelling [179, 180] are formed. If this is the case, then the study of YMC-like processes should provide a good insight into the study of other cycles. It is not known how the YMC is regulated; however, as we discussed in Subsection 1.4.5, some mutants unable to incorporate certain HM combinations do not synchronize their metabolic activity (see Fig. 1.8.c). I aim to explain cycle regulation through HMs accumulation at particular loci. The study of metabolic cycles is also relevant because metabolism-related conditions (like diabetes, obesity, cancer, cardiovascular diseases, etc.) are linked to disordered lifestyles [181], which are probably coordinated by cell clocks whose irregularity plays a role in these diseases [182]. This system may also provide a relevant insight into the normal state of human cells, since cancer cells tend to favour a particular form of metabolism [183–186].

The literature review part of this thesis finishes here. The two remaining sections of this Chapter are devoted to performing an initial exercise combining bioinformatic methods with literature knowledge in order to accurately calculate the yeast oxygen consumption from the media dissolved oxygen concentration data (Section 1.5) and to provide a brief overview of the thesis body and structure, explaining in detail the questions that this doctoral work will try to answer, establishing a relationship between the main literature conclusions discussed so far and the main concepts discussed in this thesis (Section 1.6).

1.5 Yeast respiration during the YMC

The aim of this Section is to understand yeast respiration and its relationship to metabolism and metabolic coordination using a mathematical estimate of the cellular oxygen consumption levels across the YMC.

As we discussed in the previous Section, the YMC was initially identified by observing fluctuations in the media dissolved oxygen (dO_2) and metabolite concentrations, and afterwards microarray and RNA-seq experiments revealed a transcriptional activity coordinated with these respiratory bursts. In the chemostat, dO_2 is readily monitored using a specific probe; therefore, even if, as we have seen, hundreds of metabolites oscillate during the YMC, the simplicity with which dO_2 is measured means that, for experimental purposes, the YMC phase is always identified with the chemostat probe dO_2 measurement.

However, it is obvious that the magnitude that we measure, the chemostat dissolved oxygen trace dO_2, is not directly identifiable with the cellular oxygen consumption, but must be related by some functional relationship. To understand the precise pace of cellular oxygen consumption, I perform the relevant analysis in Section 1.5.2. These calculations will be used in future chapters to elucidate some of the aspects of the YMC.

During the YMC, yeast cells uptake the chemostat media dissolved oxygen and combine it with other compounds to perform their most basic metabolic activities[3], unlike in the case of anaerobic fermentation [188, 189]. As we discussed in the previous Section, the first papers studying yeast metabolic oscillations measured pyridine nucleotides [98] and glycolytic intermediates [94] fluctuations, but shortly afterwards it was observed that dissolved gases like O_2 and CO_2 also oscillate during the YMC [100], and limited aerobic growth [104] and media dO_2 concentration [119] were shown to be correlated with YMC regulation.

The YMC has been described fundamentally as a system of dO_2 oscillations in a chemostat [190]. Since the two pioneer papers in modern YMC studies [3, 108], it has been shown that these cycles are matched by other oscillations: in gene expression [3], NAD(P)H fluorescence [19, 191], acetyl-CoA concentration [116], hundreds of metabolites' concentrations [112], and accumulation of histone covalent modifications [127]. It has also been shown that oxygen is one of the 8 kinetically stable but thermodynamically activated molecules that power yeast cell metabolism [187].

From these considerations, an obvious question arises: is oxygen coordinating the YMC? Subsection 1.5.1 will give a negative answer to this question, and then I will use fluid dynamics theory [192] to calculate the cellular oxygen consumption rate

[3]Two excellent reviews on yeast metabolic pathways and chemicals required for its survival are Cai *et. al.* [128] and Walsh *et. al.* [187].

from the chemostat dO_2 probe measurements in Subsection 1.5.2.

1.5.1 Is oxygen coordinating the YMC?

Some bacteria are able to sense the quantity of oxygen present in their surroundings and tune their metabolic activities accordingly [193–201]. The *E. coli* FNR regulon O_2 sensor [202–205], the FixL proteins in other bacteria [206–208], or the HIF-α complex in humans [209–212] can sense the surrounding oxygen concentration and use this information to decide to activate the relevant metabolic pathway. Yeast cells can sense oxygen in a wide variety of ways [213–217], but the comprehensive oxygen sensing mechanisms are not fully understood [213, 218–221].

It is a tempting idea to explain the YMC coordination in terms of the dO_2 and metabolic oscillations interplay that we have been discussing: cells sensing a dO_2 concentration decrease would boost their oxygen-intensive metabolic processes in response, with the coordinated result of consuming a sizeable fraction of the media dO_2 in a matter of minutes, therefore inducing rapidly the OX phase, as shown in Fig. 1.7. This Section is however devoted to explaining the reasons why oxygen-sensing is not related to the establishment and maintenance of intercellular synchrony, and particularly why it cannot induce the oxidative (OX) phase of the YMC.

A crucial experiment to disprove the hypothesis that dO_2 concentration coordinates the YMC consists in replacing the O_2 in the chemostat with some inert gas, assessing if this substitution has any effect on cellular coordination [157]. Figure 1.9 -a reproduction of [157, Fig. 3B]- shows WT yeast cells undergoing the YMC. At some point in the midst of the RC phase (marked with an arrow), air (oxygen) pumping is replaced with inert gas (argon) pumping for two minutes. This way, O_2 is swept from the media and dO_2 relative concentration falls from 60% to 20% in these two minutes. However, once argon pumping finishes and air pumping is switched back in, the system instantly reassumes the RC phase it was in, whose duration is not altered. Since there is no OX phase induction after exposing the cells to low dO_2 levels, it is concluded that oxygen sensing is not the YMC synchronization mechanism. This phenomenon is modelled in Subsection 5.4.2, obtaining similar results in both RNA and protein production.

Oxygen sensing is linked to both NADPH concentration [210] and the $[4Fe–4S]^{2+}$ cluster stability [204]. It is therefore possible that the NADPH oscillations observed

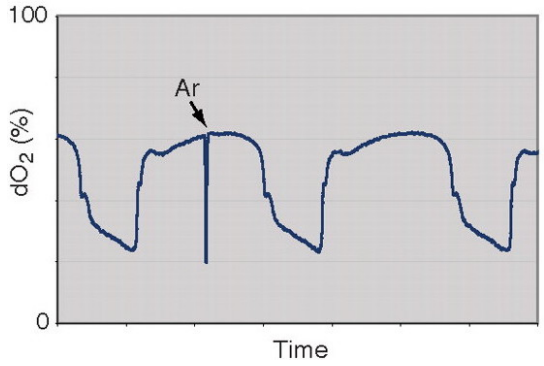


Figure 1.9: Transient administration of argon has no apparent effect on metabolic cycling. Cycling WT cells are exposed to argon rather than air applied at ≈5 psi for 2 minutes in the midst of RC phase (black arrow). y-axis: dissolved oxygen present in the culture (dO_2); x-axis: time. When argon is being applied, dO_2 falls from 60% to 20% in just two minutes. Once air pumping is switched back in, the system instantly reassumes its previous RC phase, whose duration is not altered. Taken from [157].

during the YMC [112] are a consequence of the YMC itself (maybe through some K^+ channel [210]) if it was an O_2 sensor. In this case, low levels in the $[4Fe–4S]^{2+}$ cluster could affect the cycle length, because cells would not sense the just-reached OX phase [204], implying that it would take longer for them to complete their metabolic cycles. Experimental confirmation of this idea is deferred to Section A.8, where it is shown how an increase in $[4Fe–4S]^{2+}$ clusters availability can increase the frequency of the yeast metabolic oscillations by 50%.

1.5.2 Cellular oxygen consumption calculation

This Subsection aims to establish a functional relationship between the dO_2 traces measured by the chemostat oxygen probe and the inferred cellular oxygen consumption during the respiratory bursts that form the YMC, using novel experimental data [222]. Such analysis will be used through this thesis to construct a number of bioinformatic models and simulations: for example, in Section 5.5 (see Fig. 5.7) it will be used to identify oxygen consumption with the concentration of proteins coded by genes expressed during the OX phase of the YMC.

Experiment description

Dissolved oxygen traces data were obtained using a chemostat, whose experimental setup was described in Fig. 1.3. In a chemostat, air is imputed at a constant rate, and homogeneously distributed inside the vessel using a magnetic stirrer [223]. The chemostat also monitors pH and temperature, but the point of a chemostat is to keep these two magnitudes as constant as possible [110]. Because of the metabolic oscillations during the YMC, pH would oscillate if we were not to buffer the media to achieve a constant pH of 3.5. dO_2 is automatically measured by the chemostat probe every 5 seconds, which given the time-scale of the oscillations (T$\sim$ 3.5 h) should eliminate any time-resolution concerns.

The experimental data I have used [222] to perform these calculations are initially the chemostat oxygen probe dO_2 records when the chemostat was active, as shown in Figure 1.10.a, where a number of YMC inductions can be observed. In particular, I focus on a highly regular time window of YMC oscillations: the time span from $t = 500$ h to $t = 570$ h, marked red in Fig. 1.10.a and enlarged in Figure 1.10.b.

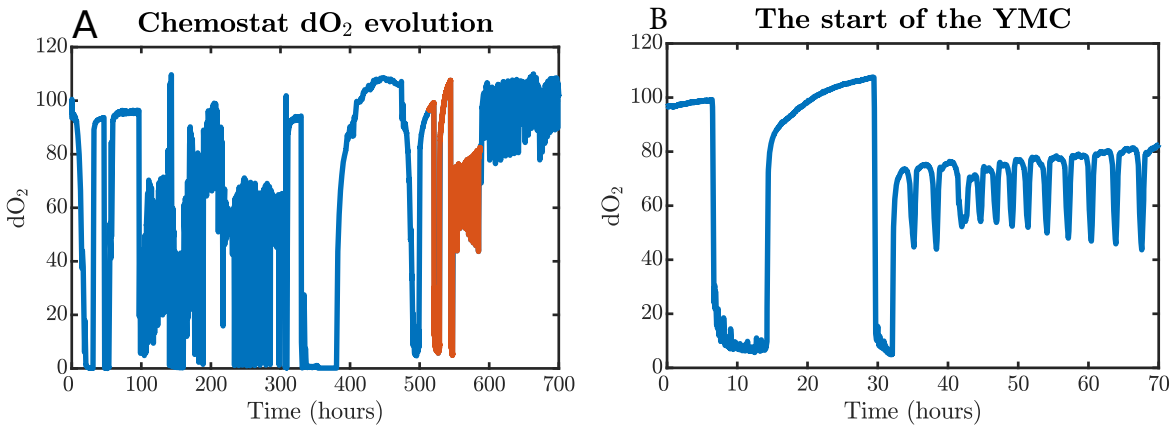


Figure 1.10: (a) Dissolved oxygen (dO_2) in the chemostat (as measured by the oxygen probe) during a period of approximately 30 days. x-axis: time; y-axis: media dO_2 level. Red: one of the regions in which the YMC takes place, $t = 500$ h to $t = 570$ h. The oxygen probe is calibrated before the beginning of the experiment using distilled water. At any temperature, water has lower oxygen solubility than the chemostat mixture of media and yeast cells; therefore, sometimes dO_2 can raise above the 100% level. (b) Expanded view of the time region marked red on the left panel. The batch culture phase is started at $t = 7$ h by adding glucose-rich media to the yeast cells in the chemostat. $t = 14$ h marks the moment at which the batch culture phase is halted and the starvation phase is induced. Glucose feeding is resumed at $t = 30$ h. Increasingly regular oscillations are observed in the $t > 32$ h region. Data from [222].

Analysis of secondary frequencies

The $T = 218$ minutes oscillations discussed in this subsection are immediately identified by eye in Figure 1.10.b. However, it is possible that these oxygen traces are generated by the overlap of several, more subtle, processes and cycles whose smaller amplitude leaves them unnoticed on a superficial analysis.

To check the possibility that different-frequency oscillations have been missed, I perform a Fourier transform analysis (MatLab `fft` function) on the original chemostat oxygen probe data. The obtained frequency spectrum decomposition identifies a strong peak associated with the $T = 218$ minutes oscillation, smaller peaks at its multiples $T = n \cdot 218$ min, and finally a wider, small-amplitude peak at the end of the spectrum, the $T < 1$ min region, which is identified with the probe measurement noise caused by air bubbles [123]. Fourier transform methodology, and its utility to detect underlying oscillations, will be discussed at greater length in Section 3.2.2 (see Fig. 3.3). MatLab autocorrelation function `autocorr` revealed a similar result: there are not secondary frequencies present in the dissolved oxygen data analyzed. Therefore, the calculations performed in this subsection can consider only the $T = 218$ minutes oscillations present in the $t > 32$ h region of Fig. 1.10.b.

Calculations

As a first approximation, I take the effective oxygen intake of the yeast cells to be a function of both the chemostat aeration rate and the media O_2 saturation level. I focus my analysis on six of the consecutive YMC oscillations shown in Fig. 1.10.b, and reproduced in greater detail in Figure 1.11.a.

Hamiltonian plots are a standard topological method in data analysis to visualize curves in phase space [224, 225]. By plotting the chemostat dissolved oxygen trace (dO_2) discussed in the previous paragraph against its time-derivative (dO_2') we obtain a useful insight into the form and regularity of these oscillations. The chemostat oxygen probe measures dO_2 every 5 seconds. Because of the inevitable noise in oxygen probe measurements, mostly caused by gas bubbles [123], I apply a moving window filter of 1 minute duration (13 time points, so $\gamma = 12$) and transform the data:

$$dO_2(t_0) = \frac{1}{\gamma + 1} \sum_{i=t_0-\gamma/2}^{t_0+\gamma/2} r(i), \tag{1.1}$$

where $r(i)$ are the raw chemostat oxygen probe measurements at time point i, and

dO_2 is the filtered, noise-free, measurement. $\mathrm{dO}'_2(t)$ can then be approximated using Newton's formula:

$$dO'_2(t_0) \approx \frac{f(dO_2 + \delta) - f(dO_2 - \delta)}{2\delta}, \tag{1.2}$$

where $\delta = 5$ seconds, and the result is shown in Fig. 1.11.b. Other δ values in the 1-30 seconds range resulted in analogous $\mathrm{dO}'_2(t)$ results.

The functional relationship between dissolved oxygen (dO_2) and cellular oxygen consumption (Q_{O_2}) is taken to be:

$$Q_{O_2} = f_{abs}(dO_2) - \frac{d(dO_2)}{dt}, \tag{1.3}$$

where $f_{abs}(dO_2)$ is a function representing the media oxygen absorption at a given oxygen saturation level introduced by the oxygen pump. Physical interpretation of Eqn. (1.3) is clear: the oxygen being consumed by the cells Q_{O_2} is the rate at which oxygen is being transferred from the media to the cells, so it must be the oxygen imputed into the media rate f_{abs} minus the rate of change of oxygen in the media $\mathrm{dO}'_2(t)$.

On a first approximation, I take $f_{abs}(dO_2)$ to be:

$$f_{abs}(dO_2) = \kappa \cdot \left(1 - \frac{dO_2}{\max(dO_2)}\right), \tag{1.4}$$

where κ is a constant dependent on the geometry of the experiment and dO_2 is expressed as a fraction of the total media oxygen solubility at that chemostat temperature. The physical meaning of Eqn. (1.4) is that absorption is maximized when there is no oxygen in the media, and the higher the saturation of oxygen molecules, the harder it is for new molecules to dissolve in the media. I justify the functional form of Eqn. (1.4) on the grounds that absorption is directly proportional to the number of available spaces for oxygen molecules to be dissolved in the media [226], and this is proportional to $(1 - \mathrm{dO}_2)$, particularly in a turbulent fluid like in a chemostat, in which the fermenter paddles keep it in constant agitation [227].

To find the value of κ, I assume that at some point during the YMC, oxygen consumption is negligible [228], the OX phase point at which the dO_2 derivative is highest. Using the data obtained from Eqn. (1.2):

$$\max_{6\ cycles}\left(\frac{d(dO_2)}{dt}\right) = \frac{3.1\%}{1\ \mathrm{min}}. \tag{1.5}$$

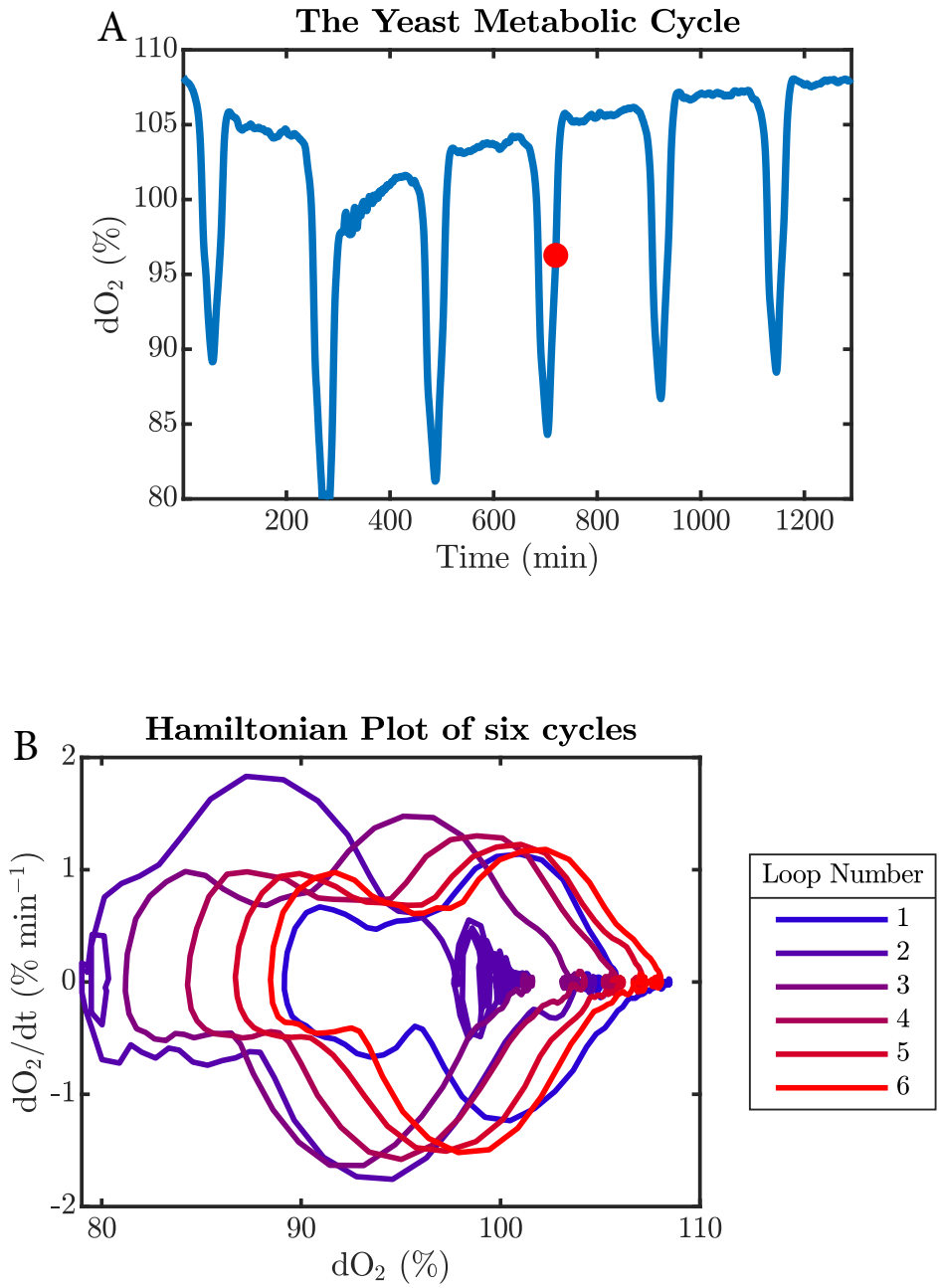


Figure 1.11: (a) Enlarged section of six respiratory oscillations in Fig. 1.10.a. x-axis: time; y-axis: media dO_2 level transformed using Eqn. (1.1). Red dot represents the highest dO_2 derivative (max dO_2') point obtained from Eqn. (1.2). (b) Hamiltonian plot displaying same data as the left panel, but in phase space. x-axis: transformed dO_2 data; y-axis: time derivative of such values, dO_2'. The six loops are displayed in different colors, blue corresponding to the first oscillation ($t = 0$ to $t = 218$ min) on the left panel and red being the last oscillation ($t = 1090$ min to $t = 1308$ min) on such panel. A colour gradient from blue to red is used to represent the other four loops. Raw data from [222].

Since dO_2 at this highest-derivative point is 96.4/107.1 = 90% (marked red in Fig. 1.11.a), I obtain $\boldsymbol{\kappa = 31\% \cdot min^{-1}}$. This means that, if there was no dO_2 in the medium, dO_2 would pass from 0% to 26% in a minute[4]. From this κ value, I use Eqn. (1.4) to calculate the functional form of $f_{abs}(dO_2)$ and then, using Eqn. (1.3), I obtain the yeast population oxygen consumption $Q_{O_2}(t)$, which is displayed in Figure 1.12, where we can appreciate the difference in oxygen consumption at the low and high oxygen consumption phases.

Error analysis

A possible source of error in the Q_{O_2} calculation given by Eqns. (1.1) to (1.5) arises from the fact that media composition changes through the YMC [113, 174], and oxygen absorption is highly dependent on small variations of the media composition [229, 230]. This should not be a concern, since the value obtained for $\kappa = 31\%$ min^{-1} means that the dO_2 stabilization time is in the order of magnitude of minutes, while oscillations in dissolved metabolites have an order of magnitude of hours.

The absorption function $f_{abs}(dO_2)$ is not only dependent on the aeration rate and on the rotor speed, but also on a number of technicalities relating to the precise geometry of both the chemostat and the oxygen valve: for example, the smaller the bubbles, the larger $f_{abs}(dO_2)$. However, this precise fact is the reason why $f_{abs}(dO_2)$ was not directly measured, but rather inferred from the dO_2 data using Eqns.(1.4) and (1.5). In principle, variations in atmospheric humidity could also affect the absorption function, since humid air is more soluble than dry air [231, 232]. But because the experiments were performed indoors and at constant room temperature, humidity fluctuations should be also negligible [233, 234].

The κ value obtained in Eqn. (1.5) is in principle greatly dependent on background noise that could greatly alter the dO_2' value obtained via Eqn. (1.2). However, similar $\kappa > 25$ min^{-1} values are obtained for the maxima in the other five oscillations. These six dO_2' maxima are obtained at the same YMC phase as the overall maximum value obtained via the Eqn. (1.5) (marked red in Fig. 1.11.a), excluding therefore anomalous high values given by filter-resistant noise during the cycle. The standard deviation of the six values obtained for each of the oscillations was $\sigma = 2.2$ min^{-1},

[4]It would not reach 31% because as oxygen concentration increases, oxygen absorption decreases.

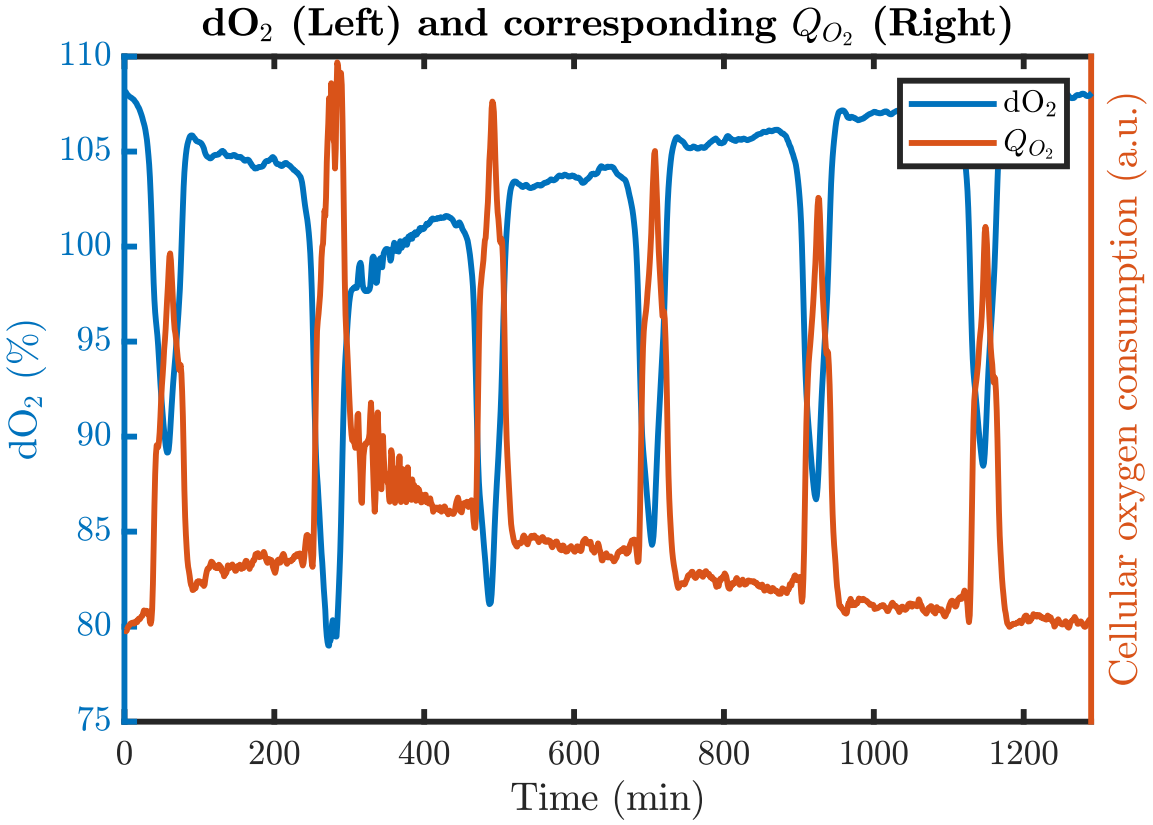


Figure 1.12: Dissolved oxygen (dO_2) and cellular oxygen consumption (Q_{O_2}) curves. Left y-axis (blue): dissolved oxygen concentration dO_2 - data transformed using Eqn. (1.1). The oxygen probe is calibrated before the beginning of the experiment using distilled water. At any temperature, water has lower oxygen solubility than the chemostat mixture of media and yeast cells occurring during the YMC; therefore, sometimes dO_2 raises above the 100% level. Right y-axis (red): cellular O_2 consumption in arbitrary units, calculated from dO_2 via Eqn. (1.4). x-axis: time.

therefore the κ value should be expressed as

$$\kappa = 31 \pm 2 \text{ min}^{-1} \tag{1.6}$$

with a 95% confidence.

1.6 Thesis aims

The overall aim of this DPhil thesis is to develop mathematical and computational methods to deconstruct, by using a multi-omics approach, the series of events and regulatory logic underlying the YMC. In other words, my main research goal is to break down and to rebuild afterwards the individual processes that control the YMC, paying particular attention to the genes that act as its regulators.

In particular, I aim to produce a mathematical description of the YMC that can quantitatively explain and predict its behaviour under a wide range of laboratory conditions, providing therefore relevant insights into the YMC regulatory machinery and allowing us to understand the fundamental processes underlying these oscillations. To achieve this, I have focused on three different datasets, whose technical details, source, and limitations are discussed in Appendix A.1: RNA-seq [235–239], as an indicator of the genes' transcriptional activity, measuring levels of transcripts and their isoforms; NET-seq [240, 241], measuring the nascent, actively transcribed RNAP II; and ChIP-seq [149, 242–246], measuring the abundance and position of histone modifications across the yeast genome, which are known to act sometimes as a component of epigenetic switches at genes [247–249], and to play a general role in gene regulation [250, 251].

1.6.1 Interplay of histone modifications, transcription, and transcript degradation in the regulation of the cycle

The regularity of yeast respiratory oscillations suggests the existence of some time-keeping mechanism, or biological clock, behind the YMC regulation. I aim to identify the nature of this oscillator and the genes that are involved in its regulation, using both RNA-seq, NET-seq, and ChIP-seq data and bioinformatic analysis tools. We know from Subsection 1.4.2 (see Fig. 1.5) that glucose concentration alters both the YMC period and the number and amplitude of cycling genes [122], suggesting that such a rhythm is intrinsically related to the yeast metabolic activity, and it is hence

potentially coupled to the yeast cell division cycle (CDC).

mRNA degradation could potentially play a role in the regulation of the cycle, and histone modifications could be involved in this process, aiming to obtain a precise dynamic calculation of the different genes' mRNA degradation rates and to investigate how these degradation rates relate to the histone modifications accumulation dynamics.

1.6.2 Investigate methods of clustering the data to identify classes of genes that could be co-regulated

I aim to individuate the precise components of the mechanism underlying the regulation of cycling metabolic activity observed during the YMC by using several bioinformatics techniques: histone modification time-evolution and mRNA degradation rates calculation in Chapter 2; and cluster analysis to discern the individual cellular regulatory components in Chapter 3, aiming to know if the YMC is chiefly regulated by TFs, by transcription, by gene expression, by metabolites, by post-transcriptional modifications of histone proteins, or by a combination of these and other means.

1.6.3 Is a transcriptional oscillator the basic control mechanism for the cycling of individual yeast cells?

It is crucial to distinguish between the individual yeast cells' transcriptional dynamics (which is studied in Chapter 4) and the yeast population's collective metabolic oscillations that form the observed respiratory bursts (studied in Chapter 5). The regularity of time-keeping mechanisms is fundamental for the proper functioning of metabolic processes [182] via the fluctuation of metabolites concentrations [174], and it is associated with the redox changes and gene expression oscillations responsible for the origin and maintenance of biological rhythms [252]. I explore the hypothesis that the metabolic cycle is controlled by a transcriptional oscillator, whose components can be identified with the retrieved clusters of genes.

1.6.4 Investigate how the cycles of individual cells are linked to form the observed synchronised cycling behaviour

The YMC takes place in a closed, stable environment, in which there are not periodic signals from outside of the chemostat. The only interactions with the outside of the chemostat (see Fig. 1.3) are the air and media inflow and outflow of the

vessel (which occur at a constant rate) and the addition of buffer to keep the pH constant (which is automatically triggered when pH in the vessel falls below 3.5, being therefore a response to the system). Temperature is also kept constant. This is probably what makes the YMC so peculiar compared to other periodic biological systems: in the YMC, under laboratory conditions, there is no entrainment from external factors like sunlight [253, 254] or temperature [255], as it is the case in plants' circadian clocks. Since non-entrained biological rhythms are in general not perfectly regular [256–258], and cells in the chemostat maintain their synchrony for indefinitely long periods of time, respiratory bursts and coordinated transcriptional activity observed in the chemostat do probably arise from some form of inter-cellular synchronization mechanism, such as ethanol concentration in the media.

Under this approach, the YMC can be explained as the result of two different, complementary, biological processes; the first is an internal transcriptional oscillator, on which periodic gene expression is achieved from a combination of regulatory feedback loops in transcription and translation, mechanisms that probably share relevant structural features with the cell division cycle [259]. This mechanism is a self-sustained oscillator on which the different cellular processes are compartmentalized in time in a periodic manner [114]. Internal clocks have been widely modelled in the context of circadian and other periodic rhythms [2, 260–262], discussing the connection of these mechanisms to metabolic rhythms [261, 263, 264], and key regulator genes like $PER2$ have been identified in a number of organisms [265, 266]; yeast is not a particular case, since their metabolic cycles share features conserved among other biological rhythms [267–269]. The precise transcription factors (and possibly other genes) acting as the components of this transcriptional oscillator will be explored in Chapter 4. However, this internal rhythm by its own cannot explain why cells are kept on a synchronized state for arbitrarily long periods of time, suggesting thus the existence of an intercellular synchronization mechanism responsible for maintaining the metabolic rhythms and transcriptional activity of yeast cells in the same culture, perhaps achieved through the secretion of signalling molecules like ethanol [270] or acetate [116, 117] to the intercellular media. In other words, this mechanism adjusts the individual cells' phases to form a collective oscillation [134] that is macroscopically observable in the form of respiratory bursts and changes in the average expression of most genes and some histone modifications (measured by RNA-seq and ChIP-seq respectively). The mechanisms underlying this group behavior have been briefly explored in Section 1.3, and they will be further investigated in Chapter 5.

A crucial aim of this thesis is to compare these two mechanisms, the cellular transcriptional oscillator and the inter-cellular coordination via signalling molecules secretion, to learn about the interplay between the self-sustained oscillator and the group behavior, specifically investigating the strength of their coupling.

1.6.5 Overview

In summary, this thesis aims to investigate how the metabolic compartmentalization is related to other cellular processes. Metabolic compartmentalization [271] is an efficient manner for cellular metabolism to solve a number of challenges, such as of toxic pathway intermediates, competing metabolic reactions, and slow turnover rates [272]. It has been linked to periodic rhythms in a number of organisms, like cyanobacteria [273], flies [274], and mice [82]. During the YMC, a number of crucial cellular processes are precisely compartmentalized in time [3], suggesting that metabolic cycling may be an efficient way to time-compartmentalize incompatible cellular processes as to increase the efficiency of the yeast metabolic machinery. In Chapter 3, gene clustering and gene ontology analysis will be used to identify the cellular processes that occur at different phases of the cycle, aiming to retrieve the functions performed within each of these compartments. In Chapter 4, this knowledge will be used to construct a single-cell transcriptional oscillator able to explain and reproduce, at a cellular level, the dynamic patterns in gene expression stemming from the compartmentalization of metabolic activity in the YMC. Finally, Chapter 5 will explore the external synchronization mechanisms from a mathematical perspective, formulating a complete model able to explain the most relevant features of the YMC, and to predict the outcome of some validating experiments in the form of precise research questions. Other, perhaps more subtle, consequences of this oscillatory phenomenon, such as inducing a highly acidic environment for short windows of time [275], or forming coordinated pulses of alcohol segregation [276] -both highly effective manners of getting rid of competitors in nutrient-scarce environments [277, 278]- will be explored along the way.

Chapter 2

Role of Histone Modifications and mRNA Degradation in the YMC

In this Chapter I discuss how Histone Modifications (HMs) and mRNA degradation are tied to YMC regulation. This Chapter aims to understand the temporal dynamics of some processes, like the deposition of covalent modifications on lysine residues in the histone H3 tail or the degradation of mRNA, known to be involved in the YMC regulation [127, 279, 280]. In particular, I aim to ascertain if, during the YMC, HMs are related to mRNA degradation rates and if HMs are instructive for transcription.

The Chapter is divided into two parts. The first part explores whether HMs can be instructive for transcription (Section 2.1) and how these HMs correlate with the local genes' RNA expression across time (Section 2.2). The second part is concerned with mRNA degradation during the YMC in general (Section 2.3), aiming to achieve a precise dynamic calculation of the different genes' mRNA degradation rates (Section 2.4) and to investigate how these degradation rates relate to the HM accumulation dynamics (Section 2.5).

Lysines in the H3 tails of nucleosomes positioned at both the Transcription Start Site (TSS) closest-neighbour[1] and inside the gene body exhibit highly dynamic changes in some of their covalent post-transcriptional modifications [282]. My goal is to study these oscillations in HM abundance to ascertain if they are somehow related to the regulation of mRNA production. Covalent modifications of the aminoacids in the flexible H3 tail are involved in transcriptional regulation [250, 251, 283, 284] by

[1]The first nucleosome after the Nucleosome Depleted Region [281].

altering the RNA polymerase II (Pol II) binding affinity to the DNA [285] and nucleosome positioning [286], with H3 lysine acetylation [287,288] and methylation [288,289] playing the most fundamental part. Pol II binding affinity to the DNA is altered because histone acetylation removes the histones positive charge [290], altering thus the neighbouring DNA geometry [286, 291], loosening the interaction between the histone and the negatively charged DNA [279], and hence altering the probability of Pol II being bound to that DNA fragment [292]. Histone methylation does not affect histone charge since the $-CH_3$ group is neutral in charge, but it creates a docking site for chromatin-associated proteins containing specific methyl histone-binding domains [279]. My goal is to extrapolate this knowledge by studying the HM-mRNA interplay through the YMC to identify its main regulatory mechanisms.

There are chiefly two methods by which mRNA concentration can be observed to peak through time. An increase in the relevant synthesis rate will increase the mRNA concentration, but so will a decrease in the mRNA degradation rate. Therefore, RNA-seq data cannot distinguish an increase in the synthesis rate from a decrease in the degradation rate -or vice versa-: mRNA degradation calculations require some measure of the mRNA synthesis rate. This rate is relatively given by the quantity of nascent transcription observed, or number of NET-seq reads [240, 293].

mRNA degradation can occur via a wide variety of mechanisms [294], chiefly glucose-regulated phosphorylation of Puf3 [295], deadenylation (through both the Pan2/Pan3 [296] and the Ccr4/Pop2/Not [297] complexes), decapping by a variety of factors [298–301], and 3' to 5' mRNA degradation by the exosome and some associated proteins [302–304]. mRNA degradation for different genes has been accurately measured for a variety of organisms, including yeast [305–308], but these experiments were performed in batch yeast, and different techniques result in completely different mRNA half-lives. For example, Farina *et al.* [305, Fig. 2] have measured the correlation coefficient between two previous calculations for batch yeast mRNA half-lives [306, 307] and found it to be only $r = 0.01$ [305]. My goal is to produce an estimate of mRNA half-lives during the YMC, and test the hypothesis that mRNA half-lives are constant [309–311] in this particular system too, in order to learn if mRNA degradation mechanisms are linked to the temporal dynamics of HMs. The mathematical estimate of the mRNA degradation processes undertaken in Section 2.4 is used in Chapter 4 to fit numerical values into a mathematical model of YMC regulation.

2.1 Are HMs involved in YMC regulation?

As we learned in Subsection 1.4.5, the YMC can be halted through a combination of histone mutants: Fig 1.8 shows how some lysine-to-arginine mutants in the H3 and H4 tails allow for yeast metabolic oscillations, while others destroy them [127, Fig. 8]. No single K→R mutation is able to stop the YMC by itself, but different combinations of 4-5 mutations can destroy the synchronized oscillations (cells still being viable), potentially disrupting the cells' internal oscillator as well, so I hypothesize some HM permutations of the histone code [147, 247, 312–315] coordinate -at least partially- the YMC regulation.

The RNA-HM interplay will be addressed in the next Section by comparing the RNA dynamics (as measured by RNA-seq) with some particular HMs appearance at relevant moments of the YMC; this section is devoted to study HMs *per se*. If HMs are instructive for transcription [316], I would expect them to arise either earlier or at the same time as mature RNA is observed, but definitely not later [317]. ChIP-seq data consisting of the six HMs measured in Boeke *et al.* [127]: H3K4me$_3$, H3K9ac, H3K14ac, H3K18ac, H3K36me$_3$, and H3K56ac. Table 2.1 provides some information on these HMs: depositing enzymes and typical positions on genes. In the future, it would be interesting to explore HM YMC regulation via other covalent modifications; for example, phosphorylation [112,318,319], ubiquitylation [279,319,320], and crotonylation [321, 322] have been suggested as promising candidates for the YMC regulatory logic.

Table 2.1: Histone modifications deposition enzymes and typical position on the genes.

Modification	**Deposited by**	**Position on gene**
H3K4me$_3$	Set1 [323]	Active promoters near the TSS [242, 324]
H3K9ac	Gcn5 [325]	Peaks at TSS [326]
H3K14ac	Gcn5 [327]	Active gene promoters [328–330]
H3K18ac	Gcn5 [331]	Peaks at TSS [332]
H3K36me$_3$	Set2 [333]	Gene Body [334, 335]
H3K56ac	Spt10 [336]	All nucleosomes [337]

Several histones H3 and H4 post-transcriptional tail modifications (acetylations and trimethylations) are known to act as genes' epigenetic switches [247–249]. HMs

can both repress and facilitate the activation of many genes [147, 338]. Figure 2.1 reveals the interplay of RNA and H3K18ac dynamics during the YMC. An intuitive way of appreciating the time synchrony is that the order of the colors for H3K18ac and RNA is exactly the same, starting from the peak: in other words, RNA abundance is concurrent with H3K18ac accumulation. However, as we see in Figure 2.2, H3K4me$_3$ accumulates some time after the RNA, and the strict time order observed for H3K18ac in Fig. 2.1 is less strict in the case of both H3K9ac and H3K14ac. In both figures, it is a remarkable result that, for every given HM, curves corresponding to different times have entirely similar shapes, indicating that acetylation and deacetylation (or trimethylation and detrimethylation) affect yeast genes homogeneously across both the gene body and the promoter region. Another general feature observed in all the HMs studied is that changes in histone tail lysine acetylation levels are more abrupt than changes in trimethylation levels.

Conclusion Of the results discussed in the previous paragraph, the ones shown in Figs. 2.1 and 2.2 will be key in the following sections and chapters: for every HM, curves corresponding to different times have entirely similar shapes, indicating that acetylation and deacetylation (or trimethylation and demethylation) affect yeast genes homogeneously across both the gene body and the promoter region. This fact means that it is correct to represent the HM accumulation level on a given gene at a given time by its average value on the promoter + body region. By doing this, we have greatly simplified our analysis: each gene-time combination has one value, rather than the ~ 1000 values it would have if this was not the case.

2.1.1 Phenelzine and the YMC

The addition of phenelzine (PZ) to cycling cells does not destroy the YMC oscillations [108]. Phenelzine is a monoamine oxidase (MAO) inhibitor [339] widely used as an antidepressive [340, 341]. Crucially, H3K4 demethylation is a target for some antidepressive medications like PZ [342] and it affects the speed at which H3 and H4 lysine residues are acetylated and de-acetylated [343]. PZ is also an inhibitor of histone demethylases [344, 345] and blocks the degradation of numerous compounds [346]. In batch yeast, histone acetylation half-lives are as short as 3 minutes [347], while we observe in Figs. 2.2.b and 2.2.c that during the YMC some H3 lysine acetylations like H3K14ac or H3K18ac are active for temporal windows of over 30 minutes. It is interesting that the YMC can be maintained in the presence of PZ [108], since PZ addition significantly modifies the epigenetic landscape of HMs.

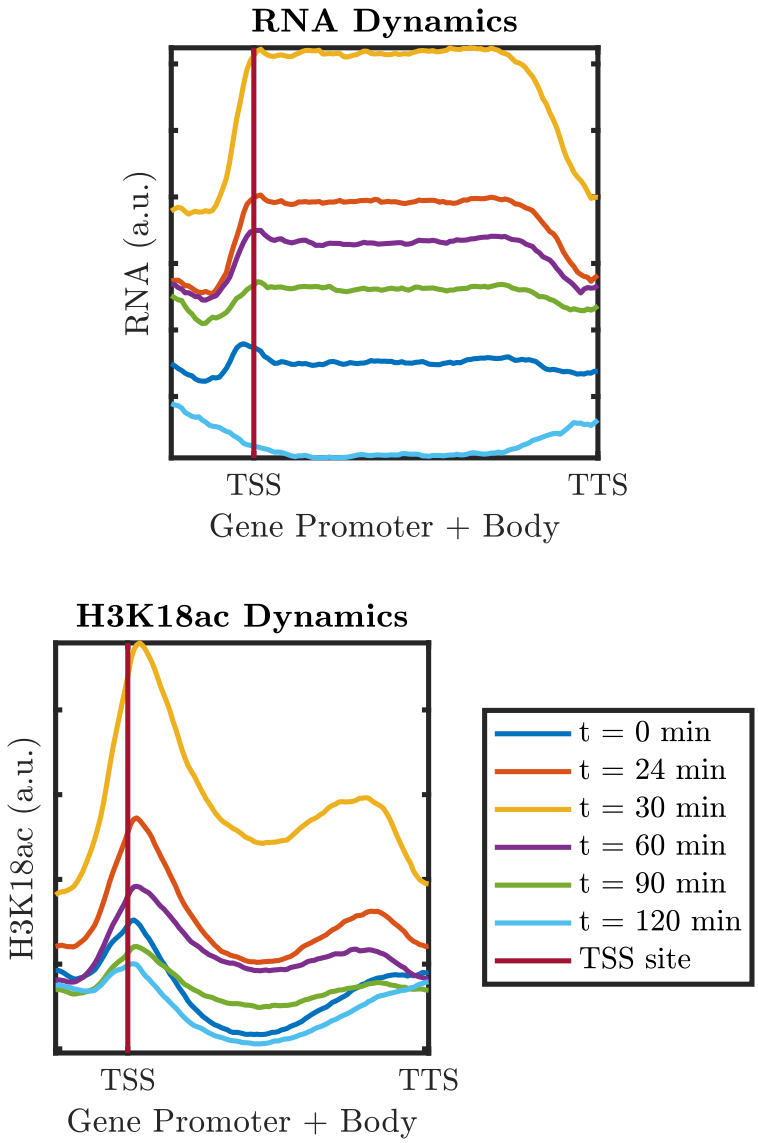


Figure 2.1: RNA and H3K18ac accumulation dynamics during the YMC for the OX cluster of genes. TSS: Transcription Start Site; TTS: Transcription Termination Site. x-axis: DNA meta-region ranging from 500 base pairs before the TSS (promoter region) to the TTS (gene body). y-axis: RNA and H3K18ac metagene profiles, expressed in arbitrary units. The metagene is defined as the average expression (in arbitrary units) at a given base pair of the genes included in the cluster; the procedure for calculating metagene is explained in Appendix A.4. Lines represent OX genes averaged RNA expression and H3K18ac accumulation levels at the YMC time indicated in the legend. These data refer to Section 2.1 and support the conclusion claim made in Subsection 2.6.2.

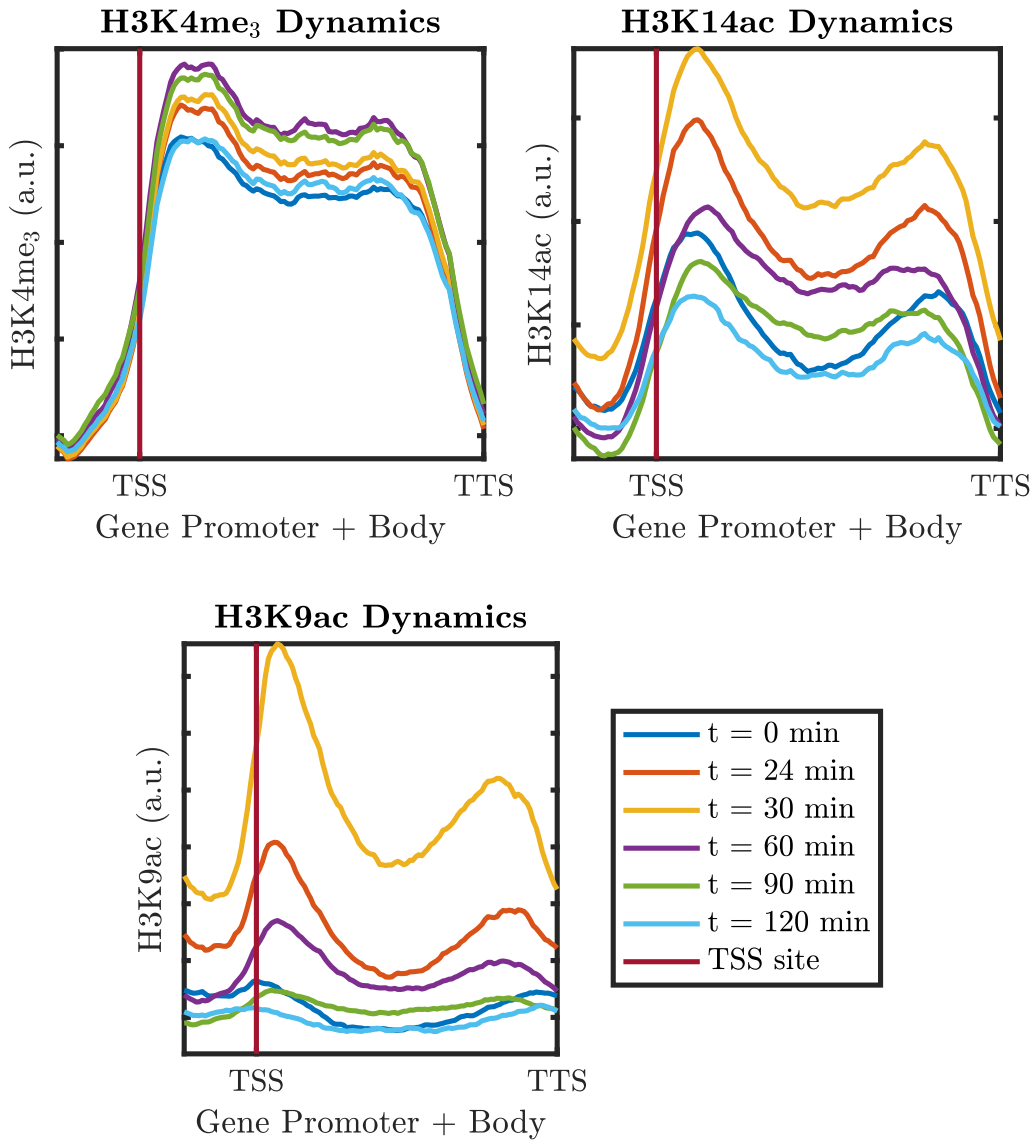


Figure 2.2: (Top Left) H3K4me$_3$, (Top Right) H3K14ac, and (Bottom) H3K9ac accumulation dynamics during the YMC for the OX cluster of genes. TSS: Transcription Start Site; TTS: Transcription Termination Site. x-axis: DNA meta-region ranging from 500 base pairs before the TSS (promoter region) to the TTS (gene body). y-axis: histone modifications metagene profiles, expressed in arbitrary units. Lines represent OX genes averaged HM accumulation levels at the YMC time indicated on the legend.

In the future, measuring H3K4 methylation and H3 lysine acetylation via ChIP-seq at some YMC points after PZ has been added to the media may reveal important differences in HM behavior, potentially eliminating some HMs that are in principle prospective YMC regulators, like H3K4me$_3$/H3K36me$_3$ and H3K9ac/H3K14ac/H3K18ac.

2.2 RNA and HMs are correlated across the YMC

The goal of this Section is to identify which HMs arise while transcription is taking place, and which ones appear at different times. To do this, I use ChIP-seq and RNA-seq data from the literature (discussed in Subsection 2.2.1), performing a bioinformatic analysis on them (Subsection 2.2.2) that will produce illustrative plots of the RNA-HM co-occurrence across the YMC (Subsection 2.2.3), and finally draw some conclusions from these results (Subsection 2.2.4).

2.2.1 ChIP-seq and RNA-seq data

The data used in this subsection are described in Appendix A.1. There are some limitations on the data quality that must be kept in mind: first of all, ChIP-seq and RNA-seq data were only sampled at 16 points during the YMC. If sample points were equally spaced, that would be one measurement every 14 minutes, which becomes -in principle- the highest time-resolution that can be achieved.

As explained at greater length in Appendix A.2, ChIP-seq data have a number of limitations. ChIP-seq reads are noisy [149, 348], even when compared to RNA-seq [349, 350]. If noise occurs nearby the peak in HMs accumulation, it may be necessary to use a filter [351, 352] before the peak can be retrieved. ChIP-seq techniques have low saturation points [246,353]; therefore, sometimes the signal intensity saturates for some time points, measuring maximum (or saturated) values across a number of time points.

The information provided by these data was condensed using gene normalizing windows in the form discussed in Appendix A.3. It is proved that the average over the region spanning from 250 base pairs (bp) upstream of the TSS to the Transcription Termination Site (TTS), [TSS-250 TTS], of the RNA or HMs expression/accumulation level is a representative metric for the measurement of overall RNA expression or HM accumulation level that accurately reflects temporal evolution within the gene.

This metric does not privilege longer genes, and rather than on intergenic HMs distribution, it is primarily concerned with capturing time evolution, which is the quantity this bioinformatics analysis and mathematical model are mostly concerned with.

Therefore, I condense the bp-resolution information provided by the datasets detailed in Section A.1 using the [TSS-250 TTS] region average. Mathematically expressed, this corresponds to the function

$$\begin{aligned} f &: \mathcal{G} \to \mathbb{R} \\ &\mathcal{G}_i \mapsto \texttt{RM}([-250 \;\; \text{TTS}]_i), \end{aligned} \tag{2.1}$$

where $\mathcal{G}$ is the set of the 5770 yeast characterized genes $\mathcal{G}_i$ and `RM` is the values mean of the specified interval.

2.2.2 Methods

To overcome the issues raised above I calculate, for each HM and each of the 5770 verified yeast genes, the Pearson correlation coefficient r [354] between the RNA and the HM signals for that gene. I have 16 pairs of RNA-HM values, and the r resulting from these 16 pairs is the gene-associated value. The idea behind this matching algorithm is that, for example, functions $sin(\theta)$ and $sin(\theta + \phi)$ have a higher Pearson correlation coefficient the smaller the phase ϕ is, and ϕ is a perfect measure of the inter-peaks time lag: even when precise information about the peak location cannot be ascertained because of the causes discussed before, ϕ is an excellent metric.

The second step is to plot these 5770 r values as a histogram, as shown in Figure 2.3, to explore how are these r values distributed in the $[-1, 1]$ interval. High $r \sim 1$ correlation values imply that the HM and RNA are co-occurring, but we can see in Fig. 2.3 that this is seldom the case, and this correlation measure, by itself, cannot individuate either the first peak, or the inter-peaks time lag. We notice a striking difference between the H3K9ac and H3K56ac distribution functions: while H3K9ac is in general positively correlated with RNA appearance, H3K56ac distribution is not significantly different from the distribution of a random scatter plot, suggesting that **H3K56ac is probably not related to the YMC dynamics.**

H3K56ac is mainly involved in DNA replication [355–358], and only approximately 20% of the cells in the chemostat are not cycling [359], therefore this result

is consistent with the knowledge in the field about H3K56ac biological functions.

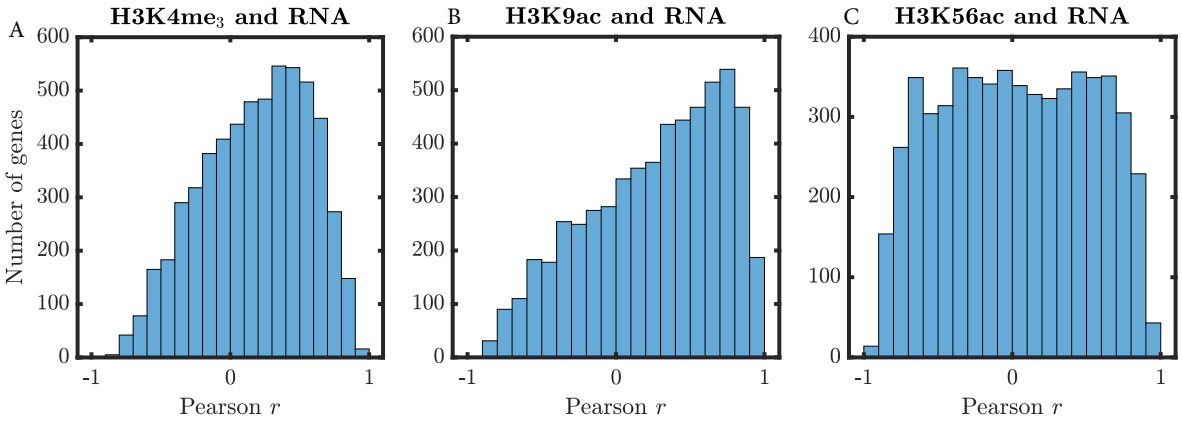


Figure 2.3: Histograms showing the Pearson correlation coefficient distribution between RNA and (a) H3K4me$_3$; (b) H3K9ac; (c) H3K56ac for individual genes across the YMC. Both RNA and HM values are taken as the average value over the gene promoter + body region, as discussed in Subsection 2.2.1. In the case of (c), the distribution obtained is not significantly different from the correlation obtained between RNA and randomly generated data, supporting the claim that H3K56ac is probably not related to the YMC dynamics.

In order to obtain an accurate measure of the time lag between the peaks in RNA and a particular HM, I start by interpolating both HM and RNA values using the MatLab cubic spline interpolation in order to have functions that are continuous in t. I calculate the correlation coefficient r between RNA and the HM when a delay τ is applied to the HM time[2], aiming to maximize the correlation coefficients distribution. For example, when exploring $\tau \pm 20$ min values I obtain Fig. 2.4.

2.2.3 Results

We observe that the high correlation region $r > 0.8$ is more densely populated after a time shift of $\tau = +20$ min for H3K4me$_3$, as shown in Figure 2.4.a, while in the case of no time shift ($t = 0$ min), this region is the most densely populated for H3K18ac, as shown in Figure 2.4.b. I scan different τ values in a -1 h to $+1$ h interval in 1 minute steps, obtaining that for H3K4me$_3$, a $\tau = +15$ min maximizes the histogram

[2]Instead of $HM(t)$, using $HM(t+\tau)$, the HM accumulation function evaluated time τ afterwards.

integral in the $r > 0.8$ region, while H3K18ac is maximized in the $\tau = 0$ case.

Similar calculations for the other four HMs resulted in $\tau = +18$ min as the maximizer value for H3K36me$_3$, and the other H3 lysine acetylations (H3K9ac, H3K14ac, H3K56ac) exhibited $|\tau| < 5$ min maximal values. Since, as I stated at the beginning of the section, the time-resolution of the experiment is approximately 15 minutes, I can take all the lysine acetylations to co-occur with transcription, at least at this level of available resolution.

2.2.4 Conclusion

Therefore, I can conclude that H3K4me$_3$ and H3K36me$_3$ are deposed approximately 15 minutes after the RNA transcripts appear. In short, this means that H3K4me$_3$ is almost certainly not instructive for transcription. H3K18ac coordinates with RNA, so it is a good candidate for being a transcription instructor, but co-occurrence can either mean cause or consequence [316, 360–363].

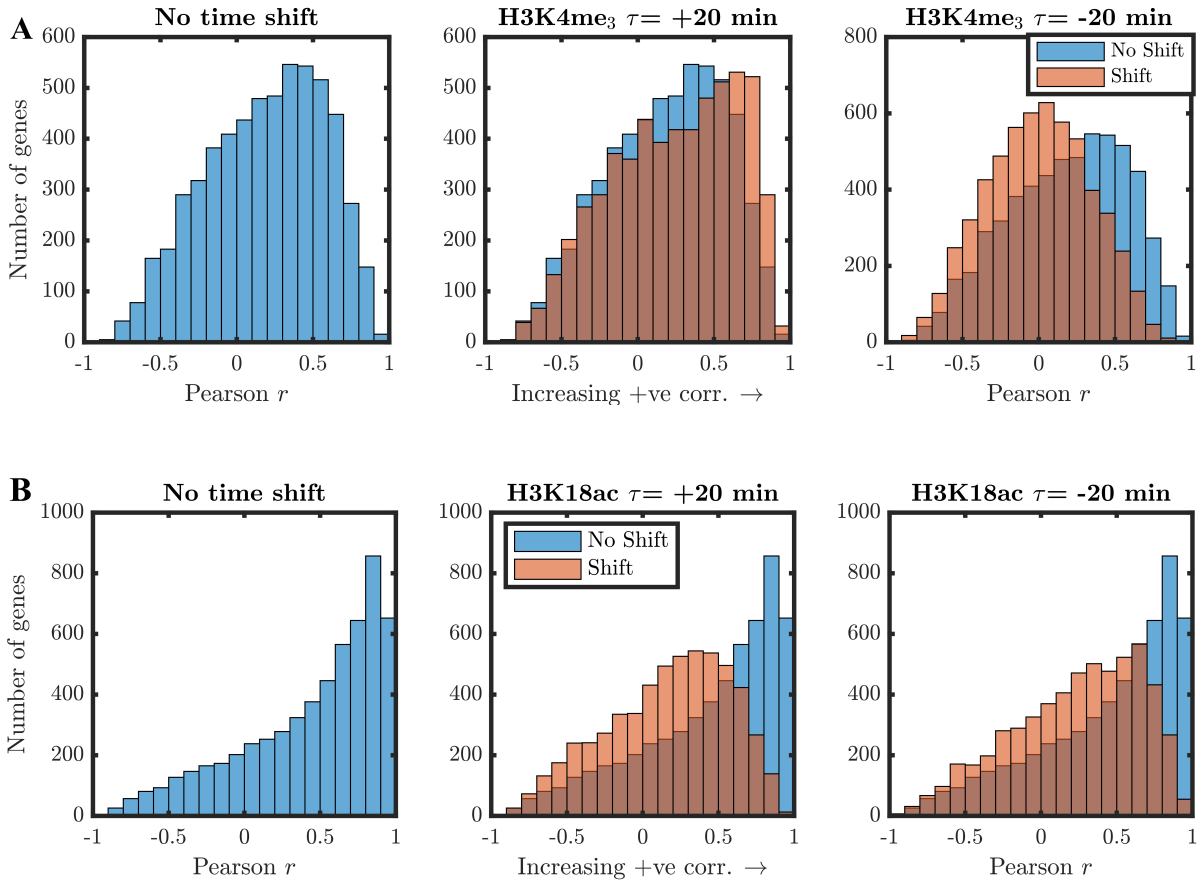


Figure 2.4: Correlations distribution between RNA and (a) H3K4me$_3$, (b) H3K18ac dynamics under three of the time shifts discussed in Subsection 2.2.3 for the 5770 yeast genes at time shifts $\tau = 0$, $\tau = +20$ min, and $\tau = -20$ min. When H3K4me$_3$ experiences a $\tau > 0$ lag, some correlation coefficients are shifted to higher r values, while for H3K18ac, the best distribution is achieved at $\tau = 0$. To facilitate comparison between the shift and the no-shift distributions, the $\tau = 0$ histogram is plotted in the background of the other two graphs.

2.3 mRNA degradation: are genes regulated at the point of synthesis or at the point of turnover?

This Section is devoted to investigating the phenomenon of mRNA degradation in yeast in general, and during the YMC in particular. My goal is to establish a relationship between the degradation processes and rhythms on one hand, and genetic regulation via transcript abundance on the other. The results of the mRNA degradation calculations performed in this Section will be used in Chapter 4 to formulate the mathematical equations describing RNA concentration dynamics through the YMC.

Since mRNA decay rates were first measured in yeast [364, 365], the interplay between mRNA production and mRNA degradation as a transcript's quantity regulation mechanism has been widely discussed [366, 367], and P-bodies have been proposed as key regulators of the production-degradation dynamic equilibrium [368]. It is worth noting that acute glucose starvation induces global mRNA stabilization while increasing ribosomal protein mRNA degradation [307].

The main caveat in mRNA decay experiments is that results are highly dependent on the protocol employed. For example, it has been found that the yeast mRNA degradation rates measured by [307] and [306] have an astonishing low ($r = 0.01$) correlation coefficient [305]. In particular, measuring mRNA degradation rates via gene expression upon transcription inhibition [306,308,369] results in completely different results to measuring mRNA degradation rates via pulse-chase RNA labelling protocols [307,370–372]. The immediate conclusion is that mRNA degradation times in yeast, while known to range from one minute to one hour, cannot be determined for individual genes with acceptable accuracy.

Figure 2.5 shows the half-lives for yeast genes as measured by Young *et al.* [370], grouped by the YMC cluster they belong to. It is observed that OX genes tend to have shorter half-lives than RB genes, whose half-lives in turn tend to be shorter than RC genes. For this analysis, I considered using the more recent Farina *et al.* [305] dataset, but it only provides data on 939 genes, of which 80 are OX, 400 RB, 140 RC, and the rest are non-cycling. This high number of RB genes probably reflects the fact that RB genes are more likely regulated via turnover, chiefly through Puf3p protein phosphorilation [295], increasing the reliability of the results obtained for this cluster. The Farina *et al.* dataset distribution functions revealed short-lived OX genes and similar profiles for the RB and RC clusters (data not shown). The precise mathematical formulation of the processes underpinning the mRNA degra-

dation machinery are not explored here; such analysis will be deferred to Subsection 4.4.2, where it will be shown to fit the proposed model based on ordinary differential equations.

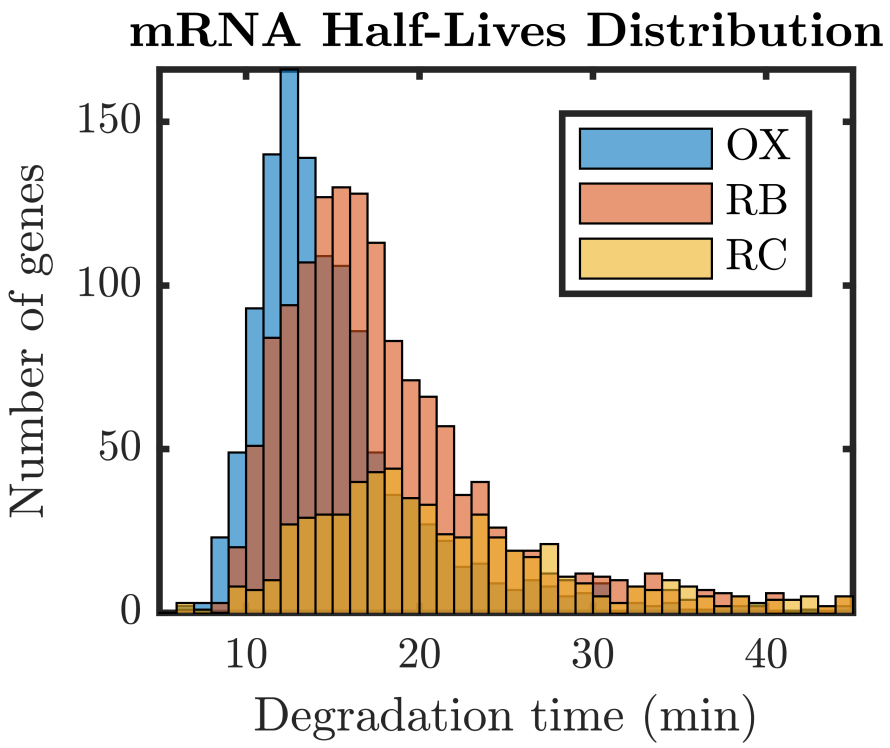


Figure 2.5: Histogram of mRNA half-lives distribution for Young *et al.* [370] in the three YMC clusters of genes (Oxidative, Reductive/Building, and Reductive/Charging). Each of the three distributions has different average and standard deviation. Bin widths are 1 minute.

Based on the results of Table 2.2, I estimate an average mRNA half life of **16.1 minutes**, which is the average of both results, with a standard deviation of **9.6 minutes**, data that will be used in the following Sections. I assume that mRNA degradation distribution is analogous in the YMC and in the batch cultures of [305, 370].

Table 2.2: mRNA degradation data: comparison of Young *et al.* [370] and Farina *et al.* [305]. All data are given in minutes. μ, σ are the average and standard deviation for all genes, max/min is the data range, and n is the number of measured genes.

Experiment	$\boldsymbol{\mu}$	$\boldsymbol{\sigma}$	**max.**	**min.**	$\boldsymbol{n}$
Young *et al.* [370]	18.8	9.8	116	5	6179
Farina *et al.* [305]	13.4	9.4	50	1.9	939

2.4 mRNA degradation calculation

In this section I aim to calculate the average mRNA degradation time for yeast genes at each of the three YMC phases. I do this by using the RNA-seq data from [127] and the NET-seq data [373] discussed in Annex A.1, in their gene-resolution form as described in Annex A.3.

RNA-seq measures the levels of transcripts [235], while NET-seq measures the density of RNA polymerase [241], which is related to the rate at which transcripts are being generated [374]. It follows that, for every gene i, I can make the approximation:

$$\dot{R}_i(t) \approx N_i(t) - D_i(t), \tag{2.2}$$

where $R_i(t)$ is the RNA-seq value measured for gene i at time t, $\dot{R}_i$ its time derivative, N_i is the NET-seq value, and D_i is the rate at which transcripts are being degraded.[3]

The goal of this Section is to obtain the value of the degradation rates D_i, since both R_i and N_i are known. If RNA- and NET-seq had defined and comparable physical units, this would be straightforward: I could set in Eqn. (2.2) the units of R_i as $[R_i] = \frac{\#\ mRNAs}{cell}$ and similarly $[N_i] = \frac{\#\ mRNAs}{cell \cdot minute}$.

Unfortunately, the use of different sequencing technologies with different sequencing depths and sensibilities (see Appendix A.2) means that both R_i and N_i have arbitrary units, and D_i cannot be directly calculated from Eqn. (2.2) as $N_i - \dot{R}_i$; some more refined form of mathematical analysis is required to learn how $\dot{R}_i$ and N_i compare gene by gene, a task performed in the next Subsection.

[3] Since data were obtained for 12-16 timepoints and Eqn. (2.2) is continuous in time, R_i and N_i are calculated using the cubic `spline` MatLab interpolation function [375, 376].

2.4.1 Methods

As we have just discussed, since both RNA and NET-seq are expressed as mean read densities in arbitrary units [241], $\dot{R}_i$ and N_i cannot be readily compared; it is necessary to make some assumptions about their dynamics that will allow us to normalize, and then compare, both magnitudes. My goal in this Subsection is to find a way of setting common units for $\dot{R}_i$ and N_i that will let us find the degradation rate D_i via Eqn. (2.2).

N is a matrix of positive elements, where $N = N_{i,t}$ is the NET-seq measurement at time t for gene i, while $\dot{R} = \dot{R}_{i,t}$ elements can be either positive or negative. Therefore, a simple elements-average calculation cannot be performed. I calculate the standard deviation σ for each gene in both R and N, and take the average across time for each of them. This way, I can calculate the average across-genes standard deviations ratio

$$\eta = \frac{Avg(\sigma(\dot{R}))}{Avg(\sigma(N))}, \tag{2.3}$$

which is found to be $\eta = 15.7$. The η parameter states that, on average, the arbitrary units on which $\dot{R}$ is expressed have a standard deviation which is 15.7 times larger than the standard deviation in matrix N. The interpretation of this result is that, because both magnitudes should have very similar standard deviation, I can make the assumption that RNA-seq Illumina HiSeq 2000 readings intensity is 15.7 times larger than the NET-seq Agilent BioAnalyser High Sensitivity readings intensity. Error analysis on the $\eta = 15.7$ value is provided at the end of this Subsection.

I then make the transformation of the (dimensionless) matrix $N \mapsto \eta \cdot N$ to have it commensurable with $\dot{R}$, so that both magnitudes can be compared and I can build the dimensionless matrix

$$D = \eta \cdot N - \dot{R}. \tag{2.4}$$

After this transformation, for each gene (row of D), I find the corresponding coefficient of variation $c_V = s/\bar{x}$ and rank the 5770 identified yeast genes from lowest to highest c_V value. Finally, I perform Gene Ontology (GO) term enrichment analysis on this ranked list of genes using GOrilla, a web-based application that identifies enriched GO terms in ranked lists of genes, without requiring the user to provide explicit target and background sets [377, 378].

GOrilla employs a flexible threshold statistical approach to discover GO terms that are significantly enriched at the top of a ranked gene list [378]. Other Gene

Ontology Consortium [379] related projects, chiefly the SGD-integrated GO Term Finder[4] [380], but also others like AmiGO [381], FatiGO [382], BiNGO [383], and the Comprehensive Yeast Genome Database (CYGD) [384] were also explored, without very different results, since the preliminary coverage of annotations approaches 100% for yeast [385].

It is important to note that there exists a risk of over-reliance on GO calculators' results [386]. For this thesis, the main concern when assessing GO results is that both Gene Ontology Consortium and CYGD databases refer to studies performed in batch culture yeast, a metabolic state that is well-known. However, YMC biological processes are known to be different from batch culture yeast processes [136, 138]. A second, minor, concern, is that over 40% of the yeast genome relates to genes without any relationship to the YMC [127]; these genes either do not share the YMC oscillations in expression, or are plainly concerned with entirely different biological processes.

Because of the propagation of errors in these calculations, I do this for both **(i)** all the yeast genome and **(ii)** only the genes described by [127] as cycling genes. Results are scaled so that mRNA degradation half-lives follow a distribution time of 16.1 ± 9.6 minutes, as obtained from Table 2.2.

Error analysis on η

As we discuss in Appendix A.2, NET-seq data are significantly more noisy than RNA-seq [387] and, because of the smaller size of the fragments imputed into the PCR, they contain more errors in the sequencing reading and mapping. Therefore, it is possible that:

- the $\sigma(N)$ value imputed in Eqn. (2.3) has been overestimated, if sequencing noise in the N measurements artificially increases $\sigma(N)$. Therefore, η would be underestimated (it is imputed in Eqn. (2.3) as the fraction denominator).
- $\sigma(N)$ has been underestimated, if the reading and mapping errors ('Inf' and 'NaN' inputs) in N are all set to zero and this makes $\sigma(N)$ artificially low.

We discuss the order of magnitude of these error sources in order to provide a confidence interval for the η parameter. From inspection of the raw data, NET-seq reading and mapping errors are present in 316 genes, 5.5% of the total. Therefore, it

[4]https://www.yeastgenome.org/cgi-bin/GO/goTermFinder.pl

is possible that $\sigma(N)$ could have been underestimated by $\approx 11\%$, since error propagation in σ is $\sim$2 larger than error in the data [388].

Noise inflating the individual genes standard deviation is more difficult to calculate. I obtained a rough estimate by dividing the genes into quintiles ranked by their individual standard deviation (a proxy variable for measurement noise) and obtained η values ranging from 11.5 to 21.3 (other three values were 17.4, 18.0, and 18.2), resulting in a standard deviation of 3.45 (or 0.23η).

Errors addition is calculated as the square root of the sum of the errors, therefore $\varepsilon_{tot} = \sqrt{0.11^2 + 0.23^2} = 26\%$ and we set $\eta = 16 \pm 4$.[5]

2.4.2 Results

The full yeast genome results GO Terms are displayed in Table 2.3. The most relevant term is proteasome-mediated ubiquitin-dependent protein catabolic process, which is linked to another 8 significant p-value ($p < 10^{-9}$) catabolic process terms. This result confirms that the processes in Table 2.3 are indeed associated with degradation, and their constant value will be used in Section 2.4 for calculating the clusters' degradation rates.

To identify the precise proteins underpinning the mechanisms regulating this transcriptional activity, I use the YEASTRACT repository [389–393] to calculate the relevant transcription factors (TFs) on the same set of genes. YEASTRACT (YEAst Search for Transcriptional Regulators And Consensus Tracking) is a well established, online, peer-reviewed repository containing most of the associations between TFs and yeast genes published in the field literature. It does not contain novel experimental information; it should be rather thought of as a library of all the TFs-genes regulatory associations that have been identified in yeast. Therefore, it should be highly reliable for our analysis of such interactions. I considered the use of other consensus TF-retrieving tools, such as TRANSFAC [394–396], TRRD [397, 398], or COMPEL [399, 400], but YEASTRACT is yeast-specific, regularly updated, and comprehensive, therefore it is more suitable for this analysis. Furthermore, literature not connected to the Teixeira group has praised the YEASTRACT repository, in particular for alleviating the problem of data scarcity [401, p. 552] and improving data reliability [402, p. 467], [403].

[5] No decimals included because of the error size, but I still use 15.7 rather than 16 for calculations.

The YEASTRACT associated TFs with significant p-value ($p < 0.01$) were: Gcr1, Abf1, Cup9, Reb1, Hap2, Tye7, Gal4, and Mcm1. Flo8 appeared in the $0.01 < p < 0.05$ p-values band. Gcr1 is a strong repressor, as we can see in Figure 2.6.a; Hap2 is in general an activator (Figure 2.6.b); and Reb1 is a repressor more moderate than Gcr1, as shown in Figure 2.7. Other TFs repress some genes and enhance some others, but the overall picture is far less clear, as we can observe in Figure 2.7.

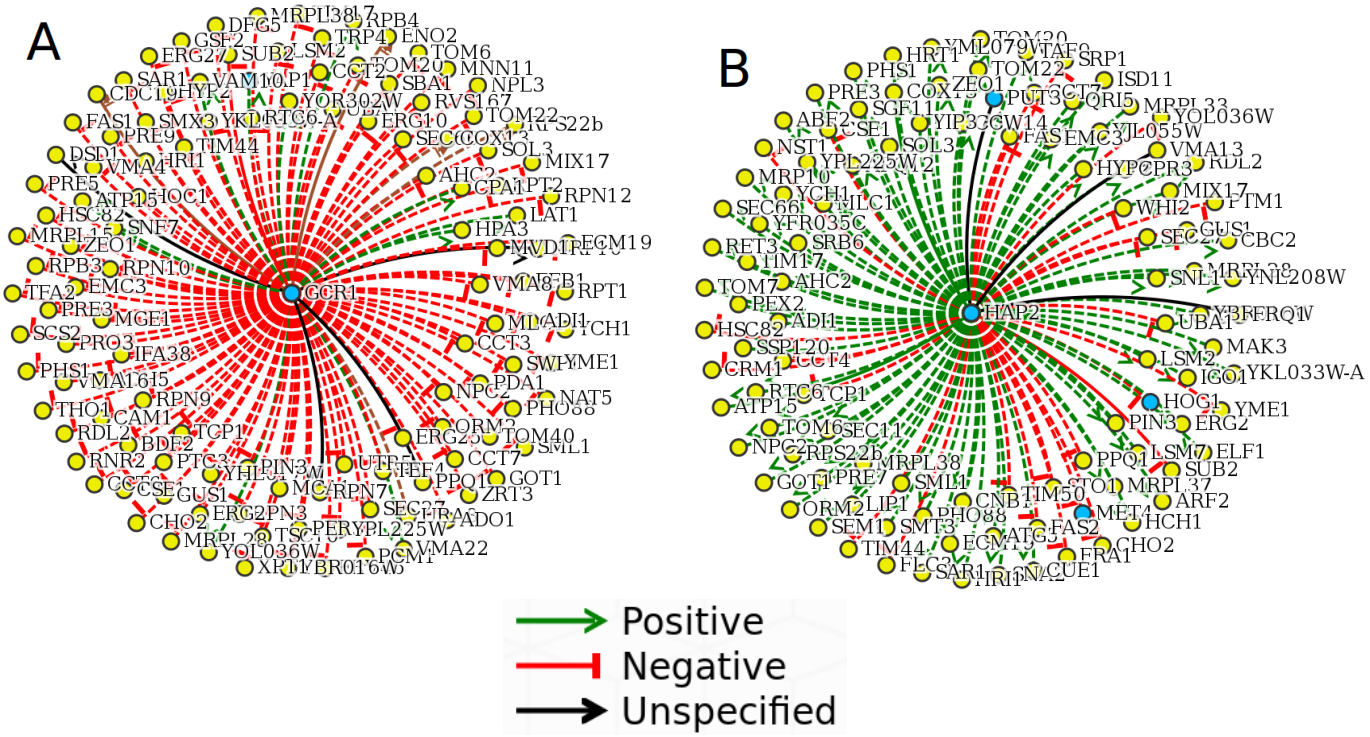


Figure 2.6: YEASTRACT Transcription Factor influence on gene expression for (a) Gcr1p and (b) Hap2p, two of the TFs most closely linked with the 300 genes with smallest c_V in the yeast genome. Green arrows $\rightarrow$ represent positive regulation, red arrows $\dashv$ represent negative regulation (repression). Yellow dots represent regular yeast genes, blue dots are TF genes.

When performing the same calculations only on the cycling genes as described by [127], the GO analysis results did not identify processes with the same confidence

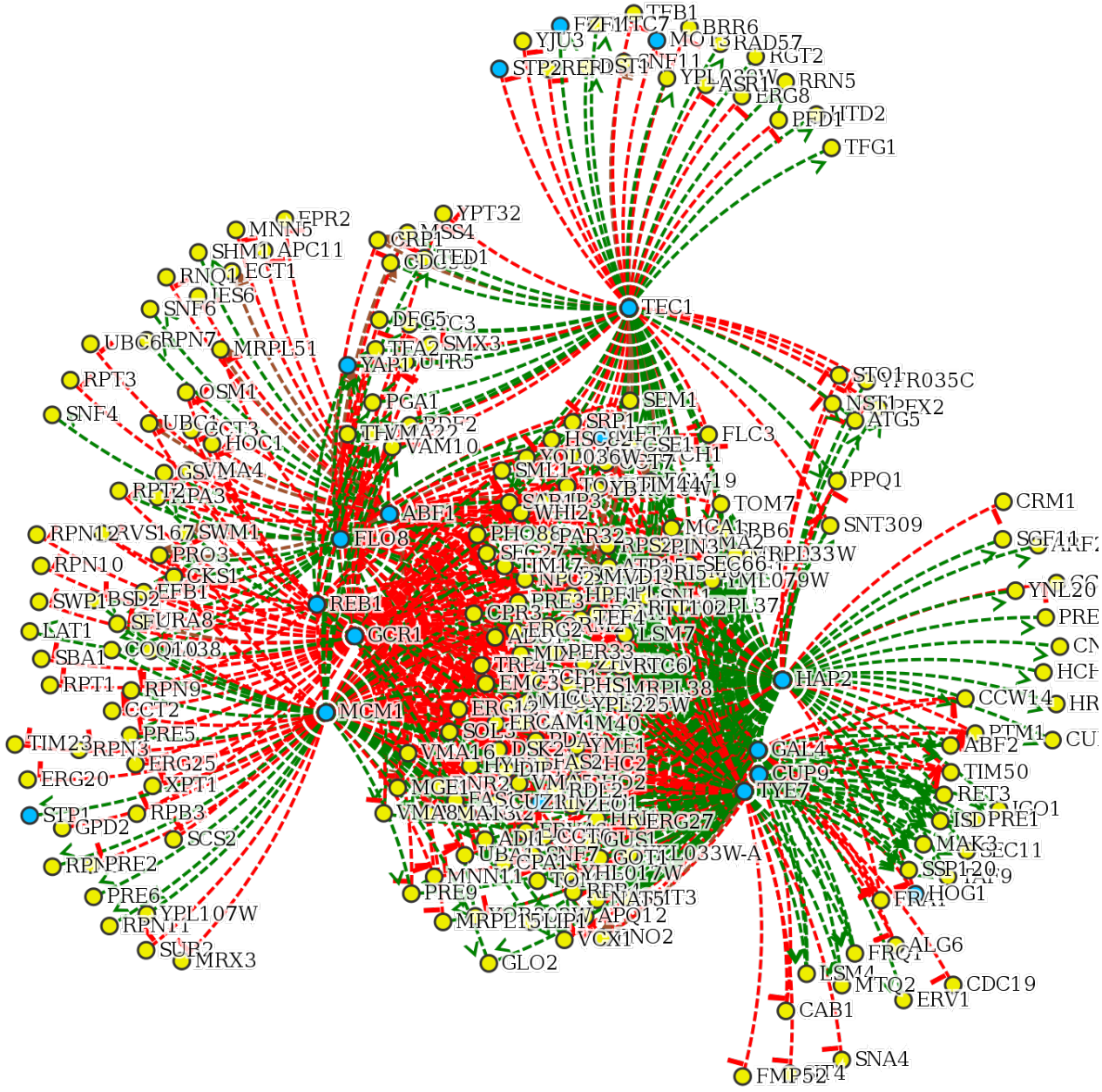


Figure 2.7: Full regulatory tree for the 300 genes with smallest c_V in the yeast genome. Yellow dots represent regular yeast genes, blue dots are TF genes. Green arrows $\rightarrow$ represent positive regulation, red arrows $\dashv$ represent negative regulation (repression). Unspecified regulations are not shown to improve readability. The overall structure of the regulatory mechanism is more complex than the one TF cases discussed in Figure 2.6.

Table 2.3: Most significant GO terms associated with least variable mRNA degradation (lowest c_V values) during the YMC, as inferred from the transcript levels of activity.

GO term	**Description**	p**-value**	**FDR**
GO:0043161	proteasome-mediated ubiquitin-dependent protein catabolic	1.99E-14	1.03E-10
GO:0006511	ubiquitin-dependent protein catabolic process	5.82E-14	1.51E-10
GO:0019941	modification-dependent protein catabolic process	9.13E-14	1.58E-10
GO:0010498	proteasomal protein catabolic process	1.05E-13	1.37E-10
GO:0030163	protein catabolic process	2.18E-13	2.27E-10
GO:0043632	modification-dependent macromolecule catabolic process	1.10E-12	9.51E-10
GO:0051603	proteolysis involved in cellular protein catabolic process	2.50E-12	1.85E-09
GO:0044265	cellular macromolecule catabolic process	1.11E-10	7.23E-08
GO:0043248	proteasome assembly	3.77E-10	2.18E-07
GO:0009057	macromolecule catabolic process	2.05E-09	1.07E-06
GO:0006508	proteolysis	2.08E-09	9.84E-07
GO:1901565	organonitrogen compound catabolic process	5.30E-09	2.29E-06
GO:0010499	proteasomal ubiquitin-independent protein catabolic process	1.60E-08	6.38E-06
GO:0071806	protein transmembrane transport	2.81E-08	1.04E-05
GO:0030150	protein import into mitochondrial matrix	7.35E-08	2.55E-05
GO:0044743	protein transmembrane import into intracellular organelle	1.10E-07	3.58E-05
GO:0065002	intracellular protein transmembrane transport	1.41E-07	4.30E-05
GO:0070585	protein localization to mitochondrion	2.68E-07	7.72E-05
GO:0072655	establishment of protein localization to mitochondrion	2.68E-07	7.32E-05
GO:0006368	transcription elongation from RNA polymerase II promoter	3.84E-07	9.97E-05

as Table 2.3 did on the previous analysis. The GOrilla software suggests [378, p. 6] a $p < 10^{-3}$ threshold for considering an outputted process as reliable, and very few of the retrieved terms had such a confidence level. A few catabolic processes as well as highly generic terms (like organonitrogen compound or protein catabolic process) were identified, but a coherent picture could not be formed (data not shown). Similarly, the analysis run on the genes with largest c_V values did not reveal any significantly GO-enriched term. One possible explanation of this result is that the additional constraint placed by the requirement of genes belonging to one of the clusters in Boeke *et. al.* [127] has excessively narrowed the possible GO combinations; in fact, if a few genes associated with a particular GO term are absent from the imputed data, it is very unlikely that the analysis will yield the relevant GO term with high confidence (i.e. low p-value) [377].

I use the yeast genome clustering produced by Boeke *et. al.* [127] to calculate the average mRNA degradation times for each of the three YMC phases (distributions shown in Fig. 2.5; average values shown in Table 2.4). We know from data in Table 2.2 that mRNA half-lives have an approximate standard deviation to average (σ/μ) ratio of 0.5 [370] or 0.7 [305], therefore these values should have a standard deviation of approximately 7 minutes.

Table 2.4: Average mRNA half-lives (in minutes) for the different YMC phases clusters as defined by Boeke *et al.* [127]. In all cases, the standard deviation is $\sigma = 9.4$ min, as per Table 2.2.

YMC Phase	OX	RB	RC
Half-life	12.3	16.5	14.6

An extension of the degradation rate calculation

The purpose of Eqn. (2.2) was to identify genes whose degradation rate $D_i(t)$ is approximately constant in time, a result corresponding to the steady-state reached when the mRNA degradation machinery is saturated. From RNA-seq analysis on such genes, we learn that these genes are active for the whole YMC duration, exhibiting significantly high values (first quintile) of gene expression in the YMC (data not shown).

mRNA degradation follows an exponential decay in both batch yeast [307, 404–

408] and other species [409–415]; assuming this is also the case in the YMC, for each gene i, mRNA molecules (whose proxy variable is RNA-seq data $R_i(t)$) degrade at a constant pace λ_i: $D_i(t) \approx \lambda_i \cdot R_i(t)$, where $D_i(t)$ is the rate at which transcripts are being degraded –as in Eqn. (2.2)– and for each gene i we can write

$$\dot{R}_i(t) = N_i(t) - \lambda_i \cdot R_i(t), \tag{2.5}$$

where $N_i(t)$ represents nascent transcription for gene i at time t (measured by its proxy variable NET-seq values), and λ_i is the gene-specific mRNA degradation constant. Other functional forms for Eqn. (2.5), like bi-exponential or multi-exponential processes, lack a clear biochemical meaning in this particular case, and do not exhibit a good fit for the data (result not shown). Using the degradation matrix D constructed in Subsection 2.4.1, the next step is to solve the linear algebra optimization problem

$$D = \boldsymbol{\lambda} \cdot R, \tag{2.6}$$

where $\boldsymbol{\lambda}$ is the vector of degradation constants $(\lambda_1, \ldots, \lambda_{5770})$ of the 5770 characterized genes in yeast and matrices D and R are as discussed in Subsection 2.4.1.

In the absence of noise or time-dependent fluctuations, the calculation of λ_i for each gene would be trivial, since $\lambda_i = D_i/R_i$ would be a constant. However, since the D_i/R_i ratio is time-dependent, we are interested in examining genes whose D_i/R_i ratio shows the lowest coefficient of variation $c_V = s/\bar{x}$ time dependency, and repeat hence the process described in Subsection 2.4.1. The results obtained for GO-enrichment are similar, and the most relevant GO term is (for the top 300 scoring genes, with a $p = 0.08$ p-value) protein transport to vacuole involved in ubiquitin-dependent protein catabolic process via the multivesicular body sorting pathway (GO:0043328), the process of directing proteins towards the vacuole that contributes to protein catabolism via the multivesicular body pathway [416]. This is an interesting result, since this biological process has recently been shown to be related to the Cdc48 ATPase in yeast [417], a cell division cycle regulator [418,419] and key element or 'power machine' in yeast protein degradation [420]. The component-GO (as opposed to the normal process GO studied) revealed some $p < 0.01$ p-value terms, like proteasome storage granule, cytosolic proteasome complex, or proteasome complex (data not shown).

2.4.3 Conclusion

Since Eqn. (2.2) resulted in more homogeneous results than Eqn. (2.5), it can be concluded that most catabolic and other degradation processes occur in fact in saturated media, where degradation is stronger and, for a given gene, degradation occurs at a constant rate. This result will be used to construct the YMC single-cell mathematical model of Chapter 4 to formulate the differential equations describing mRNA degradation, since it means that the mRNA degradation terms associated with the mRNA concentration in each of the three YMC phases can be approximated as constants, using the numerical values provided by Table 2.4.

The part of the analysis concerned with genes exhibiting the highest c_V values in Eqn. (2.5) did not retrieve any significant GO terms (data not shown). In my opinion, this result stems from the fact that the available Boeke data [127] have been processed: they are centred on the mean and normalized, which means variations in mRNA levels are exaggerated. In reality, these levels stay approximately constant, so Eqn. (2.3) should be a good approximation to processes in which the approximation of Eqn. (2.2) holds true, since they correspond to genes active during all of the YMC, and therefore with constant mRNA levels across the YMC. An alternative hypothesis is that almost all processes are described by Eqn. (2.5) except for the very few GO terms associated with catabolic and protein degradation processes. This hypothesis is supported by the fact that c_V values for Eqn. (2.5) were significantly smaller than for Eqn. (2.2).

2.5 mRNA degradation is related to some HMs accumulation levels

The goal of this Section is to ascertain whether the half-life estimations calculated in Table 2.4 correlate with HMs accumulation levels, particularly H3K18ac, across the YMC, by using both the Young *et al.* [370] and Farina *et al.* [305] measurements just discussed.

Figure 2.8 shows the correlation, for different clusters of genes, between H3K18ac and estimated half lives for each of the genes, observing how in general, H3K18ac is incompatible with stable transcripts ($r = -0.31$). Combinations are more or less frequent for different clusters of genes, but a sharp difference between the three datasets could not be observed. It is remarkable to note that *a high histone lysine acetylation*

level for one gene is **incompatible** *with a long half-life, or transcript stability, of the corresponding mRNA.* I arrive to this conclusion after examining the Farina *et al.* [305] mRNA degradation data and Boeke lab H3K18ac data. The same conclusion is retrieved when the Young *et al.* [370] mRNA degradation dataset is employed, and results for H3K9ac and H3K14ac are entirely similar (not shown).

The next part of the analysis consists in comparing the half-lives values obtained previously with some HMs data (H3K4me3, H3K9ac, H3K14ac, H3K18ac, H3K36me3, and H3K56ac data were used), exploring what the optimal shift is in HMs that maximizes the Pearson correlation coefficient r between both datasets to find the optimal lag, an algorithm which is analogous to the one presented in Section 2.2. Some studies have focused only on the relative position of the maxima in both datasets [305], but since both mRNA degradation [305] and HM measurements [387] are subject to considerable levels of noise (see Section 2.4 for mRNA degradation noise and Appendix A.2 for HMs measurement –ChIP-seq– noise), the best phase lag fit will be the one that maximizes the r coefficient between both datasets, disregarding individual genes.

HMs values were shifted in a time window of $t = \pm 75$ min, in one-minute resolution intervals. The rationale for choosing this time window is that, since the YMC has a 218 minutes period, phase shifts close to 180° (i.e. 109 min) give a loose sense of causality: if they are on anti-phase, HMs coming after degradation is as likely as the other way around. Figure 2.9 shows the results of the phases shifting calculations, while Table 2.5 displays the precise numerical values.

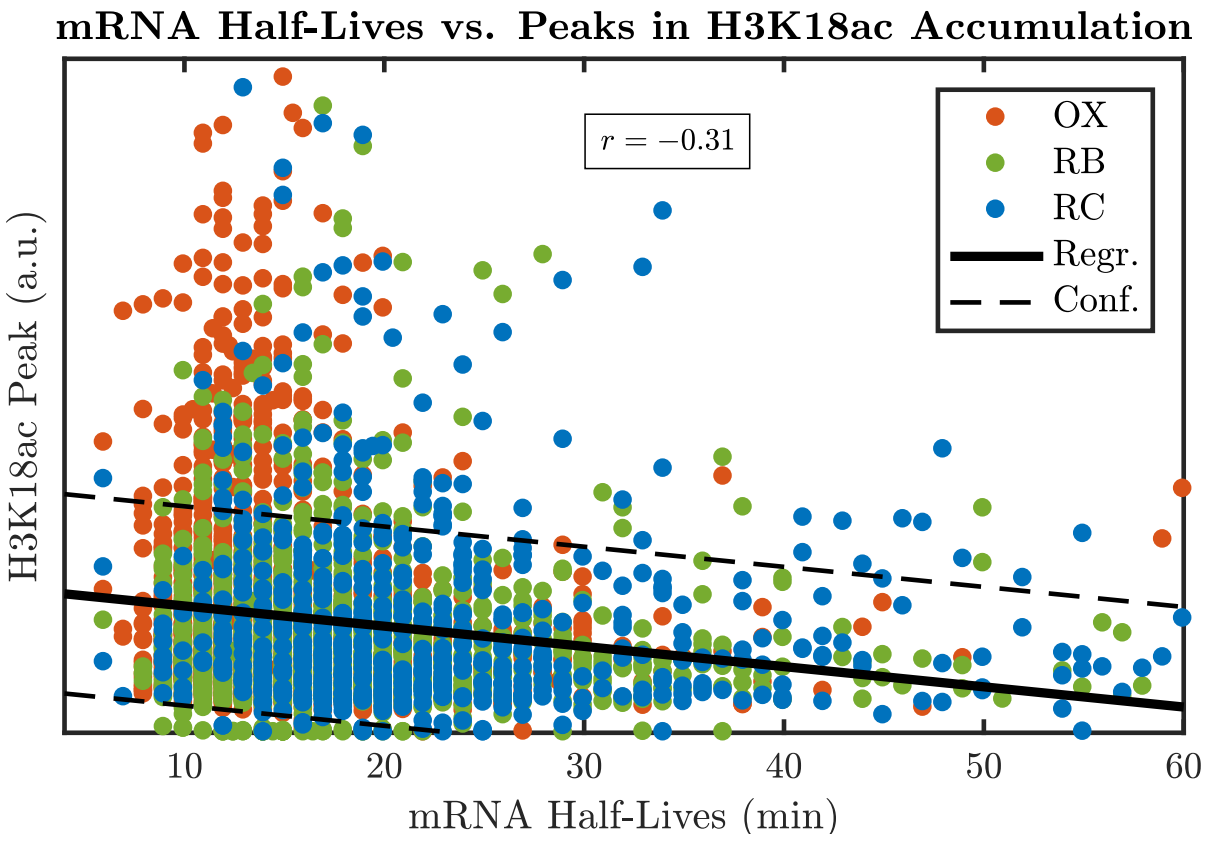


Figure 2.8: Scatter plot of H3K18ac accumulation peak compared with mRNA degradation time; data: [305]. Solid black line: Linear regression of the data; dashed lines: $\alpha = 0.15$ confidence intervals. Red dots: genes in the OX cluster; green dots: genes in the RB cluster; blue dots: genes in the RC cluster. The scatter plot exhibits a $r = -0.31$ Pearson correlation coefficient between both magnitudes. Note that high histone H3 lysine 18 acetylation levels are incompatible with long half-lives, or transcript stability, of the corresponding mRNA.

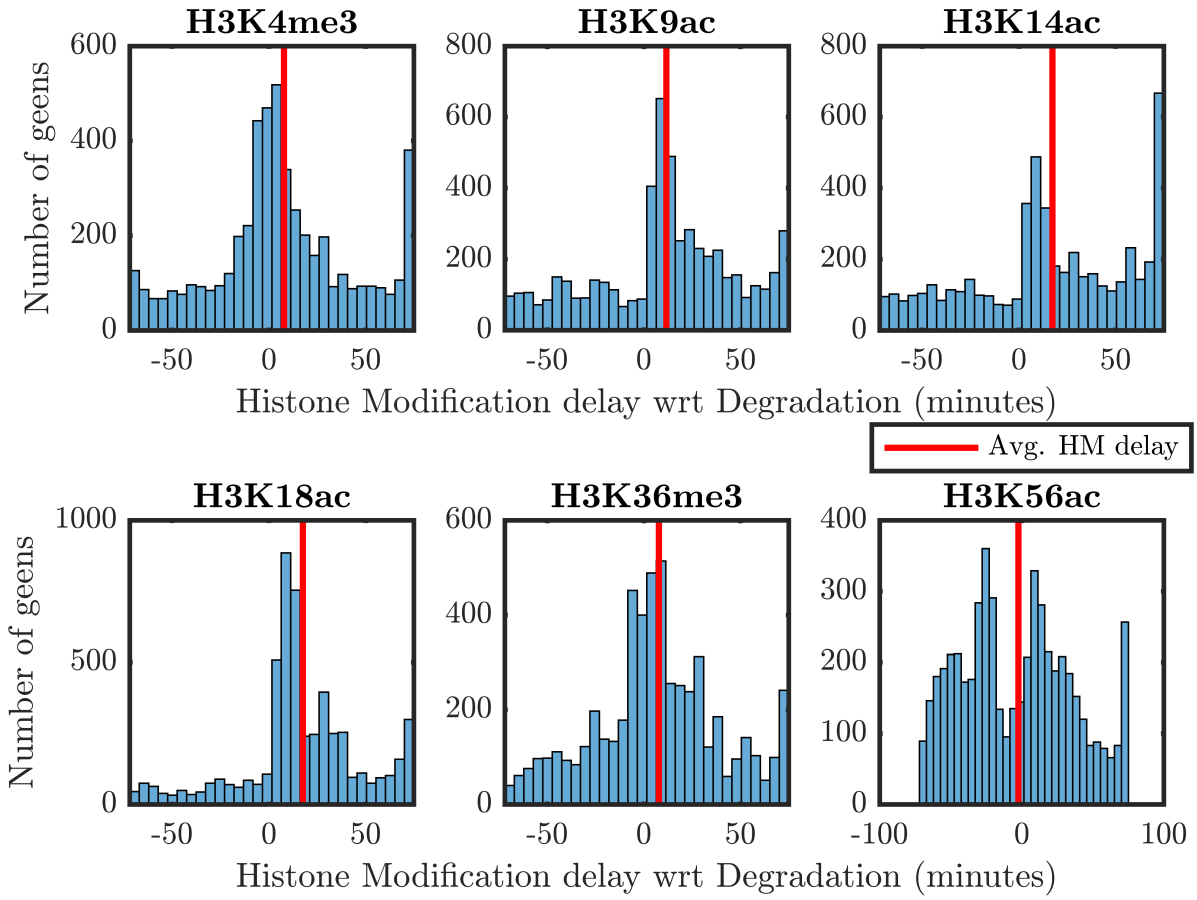


Figure 2.9: Distribution of the phase difference, or time delay, for the 5770 yeast genes between the HM accumulation function and the mRNA degradation rate dynamics function calculated in Section 2.4. Time delay is defined as the temporal shift in one of the functions which maximizes the Pearson correlation coefficient r between both functions. x-axis: time delay (bin width is 5 minutes); y-axis: number of genes exhibiting such delay. Each histogram represents, for the indicated HM, the distribution function of time delays between HM dynamics and degradation for each of the yeast genes. Red vertical lines represent the average time delay in all of the genes, which is 7.8 min for H3K4me$_3$, 11.7 min for H3K9ac, 17.4 min for H3K14ac, 17.6 min for H3K18ac, 7.8 min for H3K36me$_3$, and -2.3 min for H3K56ac. Early-tail lysine acetylations H3K9ac, H3K14ac, and H3K18ac have an entirely similar pattern, as both trimethylations H3K4me$_3$ and H3K36me$_3$ do. H3K56ac exhibits a bimodal distribution, with an average absolute value of the lag $E[|x|] = 34$ minutes.

Table 2.5: Average deposition time delays with respect to RNA expression, in minutes, for different HMs. The first column of the table is the histone modification, HM, studied. The second column collects the average delay, AD, values across all the genes (see Figure 2.9). The third column collects the average absolute values of the delays, $|AD|$, which are relevant to discard co-occurrence in some instances of low average values, as it is the case in H3K56ac. The fourth and fifth columns collect the same magnitudes when HM data are normalized by histone H3 abundance, AD/H3 & $|AD|$/H3. In the fourth column, it is observed that values are remarkably similar for all six HMs.

HM	**AD**	$\boldsymbol{\lvert AD \rvert}$	**AD/H3**	$\boldsymbol{\lvert AD \rvert}$**/H3**
H3K4me3	7.8	27.9	8.3	30.5
H3K9ac	11.7	32.2	13.4	32.9
H3K14ac	17.4	38.1	16.7	36.9
H3K18ac	17.6	28.3	17.4	30.5
H3K36me3	7.8	26.4	6.3	26.4
H3K56ac	-2.3	33.9	4.2	33.8

2.6 Conclusions

In this Chapter I have discussed the interplay between HMs and YMC regulation, identifying lysine acetylations, rather than trimethylations, as the strongest YMC regulatory candidates, as evidenced from the accumulation dynamics shown in Fig. 2.2 from the RNA-seq and ChIP-seq sequencing results. Typically, HMs accumulate at well characterized regions of genes. Such distributions, as shown in Figs. 2.1 and 2.2, are highly HM-specific. It is also a relevant conclusion that HM distributions associated with a given modification tend to be conserved from species to species. For example, H3K4me$_3$ is present at genes' 5' ends in several species [242,324], while H3K36me$_3$ is found mainly in the genes' body region [334, 335], which means it is generally not regarded as instructive for transcription [421]. In general, H3 tail acetylations tend to accumulate at the 5' end of genes [422].

I have shown how RNA and HMs correlate across the YMC: H3K4me$_3$ cycles weakly and accumulates after the RNA (see Subsection 2.6.1), while H3 tail lysine acetylations mirror RNA cycling (see Subsection 2.6.2). I have also calculated the mRNA degradation rate during the YMC, which is different from batch culture degradation, as discussed in Section 2.4, revealing that most catabolic, degradation processes, occur in fact in saturated media, where degradation is stronger and, for

a given gene, degradation occurs at a constant rate. Finally I have established a novel relationship between mRNA stability and HMs dynamics which is presented in Subsection 2.6.3.

In Chapters 4 and 5 I develop a quantitative mathematical model for the YMC regulation and dynamics, and I will use the results presented in this Chapter as the fitting parameters of the relevant systems of ODEs, which include terms like mRNA degradation rates.

2.6.1 H3K4me$_3$ cycles weakly and accumulates after the RNA

The crucial conclusion arrived at in Section 2.2 is that H3 lysine trimethylation accumulates on average 15 minutes after mRNAs have been expressed, supporting the hypothesis that the deposition of H3K4me$_3$ onto chromatin is a consequence of transcription, explaining why H3K4me$_3$ correlates positively with transcription levels [326, 330], as we can see in Figures 2.3.a and 2.4.a.

I hypothesize that the biochemical mechanism underlying this result is that kinase Kin28, a subunit of the basal transcription factor TFIIH [423], acts on some serines at the C-terminal domain (CTD) [424], activating Set1 via the Ser5p [425], which is the protein responsible for all H3K4 methylation in *S. cerevisiae* [323]. The literature has not measured the duration of such process, but my calculations in Section 2.2 reveal that the duration of this regulatory mechanism is likely to be approximately 15 minutes.

2.6.2 H3 tail lysine acetylations mirror RNA cycling

A second conclusion of Section 2.2 is that H3 early-tail lysine acetylations appear exactly at the same time as the corresponding mRNA, supporting the idea that H3K9ac correlates strongly with transcriptional activity [151].

Figures 2.1, 2.3.b, and 2.4.b, show that this is indeed the case for H3K9ac and H3K18ac, and similar data follow for H3K14ac. However, results for H3K56ac were entirely different, since the lack of a definite correlation with RNA obtained in Section 2.2, Figure 2.3.c, suggests it is not directly involved in the YMC regulation, at least not in the same way as other H3 lysine acetylations are, since it does not mirror RNA cycling.

For the early-tail acetylations H3K9ac, H3K14ac, and H3K18ac, their appearance precisely at the same temporal window as transcripts provides strong evidence for the 'just in time' acetylation supply hypothesis [127, 426]. Unfortunately, as discussed before, the employed ChIP-seq data lack the time resolution required to test this hypothesis and determine if the acetylations are transcriptional consequences or causes.

2.6.3 H3K18ac abundance is incompatible with mRNA stability

Finally, Figure 2.8 plots peaks in H3K18ac abundance versus mRNA decay rates. The clear L-shape of the distribution ($r = -0.31$) reveals that H3K18ac abundance is incompatible with mRNA stability. This results suggests that, perhaps, H3K18ac presence regulates transcript stability. What precise form this regulation could take is yet to be elucidated, but I hypothesize it could be in an 'all or nothing' manner similar to the transcript stability regulatory mechanisms of post-termination ribosome interactions with the $5'-$UTR [427, 428] and nuclear localization via the Crz1 transcription factor [429].

Chapter 3

Gene Clustering Algorithms

There are 5770 verified coding genes in *S. cerevisiae* [430–433]. Each gene follows its own RNA expression and histone-tail covalent-modification dynamics, which is subject to considerable measurement noise, therefore very little information can be extracted by analysing the dynamics of these genes individually. To identify common patterns in gene dynamics, the researcher has to group, or cluster, these 5770 genes into sets formed by genes exhibiting similar expression and covalent modification dynamics. Following this method, my aim is to unveil common features of these genes that will allow, using relevant bioinformatic analysis, elucidation of the YMC regulatory mechanisms.

The main difficulty of this task, as will be evident throughout this Chapter, is that the literature has produced and proposed a huge variety of methods for cluster analysis, without any of them being the standard tool in the field. Questions such as which clustering algorithm should be used and how many clusters should be produced do not have a straightforward answer, but do nevertheless need to be examined, otherwise the choice, and therefore its results, would be somewhat arbitrary and may not be reproducible when using a different algorithm.

Clustering analysis can be used to identify the sets of genes that are co-regulated, since they should exhibit similar RNA expression and HMs accumulation patterns [228]. This way, it is possible to search for both common transcription factors (TFs) and their upstream regulators expecting to identify the YMC control circuit components. Overall, the main goal is to identify relevant trends across the yeast genome by clustering genes in a non-biased manner.

This Chapter incorporates some mathematical formulation; to avoid confusion, it is important to differentiate a cluster (a subset of genes) from a clustering or partition (the set of clusters resulting from a given algorithm). Clustering algorithms can be exhaustive (if they cluster all the imputed data) or non-exhaustive (if they only assign clusters to some of the data).

Section 3.1 provides a brief description of the datasets used to perform the clustering analysis through the rest of the Chapter. As mentioned before, there is not a unique way of clustering genes, so the different options are discussed in Section 3.2. Clustering algorithms do not answer the question of how many clusters should a given dataset be divided into; this is a separate problem that is addressed in Section 3.3. The results from the clustering algorithms need to be evaluated, employing mathematical analysis, to assess the extent to which genes belonging to the same cluster behave in a similar manner, as done in Section 3.4. A refinement of these algorithms is presented in Section 3.5, where clustering analysis is performed on several datasets simultaneously, and afterwards these results are interpreted from a biological perspective, using GO analysis to learn which biological processes relate to each of the clusters, and examining whether different yeast strains have the same set of oscillating genes. Finally, the results obtained are discussed in Section 3.6, explaining their relevance towards building the mathematical models presented in Chapters 4 and 5.

3.1 Data to be clustered

The data analysis performed in this chapter uses a wide variety of data, obtained by different research groups under different experimental conditions, and hence will contain noise and -potentially- experimental bias or systematic error. This Section aims to describe these datasets, explaining their biological relevance, in order to provide a better understanding of their origin and limitations.

This analysis is concerned with three biochemical processes that are relevant to further our understanding of the YMC dynamics and regulatory logic: RNA expression (Subsection 3.1.1), HM accumulation (Subsection 3.1.2), and density of transcribing RNA polymerase (Subsection 3.1.3). All datasets have been processed using the algorithms discussed in Appendices A.3 and A.4.

3.1.1 RNA expression

Two of the most common methods for measuring RNA expression genome-wide are Microarrays and RNA-seq. RNA-seq is regarded to be the most reliable, offering benefits like the ability to detect novel transcripts [235, 434], a wider dynamic range [235, 434, 435], higher specificity and sensitivity [436–438], and simpler detection of rare and low-abundance transcripts [437, 438]. However, RNA-seq is a newer technology, and crucial YMC experiments from the previous decade that form the basis of our current understanding of the YMC were performed using Microarrays. An aim of this Chapter (Section 3.4) is to compare data from both sequencing technologies to learn if some systematic error may be present in either of them.

There are two reasons for which RNA expression will play a pivotal role across the analysis in this chapter: first, it is widely regarded as the most informative source of gene dynamics [439, 440], being the basic indicator of biological systems regulation [441, 442]; second, there are three papers in the literature that have measured RNA expression through the YMC [3, 108, 127], while there are only single datasets currently available of either HM accumulation dynamics or RNA polymerase density measurements during the YMC. Because biological experiments have non-negligible noise and a wide number of systematic errors, having the same magnitude compared three times under different experimental conditions may reveal some insights into its role in the YMC regulation. The three RNA expression datasets used were:

Data from the Tu lab Tu *et. al.* [3] performed Microarrray sequencing [443] during the YMC on samples obtained across 3 cycles, performing a cluster analysis based on the regularity of the oscillations in these cycles. However, RNA expression regularity across these three experiments cannot be directly compared, since one measures RNA expression only during one oscillation [127] and the other adds phenelzine (PZ) after the first oscillation is completed [108], which means regularity is destroyed at that point. It must be noted that, when the experiments were performed, Microarrays were a noisy technique [444]. An enlarged description of the experiment, sampling, and dataset is provided in Appendix A.1.1.

Data from the Li lab Li *et. al.* [108] performed an Affymetrix expression Microarray analysis at 4-min intervals through four *fast* 40 minutes cycles. After one YMC sampling, they added PZ at 1 mM. The effect of PZ on the YMC regulatory mechanisms, especially on the speed of several HMs occurrence, has been discussed in Section 2.1. An enlarged description of the experiment and dataset is provided in

Appendix A.1.1; sampling timepoints are shown in Figure A.2.

Data from the Boeke lab Boeke *et. al.* [127] performed RNA-seq at 16 timepoints during one YMC oscillation. An enlarged description of the experiment and dataset is provided in Appendix A.1.1; sampling timepoints are shown in Figure 1.7.

3.1.2 HM accumulation

HM accumulation levels have been measured using the more recent ChIP-seq technology [149]. Each histone octamer (containing two H3 histones) is wrapped around 147 bp [445], and each histone has only one tail, or N-terminus, therefore each octamer can only have two H3 covalent modifications at a given amino acid tail position. This fact, together with DNA fragments size, places a limit on the possible resolution of sequencing data. ChIP-seq technology is less reliable than RNA-seq and has a series of caveats that are discussed in Appendix A.2.

ChIP-seq was measured in the same experiment measuring RNA expression [127], measuring eight histone marks over one YMC oscillation: histone H3 lysine 9,14,27, and 56 acetylation, histone H4 5 and 16 lysine acetylation, and histone H3 lysine 4 and 36 trimethylation.

3.1.3 RNA polymerase density

The density of transcribing RNA polymerase has been measured genome-wide using NET-seq [373], a technique consisting of deep sequencing of $3'$ ends of nascent transcripts associated with RNA polymerase, to monitor transcription at nucleotide resolution [240]. A full discussion of the sampling techniques and experimental conditions for this dataset is presented in Appendix A.1.3; sampling timepoints across the YMC are shown in Figure A.3. Table 3.1 explains the most important experimental conditions under which each of these experiments were performed:

Table 3.1: Different YMC experimental conditions comparison. Most significant differences between the conditions in the different datasets used for this study. CEN.PK refers to the CEN.PK113-7D strain [446, 447].

Feature	**Tu**	**Li**	**Boeke**	**Mellor**
Reference	[3]	[108]	[127]	[373]
Period (h)	5	0.7	4	3.2
Doubling T (h)	8	8	7	≈ 6
Cycles	3	4	1	1
Samples	36	48	16	12
Drug added	No	PZ	No	No
Cell density	$\mathrm{OD}_{600} \sim 8.5$	$3 \cdot 10^7$ cell/L	Unspecified	$\mathrm{OD}_{600} \sim 8.5$
Yeast strain	CEN.PK	IFO 0233	CEN.PK	CEN.PK
Type	Diploid	Diploid	Haploid	Haploid
Sequencing	MicroArray	MicroArray	RNA-seq	$3'$ End-seq

3.2 Clustering algorithms

This Section is devoted to exploring different ways of clustering the yeast genome, by taking into consideration different aspects of the gene expression dynamics that will be interrelated with HM and nascent transcription dynamics. As mentioned in the introduction, this is not an obvious task, since in bioinformatics a unique, universally accepted, clustering algorithm does not exist [448]; rather, the literature has pointed at the problem of multiple clustering algorithms as a frequent source of confusion [448–450], particularly since each of them utilizes different distance/similarity measures and/or objective functions [451], a problem known in the field as the cluster ensemble selection problem [452–455]. Therefore, to obtain meaningful clustering results we cannot just apply a mechanical recipe to each of the datasets: it is necessary to examine the biological meaning of the data presented, and explore which clustering algorithms better present the relevant features, and hence to solve the ensemble selection problem.

In this Section, I propose three different algorithms to cluster individual datasets: expression peak clustering in Subsection 3.2.1, oscillation pattern clustering in Subsection 3.2.2, and k-means clustering in Subsection 3.2.3. The more complex task of clustering the yeast genome using several datasets simultaneously is deferred to Section 3.5.

3.2.1 Expression peak clustering

The YMC is an oscillatory system in which over 60% of the genome oscillates, and many of these genes are only active for short windows of time. Prima facie, oscillating genes are better candidates to explain the YMC than genes whose expression levels do not vary significantly through the cycle. This clustering algorithm aims to explore the common oscillatory patterns of such genes, by clustering together genes being contemporaneously activated in the YMC.

Expression peak clustering is non-exhaustive, since it only includes cycling genes whose expression peak is clearly defined at some point of the cycle. Genes are clustered by the YMC time at which they peak, considering a peak as identified if the ratio between maximum and minimum expression levels across one oscillation is larger than 2, a criterion frequently used in the literature [456–458]. The rationale for the selected ratio > 2 threshold is a trade-off between obtaining definite peak signals and forming clusters large enough that biological analysis can be performed on their components. A lower threshold would not be suitable because strong noise in the sequencing signals, plus a possibly low threshold in the sequencing saturation (see Appendix A.2) would imply a large type II error, *i.e.* false negatives would be included in the clusters. A higher threshold would result in a large type I error, since a sizeable number of cycling genes would not be clustered.

Results

RNA expression clusters were produced, obtaining results largely matching previous YMC clustering analysis using RNA expression [172, 459]. With this method, a total of 3762 genes (65% of the identified genome) were observed to oscillate; 1275 genes peaked during the YMC OX phase, 1634 during RB, and 853 during RC.

HMs clustering results are more interesting, since they revealed, for H4K5ac and H4K16ac, approximately zero accumulation at time point $t = 218$ min for almost all of the genes. Excluding this time point, the amplitude of the oscillations was, for both HMs, less than 10% of the average accumulation level at the vast majority of genes, and closer inspection of the data revealed jarring sudden transitions at some samples, multiple 'NaN' inputs and $> 20\%$ of the genes not expressing either of this HMs.

This non-cyclic behaviour may be caused either by a measurement error or because H4K5ac and H4K16ac are not linked to the YMC machinery, and therefore

they do not cycle with it.

H4K16ac controls chromatin structure and protein interactions [460], is involved in transcriptional elongation [461, 462], and regulates DNA repair genes in response to oxidative stress [463]; therefore, it is very unlikely that it is not involved in YMC regulation.

I hypothesize that H4K5ac and H4K16ac have been incorrectly measured in Boeke *et. al.* [127]. This may be due to deficiencies in the antibody used to perform the sequencing. Antibody adherence is a recurring problem in the field of sequencing [464], being responsible for many paper retractions [465] and wrongly conducted experiments [466]. The Millipore 07-327 and Millipore 07-329 Merck antibodies used to sequence H4K5ac and H4K16ac have demonstrated some crossreactivity toward H3K9ac [332]. Furthermore, Boeke *et. al.* [127] did not perform a specific antibodies validation, even if the quality of antibodies can vary among batches of the same antibody [149]. For such reasons, I have excluded H4K5ac and H4K16ac ChIP-seq datasets from this analysis.

3.2.2 Oscillatory pattern clustering

The clustering-by-expression-peaks algorithm presented in Subsection 3.2.1 provides some relevant insights into data reliability, as we have discussed, but by considering only maximum and minimum expression levels it entails a loss of information about the oscillations' shapes that this Subsection aims to repair. The oscillation-shape clustering rationale is to group genes by the form of their oscillations. It is non-exhaustive, since it only produces clusters of genes on which the average standard deviation from the cluster metagene to the genes profiles is within a certain threshold.

The rationale behind this novel technique is to use a recurring algorithm that, at each stage, removes the observations that seem not to bear any relationship with the YMC, in order to obtain a coherent result [467]. This algorithm removes also stand-alone genes, a desired outcome because Microarray [468, 469] and RNA-seq [470, 471] are noisy techniques; therefore, measuring the regularity of oscillations at the single-gene level includes a large number of false negatives: genes cycling regularly that, because of the noise distortions, are not classed as such in the analysis. Other clustering algorithms [3, 472] are highly inefficient for pulling out a reduced set of genes -75 in [3, Fig. 3], 13 in [472, Table 1]- because of the huge number of false positives outnumber the group of potentially relevant YMC regulatory elements, a problem

that this algorithm aims to resolve.

The oscillation-shape clustering algorithm is as follows:

Algorithm

1. For each gene i, assign the number N_i of genes $j \neq i$ such that $r_{i,j} > 0.9$, where r is the Pearson correlation coefficient between RNA expression data (see Appendix A.3) of i and j.

2. Rank genes by decreasing N_i value. The 16 genes with largest N_i values are presented in Figure 3.1, along with their associated N_i genes.

3. For these 16 genes, form the correlations matrix $R = r_{i,j}$, where $r_{i,j}$ is the Pearson correlation coefficient between i and j.

4. The two genes corresponding to the largest value in R are merged into the same cluster, along with their associated genes.

5. Step 4 is iterated 12 times, until only 4 genes remain.

6. The genes that have been assigned to two clusters in Step 5 are removed from the lower N_i value cluster.

The algorithm results are shown in Figures 3.2.a (RNA expression) and 3.2.b (H3K18ac accumulation for the corresponding clusters shown in Fig. 3.2.a). Final number of clusters and the $r > 0.9$ cutoff in Step 1 are justified as a trade-off between producing as few clusters as possible and obtaining intra-cluster standard deviations as low as possible.

These clusters have smaller sizes than the Section 3.2.1 peaks clustering, since they require not only definite peaks, but also similar expression levels throughout the YMC.

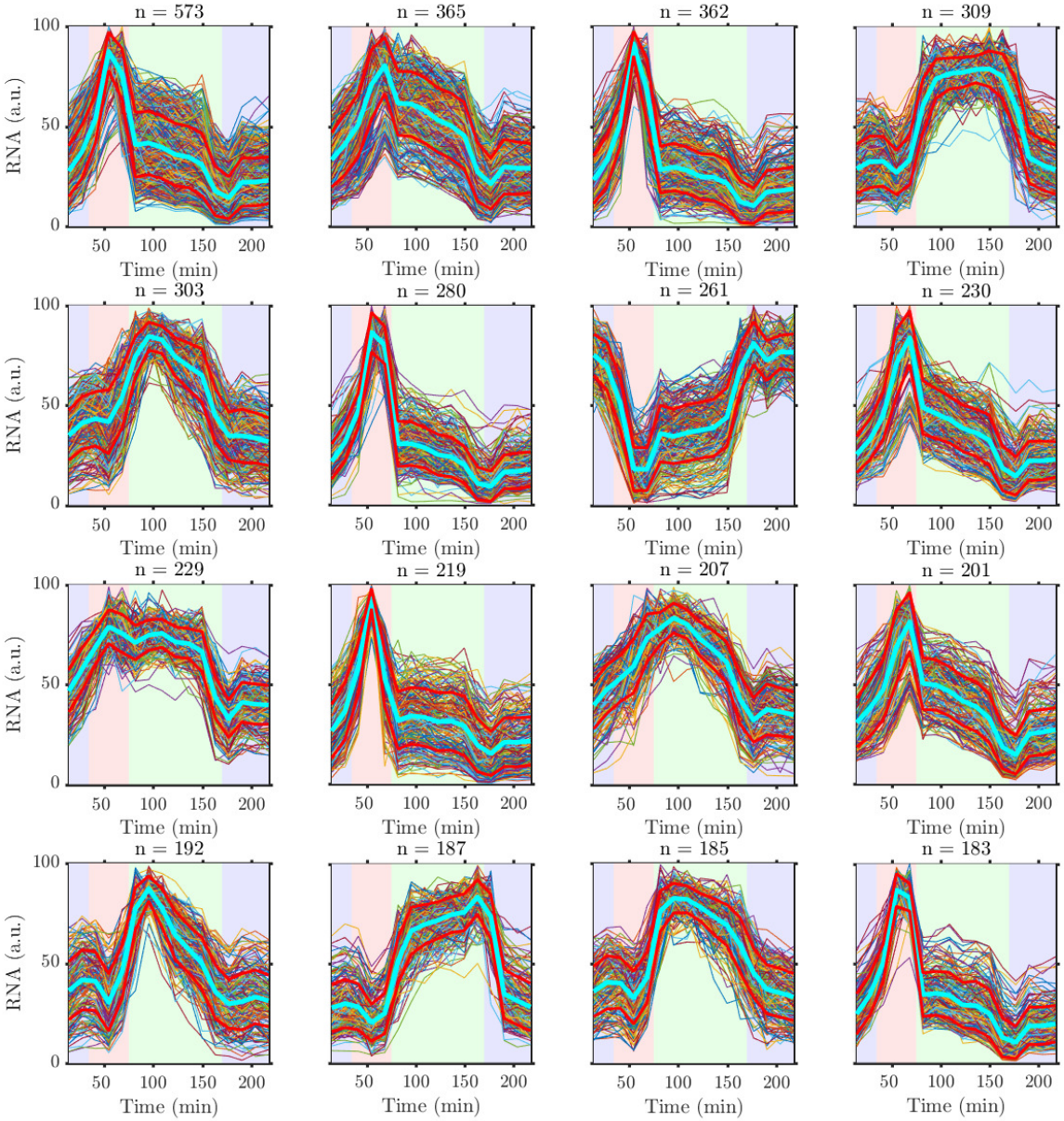


Figure 3.1: The sixteen most frequent RNA expression patterns, as determined by Step 2 of the oscillation-shape clustering algorithm described in Section 3.2.2. x-axis: time in minutes; y-axis: RNA expression level in arbitrary units. Cyan line: average RNA expression level in the cluster μ, or metagene; Red lines: average level plus and minus the standard deviation $\mu \pm \sigma$. The background colors provide a visual guide to the three YMC phases: RC (blue), OX (red), and RB (green).

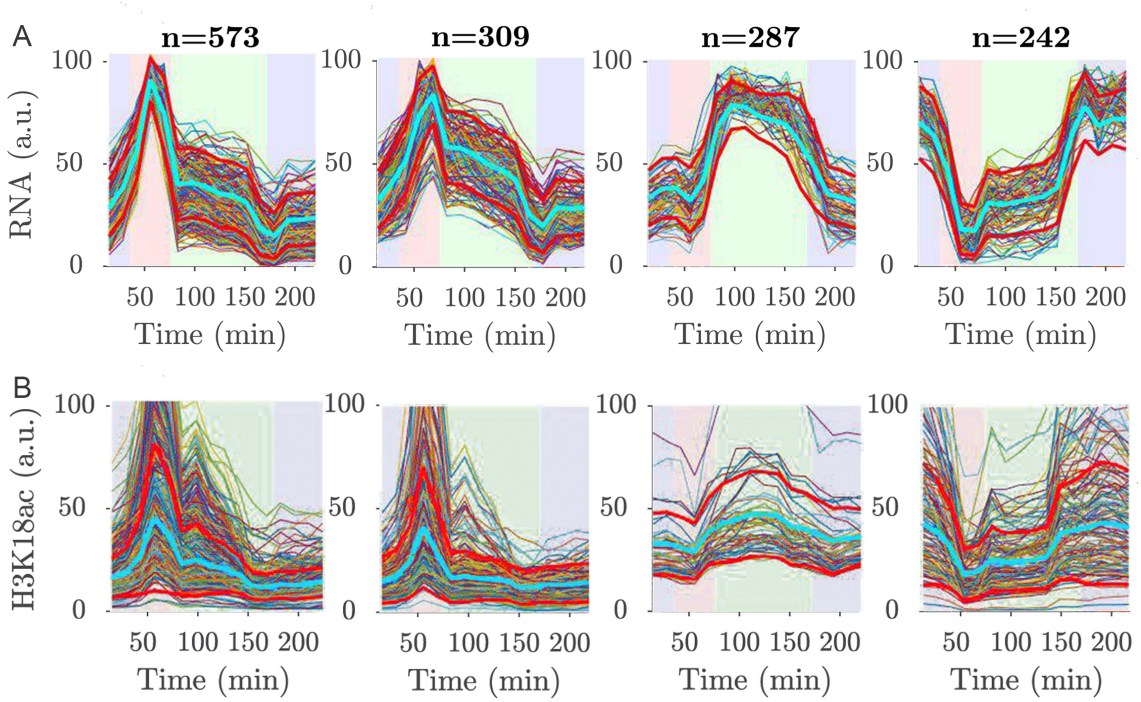


Figure 3.2: Four main clusters resulting from the algorithm in Section 3.2.2 for (a) RNA expression and (b) corresponding H3K18ac accumulation dynamics of the same clusters. The background colors provide a visual guide to the three YMC phases. The cyan line is the cluster average μ, or metagene, red lines show the standard deviation $\mu \pm \sigma$. Lysine acetylations at H3 tails synchronize with transcripts during OX phase. Clusters of genes peaking during the OX phase have lower standard deviations. For H3K18ac accumulation, the result is not significantly different if H3-normalized values are used instead of ChIP-seq reads (data not shown).

Fourier Analysis

Genes within the clusters represented in Figure 3.2.a follow a clear oscillatory dynamics. Their oscillations being highly regular [3], it is a natural question to pose if such oscillations are formed by a unique underlying cycle, or if they are the result of several, overlapping, processes of different frequencies.

To answer this question, these oscillations are decomposed into their fundamental harmonics by using the Discrete Fourier Transform (DFT) [473–475], a useful tool for identifying periodic changes in mRNA expression [77, 476–480]. Since the DFT requires measurements to be taken at equally spaced time points [481], data are first interpolated to 1 minute resolution using the MatLab cubic spline interpolation function [375]. Then, for each vector $\vec{x} = x_1, \ldots, x_N$ of points (RNA expression measurements), the DFT is calculated as

$$X_k = \sum_{n=0}^{N-1} x_n \cdot e^{-\frac{i2\pi}{N}kn}, \tag{3.1}$$

where the vector X represents the decomposition of the original time series into an offset value and a series of harmonic oscillations around this offset with periods $P_k = (t_{N-1} - t_0)/k$. This calculation was performed using the MatLab `fft` function [482, 483]. The amplitude of a given DFT component k is then $A_k = |X_k|/N$, and functional form of the decomposed waves is $f_k(t) = A_k \cdot sin(\omega_k t + \phi_k)$.

The analysis result, shown in Figure 3.3, is that the YMC is not formed by the superposition of different frequencies processes, since the wave amplitude of the main Fourier frequency associated with each of these clusters is eight times smaller than the amplitude of the metagenic oscillation, revealing thus that numerous ($n > 10$) harmonics are required to recover the initial metagene oscillation. This means a fundamental rhythm of smaller period than the YMC does not exist.

The Fourier analysis of the chemostat dissolved oxygen traces performed in Subsection 1.5.2 revealed a similar result. Therefore, processes and cycles exhibiting different frequencies than the main YMC frequency of $7.6 \cdot 10^{-5}$ Hz[1] can be excluded from the YMC regulatory mechanism. MatLab autocorrelation function `autocorr` revealed similar results.

[1]The inverse of the YMC period, 218 minutes.

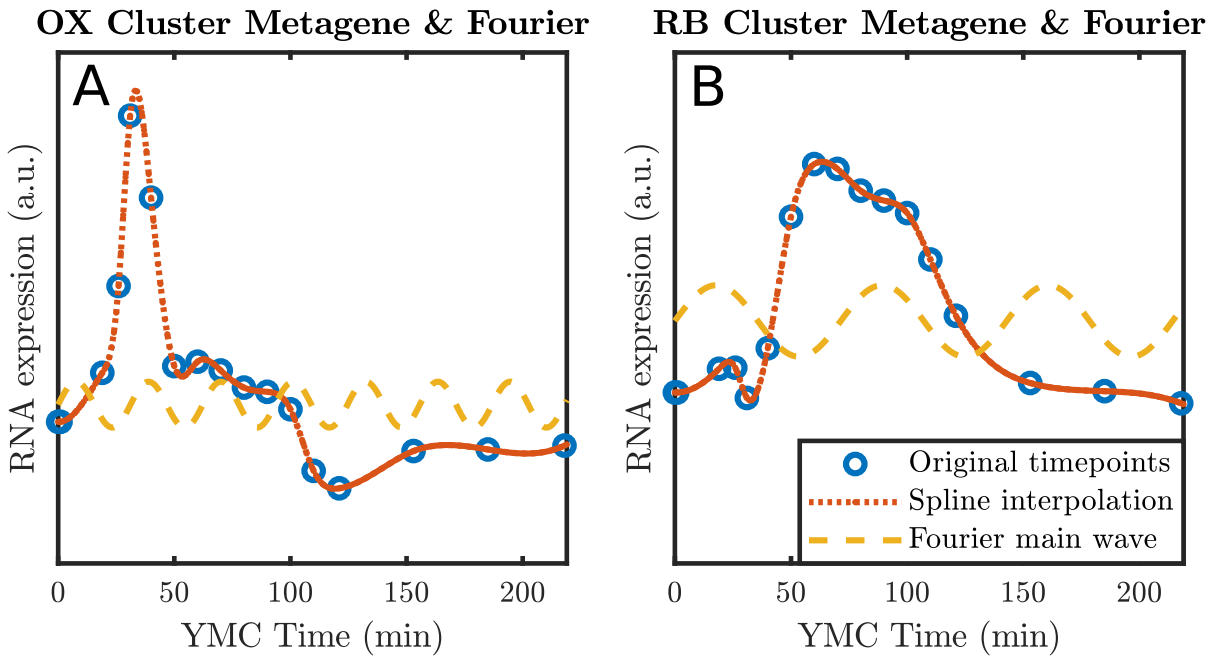


Figure 3.3: Fourier transform waves of (a) Cluster 1 (genes peaking during early OX phase) and (b) Cluster 3 (RB phase peak) in Figure 3.2.a metagenes. Blue dots are the metagene 16 time points. Discrete Fourier Transform requires equally spaced time points [481], so these are interpolated into the red line using the MatLab cubic spline function. Yellow dashed line corresponds to the first harmonic of the Fast Fourier Transform wave deconvolution. Because the amplitude of such line is eight times smaller than the amplitude of the RNA expression oscillations, a fundamental rhythm of smaller period than the YMC, or subdivisional process, does not exist. The second harmonic amplitude is negligible, and therefore it is not displayed.

It is relevant to compare this result with the conclusions drawn in Chapter 2 regarding the interplay between HMs and RNA: the existence of higher-order (second and successive Fourier harmonics) processes in RNA expression could have suggested a more complex YMC regulation via epigenetic switches. Such a result may have questioned the Subsection 2.6.2 claim that H3 tail lysine acetylations mirror RNA cycling, since they could be following independent, different-frequency oscillatory processes, as the analysis of Chapter 2 would have not been able to differentiate between these hypotheses.

The OX/RB interphase

An interesting result of the analysis performed in this Subsection is obtained by observing that the clusters in Figure 3.2.a correspond to very clear OX (Left), RB (Middle Right), and RC (Right) clusters, as defined by both the dO_2 trace shape and the literature [115, 127]. The only cluster not following into this classification is (Middle Left), a rich ($n = 303$) cluster of genes peaking at the OX/RB boundary.

During the YMC, H3 acetylation marks H3K9ac, H3K14ac, H3K23ac, and H3K27ac [116, 118, 128] as well as H4 acetylation marks H4K5ac, and H4K16ac [279] only appear over a brief $30 - 45$ minutes window of time corresponding precisely with the OX/RB boundary, which is coincident with the peak abundance of intracellular acetyl-CoA [116]. From the calculations performed in Subsection 1.5.2, this corresponds to the YMC high oxygen consumption phase, where oxygen consumption Q_{O_2} is maximized.

I investigate on which genes HM accumulation levels peak during the OX/RB interphase. For each of the six HMs studied, approximately 130 genes were found to exhibit definite HM accumulation peaks precisely at this part of the YMC. Using the GOrilla algorithm to find the GO terms associated with the OX/RB interphase cluster, it is found that ribosomal biogenesis and rRNA processing are the chief two GOID families of identified biological processes.

Similar searches were performed for the groups of genes with HMs accumulation levels peaking at either the RC/OX or RB/RC boundaries, but no meaningful GO results (GOrilla) were found. It has been claimed that Gcn5, a catalytic subunit of ADA and SAGA histone acetyltransferase complexes [287], is the enzyme responsible for transient acetylation of growth genes at the OX/RB transition [117, 279]. My

results[2] for this gene and related HAT complexes [484,485] did only show a significant H3K18ac peak in the middle of the RB phase, with other HMs not showing clear patterns or remarkable accumulation at the OX/RB interphase.

3.2.3 k-means clustering

k-means clustering [486–488] is widely used in bioinformatics analysis of genomic sequencing data [489–493]. k-means is a vector quantization algorithm to partition the observation space into k clusters minimizing the total sum of point to nearest centroid distances [494]. Its computational implementation consists in taking the expression values for a given dataset and producing the corresponding clusters using the MatLab `kmeans` function [495, 496], which uses the `k-means++` algorithm [497].

To determine the optimal k value is a difficult statistical problem [498–500], and it is discussed in Section 3.3. k-means analysis is dependent on the initial algorithm starting point, so I run the MatLab `kmeans++` 10000 times and select the clustering with smallest within-cluster sums of point-to-centroid distances.

3.2.4 Other clustering methods

In the future it would be interesting to complement this analysis with hierarchical clustering algorithms [501, 502], which are applicable to cluster sequencing results [501, 503–506]. Some interesting modifications to the algorithms presented would be to change the Euclidean distance of the k-means algorithm to a new metric reflecting the complexities of the YMC and to follow some noisy dataset analysis algorithms [507].

3.3 Optimal number of clusters

Clustering algorithms follow different criteria for grouping together data points (genes) in a given number of clusters, but they say nothing about how many clusters a given dataset should be partitioned into. Determining the appropriate number of clusters for a given algorithm is a highly contended question in the literature [508–515]; there is a trade-off between increasing the number of clusters (and hence obtaining more similar elements within each cluster) and effectively recovering the similar features of

[2] Using high-resolution ChIP-seq from Boeke *et. al.* [127].

its elements. The optimal cluster size must balance the increased accuracy resulting from a higher number of clusters and the achievement of an adequate data compression, which is the aim of every clustering algorithm [500, 516, 517].

For example, the oscillation-shape clustering algorithm presented in Section 3.2.2 can cluster RNA expression data into 16 clusters (as in Fig. 3.1), into more, or into fewer, by for example merging peak points (as in Fig. 3.2), a process that without the analysis of this Section can only be done via an ansatz, or educated guess, with its corresponding degree of discretion [518, 519].

The aim of this Section is to examine different statistical indicators to learn which number of clusters provides, in each circumstance, a better data compression, information which is crucial to perform any robust bioinformatics analysis. I examine the four criteria more frequently used in the bioinformatics literature [520]: the Bayesian Information Criterion (Subsection 3.3.1), the Akaike Information Criterion (Subsection 3.3.2), the Silhouette Criterion (Subsection 3.3.3), and the ELBOW method (Subsection 3.3.4); afterwards I discuss how the different criteria results compare to each other for the particular case of the datasets presented in Section 3.1 and the clustering algorithms discussed in Section 3.2.

3.3.1 Bayesian Information Criterion

The Bayesian Information Criterion (BIC) [498] is a metric of the trade-off between the information gain derived from the addition of an extra cluster to the partition, and the expected 'average gain' expected from such addition in a random distribution. The optimal number of clusters under the BIC estimator k_{opt}^{BIC} is defined as the absolute minimum of the function $BIC(k)$:

$$k_{opt}^{BIC} = \min_{k \in \mathbb{N}} \{BIC(k)\}, \text{ where} \tag{3.2}$$

$$BIC(k) = -2 \cdot \ln\left(\sum_{i=1}^{k}\sum_{j=1}^{n_i} \|x_j - C_i\|^2\right) + k \cdot \ln(n), \tag{3.3}$$

where k is the number of clusters, ln is the natural logarithm, n_i is the cardinality of cluster k_i, $x_1, \ldots, x_{n_i}$ are the elements of cluster k_i, C_i is the center of gravity of k_i defined as the arithmetic mean of $x_1, \ldots, x_{n_i}$ across each dimension, $\|v\|$ is the euclidean norm of vector v, and n is the total number of data points (genes). Large

k values are penalized via the $k \cdot \ln(n)$ term, while small k values will have large point-to-centroid distances $\|x_j - C_i\|$.

3.3.2 Akaike Information Criterion

The Akaike Information Criterion (AIC) [521] is very similar to the BIC, but with the difference that it does not depend directly on the size of the sample to be clustered, since the 'penalty term' $k \cdot \ln(n)$ is replaced by $2 \cdot k$. The optimal number of clusters under the AIC estimator k_{opt}^{AIC} is defined as the absolute minimum of the function $AIC(k)$:

$$AIC(k) = -2 \cdot \ln\left(\sum_{i=1}^{k}\sum_{j=1}^{n_i}\|x_j - C_i\|^2\right) + 2 \cdot k. \tag{3.4}$$

Figure 3.4 displays the functions $AIC(k)$ and $BIC(k)$ for clustering the Li *et. al.* [108] dataset using the k-means algorithm to show both minima, $k_{opt}^{BIC} = 11$ and $k_{opt}^{AIC} = 5$, for that particular dataset.

3.3.3 Silhouette

The Silhouette Criterion [522–524] aims to provide a graphical guide to the partitioning of the clusters that can reveal which genes lie well within their cluster, and which ones are merely somewhere in between clusters [522]. Cluster silhouettes are used in bioinformatics to assess the fitness function for a given dataset [525–527]. The average silhouette width provides an evaluation of clustering validity, and can be used to select the appropriate number of clusters [522, 528].

Like the BIC and AIC criteria, the Silhouette index evaluates the point-to-centroid distances within each cluster, but it also considers the average distances to each point from the points in the other clusters. For each point p in the dataset, the mean distance between p and all the other data points in the same cluster C_i is calculated as

$$a(p) = \frac{1}{|C_i| - 1}\sum_{j \in C_i, j \neq p}\|p - j\|, \tag{3.5}$$

where $|C_i|$ is the cardinality of C_i and $\|p - j\|$ is the euclidean norm of vector $p - j$. The minimum of the average distances to the point from the points in the other

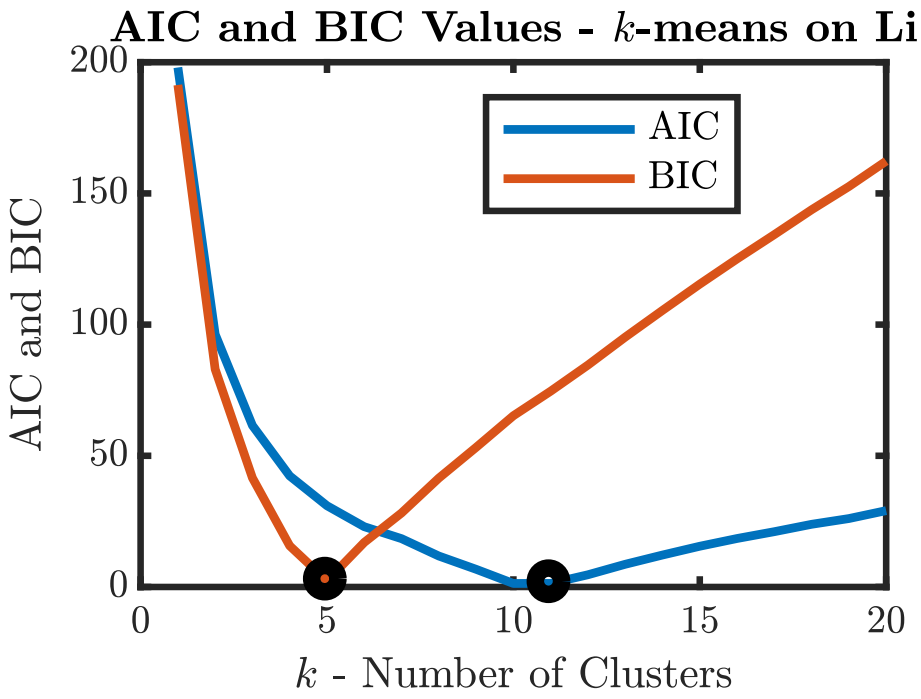


Figure 3.4: AIC and BIC values for the Li *et. al.* dataset [108] under the k-means algorithm. The BIC statistic [Eqn. (3.3)] is minimized at $k = 5$, where $BIC(5) = 2.3$, while AIC [Eqn. (3.4)] optimum value is at $k = 11$, where $AIC(11) = 1.3$ (both marked with black dots). Following Eqns. (3.3) and (3.4), we observe oblique asymptotes of slope $\ln(5770) \approx 8.7$ (for BIC) and 2 (for AIC) when $k \to \infty$.

clusters is

$$b(p) = \min_{k \neq i} \frac{1}{|C_k|} \sum_{j \in C_k} \|p - j\|. \tag{3.6}$$

Then, the silhouette width s of point p is defined as

$$s(p) = \frac{b(p) - a(p)}{\max\{a(p), b(p)\}}, \tag{3.7}$$

and the Silhouette Coefficient $SC(k)$ is defined as the mean of all the $s(p)$ values in the dataset:

$$SC(k) = \frac{1}{|n|} \sum_{p \in C} s(p), \tag{3.8}$$

where k is the number of clusters and n is the number of the data points (genes) in dataset C. The optimal number of clusters according to the Silhouette criterion is given by the maximum of the function $SC(k)$.

3.3.4 ELBOW

The *ELBOW* method [508, 529] considers the percentage of variance that can be explained through the number of clusters. The point at which the marginal increase of explanatory power drops is called the elbow. There is not a mathematical formula for obtaining the elbow point, which has to be determined 'by eye' on a case by case basis, hence this method has not been widely employed in bioinformatics analysis. In some circumstances, the marginal increase in information gain is approximately uniform for large intervals, leading to a disproportionately large optimal number of clusters [530, 531], as it is the case for the Li *et. al.* dataset shown in Table 3.2. Sometimes, the elbow point may not be unique [532], suggesting that there may exist several *natural* forms of clustering the data.

3.3.5 Comparison

Table 3.2 shows the optimal k value for the k-means algorithm implementation according to the BIC, AIC, Silhouette, and ELBOW number of clusters criteria for each of the datasets presented in Section 3.1.

From Eqns. (3.3) and (3.4) we observe that the only difference between the BIC and AIC criteria is the 'penalty term', $2 \cdot \ln(n)$ for the BIC and $2 \cdot k$ for the AIC. Since

Table 3.2: Optimal k-value for the k-means algorithm implementation according to the BIC, AIC, Silhouette, and ELBOW number of clusters criteria. The AIC and Silhouette agreement for the Tu and Boeke datasets leads us to propose that $k = 5$ is the optimal number of clusters for the Tu dataset, while the Boeke dataset is best partitioned into $k = 8$ clusters. The AIC criterion always results in a larger number of clusters than the BIC, since $k \cdot \ln(5770) > 2 \cdot k$. The anomalous result for the Li dataset under the ELBOW criterion is explained in Subsection 3.3.4.

Method	**Li**	**Tu**	**Boeke**
BIC	5	3	4
AIC	11	5	8
Silhouette	8	5	8
ELBOW	101	5	18

our datasets contain $n = 5770$ observations, it must be the case that $k_{opt}^{AIC} > k_{opt}^{BIC}$, since $\ln(5770) \approx 8.7 > 2$. These penalty terms are also responsible for the oblique asymptotes of slope $\ln(5770) \approx 8.7$ (for BIC) and 2 (for AIC) observed in Figure 3.4 when $k \to \infty$.

The optimal value of 101 clusters obtained for the Li *et. al.* dataset is remarkable, being much larger than any other value in the table. We hypothesize that this can be due to the following causes: first, under this criterion, the 'elbow point' cannot be always identified unambiguously [531]; second, this dataset comprises 48 observations that, because of the PZ addition after the first cycle, do not show the regular oscillatory pattern that the Tu dataset does, potentially requiring a much larger number of clusters to explain the expression divergence after the PZ addition point. The large number of clusters required to explain the Li dataset leads us to hypothesize that the ELBOW criterion is not adequate for genomic data clustering.

The AIC and Silhouette k_{opt} agreement for the Tu and Boeke datasets leads us to propose that $k = 5$ is the optimal number of clusters for the Tu dataset, while the Boeke dataset is best partitioned into $k = 8$ clusters.

3.4 Measures of clustering similarity

In the previous Sections, we have discussed and compared different forms of clustering of the yeast genome; however, if we were to obtain a unique clustering of the

genome, three questions must be resolved: which dataset to use, which clustering algorithm to apply, and how many clusters to produce. The aim of this Section is to compare the clustering results obtained in Sections 3.2 and 3.3 to learn how clusters compare to each other. By finding similar clustering distributions we expect to retrieve a result similar to the one obtained in Section 2.1 for HM metagene construction.

Following this aim, Subsection 3.4.1 is devoted to discussing the literature in measures of cluster similarity; this information is then used in Subsection 3.4.2 to build a novel metric that is particularly suitable to answer the questions raised in previous Sections, and it is applied to RNA expression data in Subsection 3.4.3 to explore the differences between RNA-seq and Microarray results on the YMC.

3.4.1 Cluster comparison criteria

Comparing the outcomes of different clustering algorithms is an old problem in bioinformatics [533–536]; it has, however, been much less explored than other problems in the field, like choice of a clustering algorithm or determination of the number of clusters [537–539].

An intuitive criterion, proposed by Rand in 1971 [540], compares different clustering results by evaluating how likely they are to cluster together each pair of objects. Following its lead, several more criteria were suggested, which can be roughly divided into two groups: set matching methods [541–544], on which the overlap between clusters belonging to different partitions is measured, and variation of information methods [536, 545–548], which measure the amount of information that is lost or gained in changing from one clustering to another [545]. The method followed in the next Subsection belongs to the class of set matching methods.

The consensus in the field seems to be that, because of the huge number of competing criteria used in the literature, it is probably meaningless to search for the best clustering algorithm or clusters comparison criterion [549–552]; the aim of cluster analysis should be to match the task at hand [549, 552].

3.4.2 A novel cluster dissimilarity measure

The aim of the Subsection is to produce and explain a novel metric particularly suitable to measure the degree of overlap between the clustering results obtained in the previous Sections of this Chapter.

Consider a dataset formed by the RNA expression data at each of the 5770 verified genes belonging to the yeast genome. Call such genome $G = \{g_1, \ldots, g_N\}$, where $N = 5770$ and g_i are the genes forming it, and suppose two clustering results C_1 and C_2 are obtained for genome G. These clusterings may have been produced by the application of any algorithm on any datasets (RNA-seq or Microarrays) with any k-value or number of clusters: the important point is that it is the genome of the same organism (yeast in this case) that is being clustered.

Let n and m be the number of clusters forming clusterings C_1 and C_2 respectively; we have that $C_1 = \{S_1, \ldots, S_n\}$ and $C_2 = \{S'_1, \ldots, S'_m\}$.

We define the *distance* d between set $S_i \in C_1$ and set $S'_j \in C_2$, $d(S_i, S'_j)$ as

$$d(S_i, S'_j) = |S_i| + |S'_j| - 2 \cdot |S_i \cap S'_j|, \tag{3.9}$$

where $|S_i|$ is the cardinality of set S_i and $S_i \cap S'_j$ is the set formed by the intersection of S_i and S'_j.

Theorem 1. *Function d, as defined by Eqn. (3.9), verifies the properties of a distance measure: non-negativity, symmetry, and the triangle inequality. Proof: see Appendix A.5.*

Different functional forms have been proposed for the distance function defined in Eqn. (3.9), like non-symmetric entropy-based measures [548], metrics including a distance-to-centroid term [553], or distances motivated by the assumption of an underlying interaction probability distribution [538].

This metric captures the concept of 'cluster similarity', but it is dependent on the size of the clusters: the larger the set size, the larger the d value. In order to obtain a metric that does not scale with cluster size, allowing thus for different size clusterings to be compared, Eqn. (3.9) is normalized dividing by $|S_i \cup S'_j|$ and using

the identity $|A| + |B| = |A \cup B| + |A \cap B|$ to obtain:

$$d(S_i, S'_j) = 1 - \frac{|S_i \cap S'_j|}{|S_i \cup S'_j|}. \tag{3.10}$$

We define the *associated cluster* or *best fit cluster* of each $S_i \in C_1$, $S'(S_i) \in C_2$ to be given by the function

$$\begin{aligned} f &: C_1 \to C_2 \\ &S_i \mapsto f(S_i) = S'(S_i), \quad \text{where} \end{aligned}$$

$$S'(S_i) = \arg \min_{S'_j \in C_2} d(S_i, S'_j). \tag{3.11}$$

The intuition behind this definition is to find the cluster $S'_j \in C_2$ that is most similar to S_i. Because d is a symmetric function (Theorem 1), *associated clusters* for each $S'_j \in C_2$ are similarly defined: $S(S'_j) = \arg\min_{S_i \in C_1} d(S_i, S'_j)$.

Equation (3.10) provides a metric for the distance between clusters belonging to two different clusterings C_1 and C_2. The next logical step is to define an index for the *dissimilarity* between both clusterings, $D(C_1, C_2)$:

$$\begin{aligned} D &: \{C_1, C_2\} \to \mathbb{R}_0^+ \\ &\{C_1, C_2\} \mapsto D(C_1, C_2), \text{where} \end{aligned}$$

$$D(C_1, C_2) = \frac{1}{n+m} \left(\sum_{i=1}^{n} d(S_i, S'(S_i)) + \sum_{j=1}^{m} d(S'_j, S(S'_j)) \right), \tag{3.12}$$

keeping in mind that S_i are the n elements of C_1 and S'_j are the m elements of C_2. Index D does not assume that clusterings C_1 and C_2 are exhaustive, i.e. $\forall\, g_i \in G \exists j \mid g_i \in S_j \in C_1$; it can be used indistinctly on exhaustive and non-exhaustive clustering algorithms. Similarly, C_1 and C_2 can be overlapping or not, i.e. $S_i \cap S'_j = \emptyset \;\; \forall\, i \neq j$.

Theorem 2. *Index D, as defined by Eqn.* (3.12)*, verifies the properties of non-negativity and symmetry. Proof: see Appendix A.5.*

The rationale behind this definition is to consider the distance between two clusterings as the sum of the distances from each cluster in C_1 to its associated cluster ex Eqn. (3.11) in C_2, $\sum_{i=1}^{n} d(S_i, S'(S_i))$. To ensure that D meets the requirement of being symmetric, the sum of distances from clusters in C_2 to their associates in C_1 is also included; finally, $1/(n+m)$ is a normalizing term included to enable comparison

between different-sized pairs of clusterings.

It is important to note that, in general, index D will not be a distance measure, since it fails to verify the triangle inequality. Other similarity indices, like the the Jaccard index [554], the Rand index [540], the Wallace index [555], or the normalized Lerman index [556] are more concerned with the idea of accuracy, subtly implying the existence of a 'correct' clustering C^* that is used as the index benchmark, which is not the case in this analysis.

A second class of techniques, based on the Hamming distance [557], is concerned with the number of individual permutations required to obtain C_2 from C_1; these include the Hausdorff distance [558–560], often applied in image data analysis [561], the Mahalanobis distance [562, 563] and recent metrics on average distance between sets [564, 565]. It is very often the case that these functions do not necessarily satisfy symmetry and/or the triangle inequality [564, 566]. The Hamming distance-based functions recently proposed by the literature that verify the triangle inequality [567–569] are complex, non-intuitive, and hard to interpret from a biochemical perspective: their scope extends well beyond this doctoral work in bioinformatics. The rationale for establishing index D is to use a metric that can be easily interpreted in the context of a standard bioinformatics analysis, at the expense of verifying the triangle inequality, as we do in the next Subsection.

3.4.3 Index application to genomic data

Robust metrics are independent on fine-graining clustering, since they provide similar results when large clusters are matched with large clusters and when small clusters are matched with small clusters [570], a result that we have achieved with the normalization factors $|S_i \cup S'_j|$ in Eqn. (3.10) and $1/(n+m)$ in Eqn. (3.12). Empirical confirmation of the robustness of the index presented in Eqn. (3.12) is provided in Figure 3.5, where $D(C_1, C_2)$ values are calculated for three different RNA expression datasets, using the k-means algorithm and different k values.

We learn from Figure 3.5 that Affymetrix Microarray results are most similar between them than compared to RNA-seq. I hypothesize that this statistically significant divergence is due to the different properties of both sequencing techniques: for example, Microarrays have lower signal saturation thresholds [571], while for RNA-seq smaller DNA fragments mean their expression value cannot be accurately determined in some replicates [572].

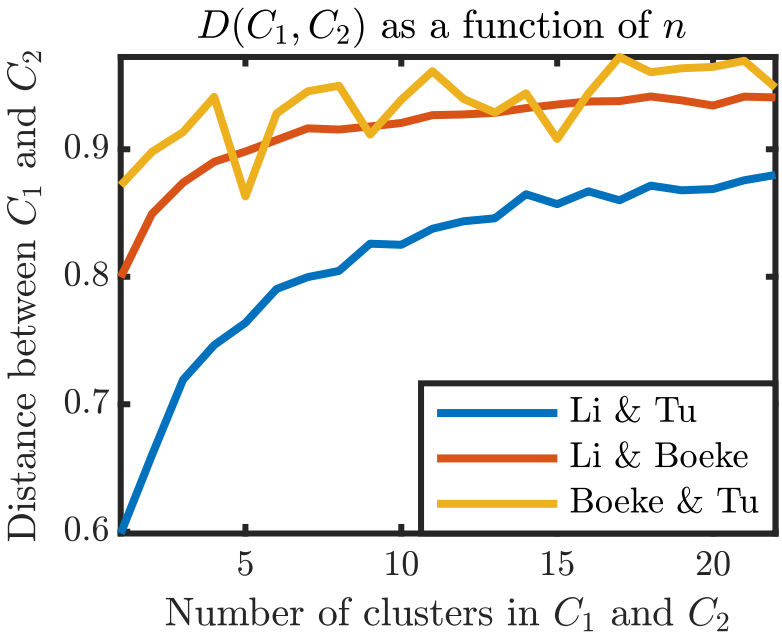


Figure 3.5: Dissimilarity index, D, values, as defined by Eqn. (3.12), between k-means clusterings comparing the three RNA expression datasets presented in Subsection 3.1.1: Tu (Microarray), Li (also Microarray), and Boeke (RNA-seq). Note that we have set an equal number of clusters in each of the clusterings compared, C_1 and C_2: $|C_1| = |C_2| = n$. The index value between the Li and Tu datasets D(Tu, Li) is consistently the lowest of the three pairs. From this fact, we learn that Microarray results are most similar between them than compared to RNA-seq. When clusterings are randomly generated, the typical value of the index is $D = 0.975 \pm 0.007$.

3.5 Clustering including multiple datasets

Sequencing data are noisy. The sequencing results of RNA-seq [573, 574], NET-seq [240,575], and ChIP-seq [348,576] discussed in this thesis require (when examined at single-gene level) a 2-fold change in expression to be regarded as statistically significant oscillations [577, 578]. Such a bioinformatic analysis constraint has produced a rich variety of tools aimed at the analysis of sequencing data, like ANOVA-SCA [579], Cuffdiff 2 [470], DESeq [580], GFOLD [581], and many others. As it is inevitable, statistical reliance comes at the expense of information loss.

The aim of this Section is to overcome the limitations that sequencing noise poses to the clustering techniques presented in this Chapter. This is achieved using a multi-omics approach: sequencing data from different sources are combined to form a single dataset, and then clustering analysis in performed on these larger datasets. Larger datasets have -proportionally- less error noise, and by combining data obtained at different experimental conditions, the influence that systematic errors may have in an individual dataset is diluted into the larger dataset, making the overall analysis more reliable.

This Section starts by presenting the bioinformatic techniques used to cluster together multiple datasets, and explaining why such co-clustering is effective in reducing uncertainty levels, increasing hence the clustering analysis reliability (Subsection 3.5.1). Then, these combined datasets are clustered using the knowledge acquired in the previous Sections of this Chapter, exploring first exhaustive[3] clustering (Subsection 3.5.2) and then non-exhaustive clustering (Subsection 3.5.3). Both methods are non overlapping clusterings: a given data point is mapped, at most, into one cluster. From this analysis, we aim to elucidate if HMs could affect transcription or vice versa, and ascertain if mRNA and HMs have similar dynamic regulation mechanisms.

3.5.1 Multiple dataset clustering technique

The goal of this Subsection is to extend the k-means clustering algorithm described and employed in Subsection 3.2.3 to account for more that one imputed dataset, taking into account the different nature and span of the different sequencing results. To achieve consistent, comparable, and reproducible results, I select, from the range of algorithms presented in Section 3.2, the k-means clustering algorithm. This al-

[3]Recall that exhaustive clustering is an algorithm in which every single data point (gene), is mapped into one cluster.

gorithm is well-known and easily implemented in bioinformatics software, such as MatLab or R.

The first step in our task is to normalize data: as we have discussed in Section 2.4, the sequencing reading intensity is given in arbitrary units, and the magnitudes sequenced do have different units. The introduction of a normalizing factor ensures that all datasets have the same data range and similar means and standard deviations; otherwise, the joint k-means calculation would be remarkably biased towards the highest-average dataset, probably NET-seq, because its readings are spiky [241, 504] and there is a relatively smaller overall number of reads [582, 583]. These considerations lead me to exclude from the analysis genes with very low NET-seq reads (lowest 10% percentile) in order to avoid possible distortion caused by normalization. This is justified because RNAPII RNAs are continuously being actively transcribed [584], therefore very low reads seem to imply a sequencing error rather than a lack of transcription.

Consider a genome $G = \{g_1, \ldots, g_N\}$ formed by genes g_i. Let X be a dataset containing, for each gene g_i, observations at T time points $(t_1, \ldots, t_T)$. We have that

$$X = (x_{it}) = \begin{bmatrix} X_1 \\ \vdots \\ X_N \end{bmatrix} = \begin{bmatrix} g_1(t_1) & \cdots & g_1(t_T) \\ \vdots & \ddots & \vdots \\ g_N(t_1) & \cdots & g_N(t_T) \end{bmatrix},$$

where $g_i(t_j)$ is the sequencing value measured for gene g_i at time t_j, and X_i are the data collected for each gene across the time course.

Each row in X, $X_i = (x_{i1}, \ldots, x_{iT})$ is normalized by applying the transformation

$$X_i(t) \mapsto X_i'(t) = \frac{X_i(t) - \mathrm{Me}(X_i)}{\sigma(X_i)}, \tag{3.13}$$

where $\mathrm{Me}(X_i)$ is the median element of X_i, and $\sigma(X_i)$ is the standard deviation of the elements in X_i. Employing the median rather than the mean for normalizing purposes is a common technique in bioinformatics [585–591] aimed to reduce the distortions caused by outliers.

Because each dataset X is sampled at different timepoints $(t_1, \ldots, t_T)$, data obtained from transformation (3.13) are interpolated, using the cubic spline function

as in Subsection 3.2.2, to estimate their values at the same timepoints in all datasets.

Once the three datasets are expressed in a standardized manner, they are merged into one using the Column-oriented DBMS Database Systems technique [592–594], a system commonly used to increase the computational efficiency of the k-means algorithm [594, 595].

This way, we obtain the concatenated matrix

$$\mathcal{X} = [X\ Y\ Z],$$

where X are the RNA-seq data, Y is NET-seq, and Z is ChIP-seq for H3K9ac. The reason for choosing H3K9ac as the key HM rather than any of the other five HMs available in Boeke *et al.* [127] is that H3K9ac, which is the most periodic HM [118], was independently measured twice [127, Online methods - Batch effect removal], providing therefore more reliable data. Based on the optimal number of clusters discussion of Section 3.3 displayed in Table 3.2, I set $k = 8$ for the $\mathcal{X}$ k-means clustering.

Error reduction analysis

The aim of multilevel clustering algorithms is to use the rich multi-omics datasets information to reduce the sequencing noise of individual experiments. This paragraph is devoted to estimate the extent to which sequencing noise is reduced as a result of the multi-omics approach.

White noise propagates with $\sqrt{n}$ in a sum of variables [596–598]: if distributions $D_1, \ldots, D_n$ each have a white noise term $D_i = D_i^0 + \eta$, where D_i^0 is the noise-free distribution and η is white noise such that $E(\eta) = 0$, then

$$\sum_{i=1}^{n} D_i \approx \sum_{i=1}^{n} D_i^0 + \sqrt{n} \cdot \eta,$$

which means that the joint distribution $\sum_i D_i$ has error $\sqrt{n} \cdot \eta$, and its average $\bar{D}$ has error $\sqrt{n} \cdot \eta / n$. Hence, since $1 - \sqrt{3}/3 = 0.42$, increasing the analysis from $n = 1$ to $n = 3$ datasets reduces the average noise by approximately 40%.

The result is even more stark in the case of systematic errors: if one of the distributions D_i has been measured with a systematic error ϵ_S so that $D_i = D_i^0 + \epsilon_S$,

where $E(\eta_i) \neq 0$, then $\bar{D}$ has error ϵ_S/n, and noise is reduced by approximately 67%.

3.5.2 Exhaustive clustering results

Figure 3.6 shows the multiple-omics clustering results obtained from applying the algorithm presented in the previous Subsection. It is interesting to compare Clusters 3 and 8: genes belonging to these two clusters exhibit very similar RNA-seq patterns: a sharp expression peak during the YMC OX phase followed by lower values during RB phase, a strong dip at the RB/RC interphase and a plateau during the RC phase; the correlation coefficient between both metagenes is $r = 0.96$. A similar result is obtained when comparing both H3K9ac ChIP-seq results: the accumulation peak is somewhat stronger in Cluster 3, but the correlation coefficient $r = 0.87$ is still high. The key difference is the entirely different NET-seq data, which seem to be out of phase between both clusters ($r = -0.68$): in Cluster 3, nascent transcription roughly mirrors RNA expression and H3K9ac accumulation, but in Cluster 8 it is negatively correlated with both magnitudes, suggesting perhaps pervasive anti-sense transcription [599, 600] for this cluster of genes, which is not being detected by non strand-specific sequencing techniques.

To draw reliable conclusions, it is necessary to verify that this result is not just an artefact introduced by either the clustering algorithm, the construction of the metagenes, the normalization techniques, or some unaccounted factor. The raw sequencing data of individual genes at each of the clusters were inspected using the IGV browser, a desktop tool for interactive visual exploration of diverse, large-scale genomic data sets [601, 602]. These raw data, as visualized under the IGV software, are shown in Figures 3.7 ($DBP2$, belonging to Cluster 3) and 3.8 ($UTP21$, belonging to Cluster 8), confirming that the results obtained in the previous paragraph are not an artefact. These two genes were selected from the cluster components because their correlation coefficient with the corresponding cluster metagenes were highest, making them the most representative example of the individual genes dynamics. Observing these single-gene data, in both Figs. 3.7 and 3.8 there is a strong H3K9ac *kick* located at the TSS nucleosome, which is only switched on during the OX phase, hinting at an HM-controlled switch for RNA expression. Similar results were observed for other H3 lysine acetylations, but not for H3 lysine trimethylations (data not shown).

To learn more about the function of genes in Clusters 3 and 8, GO (GOrilla) and TFs (YEASTRACT) analyses were performed on them. GOrilla software [377, 378]

was discussed in Subsection 2.4.1, while the YEASTRACT algorithm [389–393] was discussed in Subsection 2.4.2. Table 3.3 shows the TFs linked to the regulation of both clusters of genes according to the YEASTRACT algorithm, as well as the GO terms obtained from GOrilla for the most likely biological processes regulated by both clusters, which are found to be very similar. However, it is interesting to note how their regulating TFs, according to YEASTRACT, are remarkably different: Cst6p and Ace2p are the only two common TFs.

Figure 3.10 is a *GO tree* exploring the functional interrelation between the most relevant biological processes associated to the genes in the cluster. *GO trees* are schematic representations of the GO hierarchy using the GraphViz tool [603] of all significantly enriched terms obtained from a GOrilla analysis. The tree should be read from top to bottom: upper lines represent more generic biological processes, and at each node in the graphic, more specific processes are individuated.

Other HMs The discussion of this Subsection uses H3K9ac ChIP-seq data, rather than other HMs, because, as we discussed in the previous Subsection, it is the most reliable set of measurements. However, similar results were obtained when other H3 lysine acetylations, H3K14ac and H3K18ac, were used. An example of this is shown in Figure 3.11, where ChIP-seq data for other HMs in genes belonging to Cluster 3 of Figure 3.6 are represented. It is remarkable that both $H3K4me_3$ and $H3K36me_3$ start to accumulate when RNA and the lysine acetylations peak, and they remain deposed for all of the RB phase, suggesting that these trimethylations are an epigenetic mark signalling that transcription has already taken place at that gene: they could be repressive to stop reinitiation of transcription during a permissive window of time at the YMC high O_2 consumption phase.

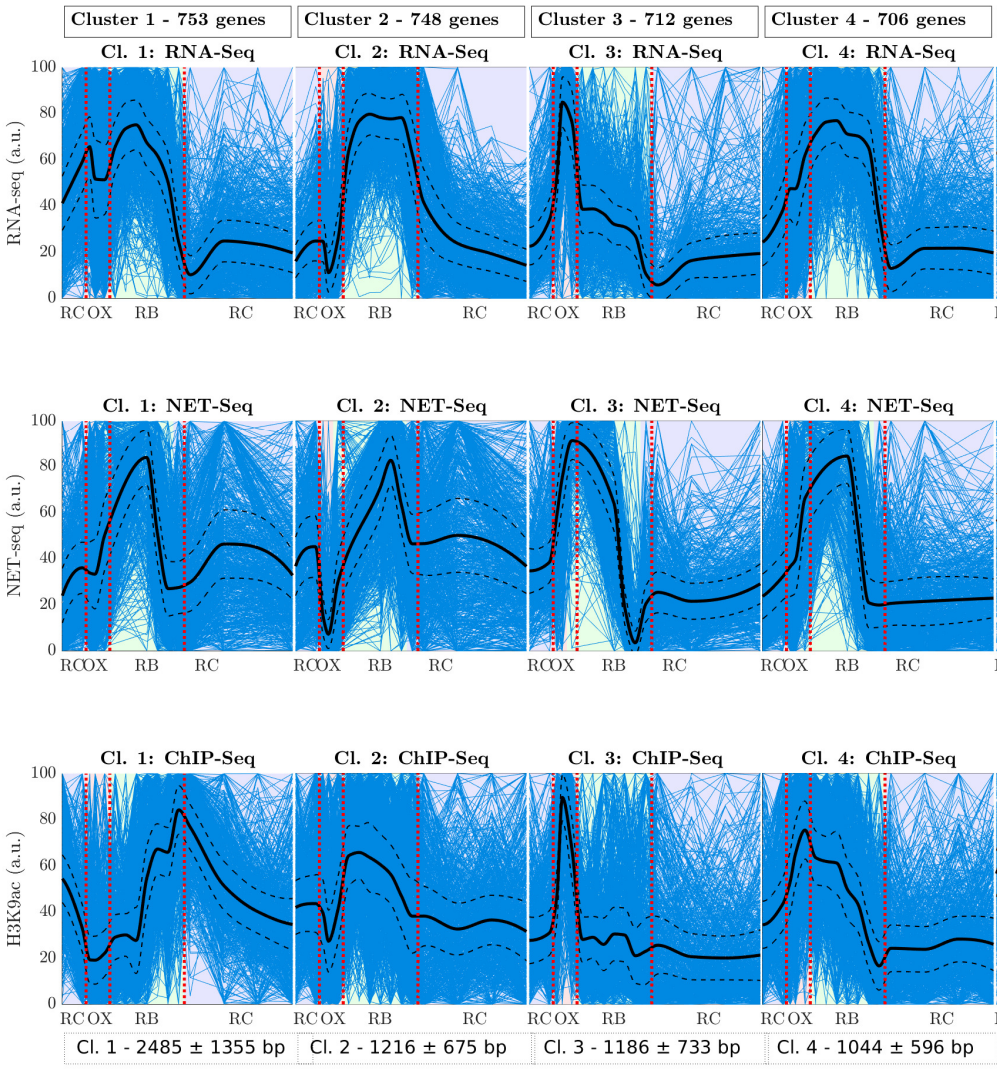


Figure 3.6: Eight clusters of genes obtained by combined k-means algorithm across time in mRNA, nascent transcription, and H3K9ac accumulation levels during the YMC. [First part of the figure, second part is on the next page]. First row plots the normalized RNA-seq expression values, second is NET-seq, and third is H3K9ac ChIP-seq. Each column is a different cluster whose size is indicated on the top. Bottom boxes indicate average length ± standard deviation (in bp) for genes belonging to that cluster. All clusters show an identical Poisson distribution [604] in genes length, the λ parameter being approximately the average gene length (data not shown), excluding therefore some statistical artefact for which gene length is relevant towards clustering. Blue lines represent individual genes' dynamics. Solid black lines represent the averaged metagene μ; dashed black lines represent $\mu \pm \sigma$, they are a measure of inter-cluster dispersion. Dashed vertical red lines signal the YMC phases transition in the order RC to OX to RB to RC. The most striking result -see main text- is the comparison of Clusters 3 and 8, where entirely similar RNA-seq and H3K9ac ChIP-seq readings are matched by different NET-seq reads: in Cluster 3, NET-seq is on phase with both RNA-seq and HM ChIP-seq, but in Cluster 8 it is out of phase with them. Similar results were obtained for other HMs (data not shown).

Note: this figure continues on the next page.

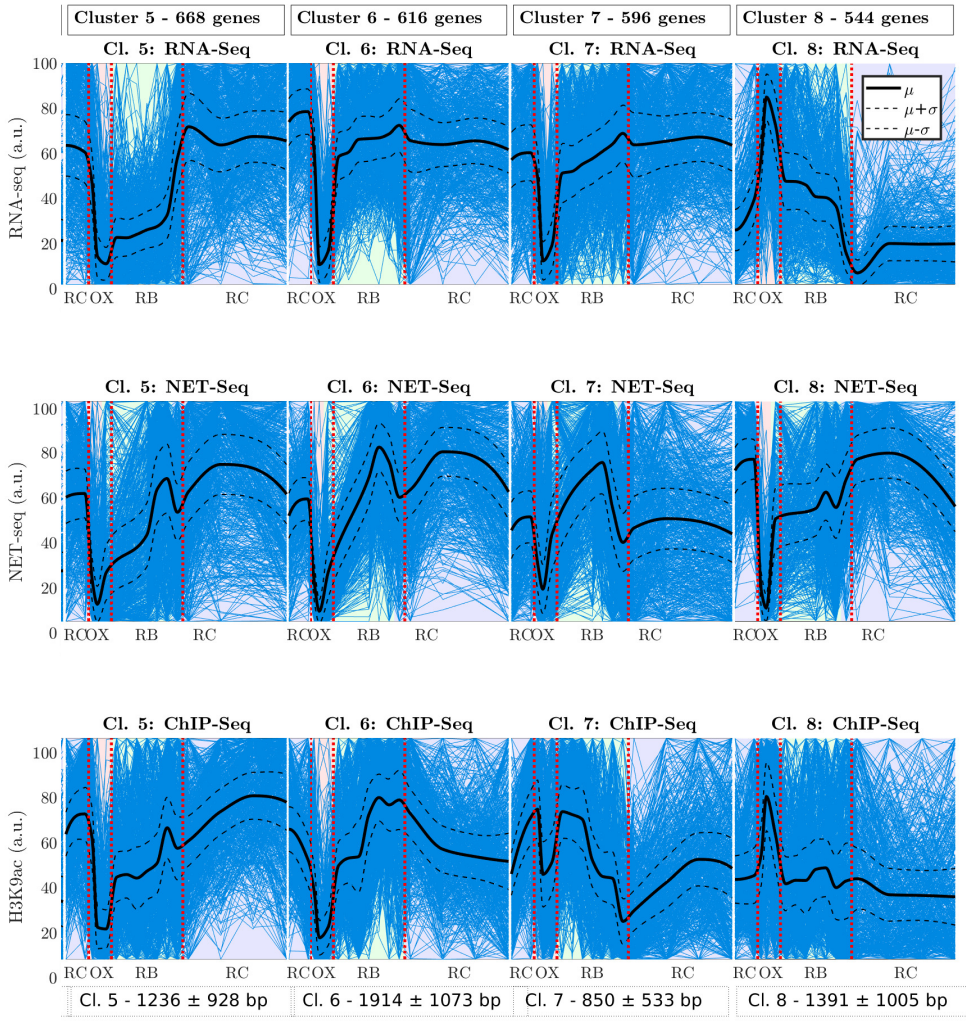


Figure 3.6 (continued).

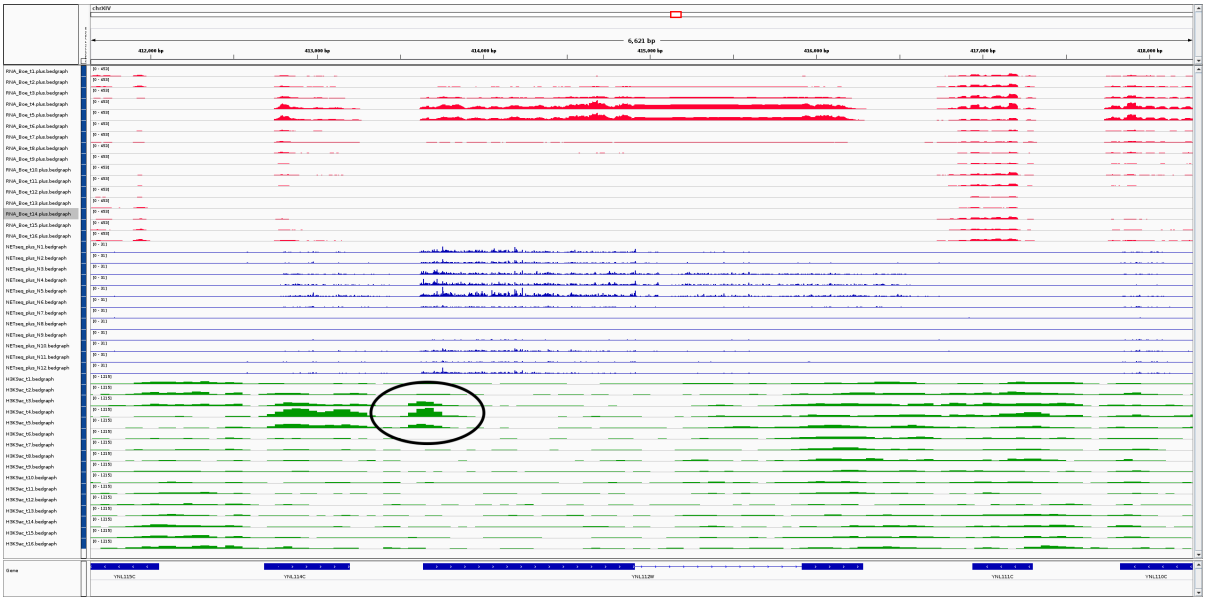


Figure 3.7: Raw sequencing data for gene YLN122W (*DBP*2) using the IGV display. *DBP*2 is the gene belonging to Cluster 3 in this analysis (see Fig. 3.6) that is most similar to the clusters' metagenes, as measured by the Pearson correlation coefficient.
x-axis: base pairs of gene YLN122W (*DBP*2) and its neighbouring genes in Chromosome XIV. Total window span: 6621 bp. Blue boxes at the bottom mark the gene bodies; the narrow white arrows inside them mark the sense of transcription.
y-axis: **Red**: RNA-seq raw data, ordered from sample point t_1 (top) to t_{16} (bottom). At each base pair, the height of the red bar represents the amount of RNA expression measured in arbitrary units.
Blue: NET-seq data, ordered from sample point t_1 (top) to t_{12} (bottom).
Green: H3K9ac ChIP-seq data, ordered from sample point t_1 (top) to t_{16} (bottom). It is worth observing the tight H3K9ac regulation, which is closely matching RNA-seq, and they both are roughly on phase with NET-seq. A strong H3K9ac signal -marked with a **black** oval and enlarged in Figure 3.9- appears at the TSS nucleosome, and it is only briefly active during the OX phase. Similar results were observed for other H3 lysine acetylations (data not shown). YLN122W (*DBP*2) is an ATP-dependent RNA helicase of the DEAD-box protein family [605, 606] which is involved in mRNA decay [607] and rRNA processing [608] and regulates GAL lncRNA-R-loop formation [609]. DBP2's YMC relevance is discussed by Laxman *et. al.* [136].

Note: This figure is too large to be included in one page while being readable; it is reproduced at double page in the offprint at the end of the thesis. The strong H3K9ac signal marked with a black oval is enlarged in Figure 3.9.

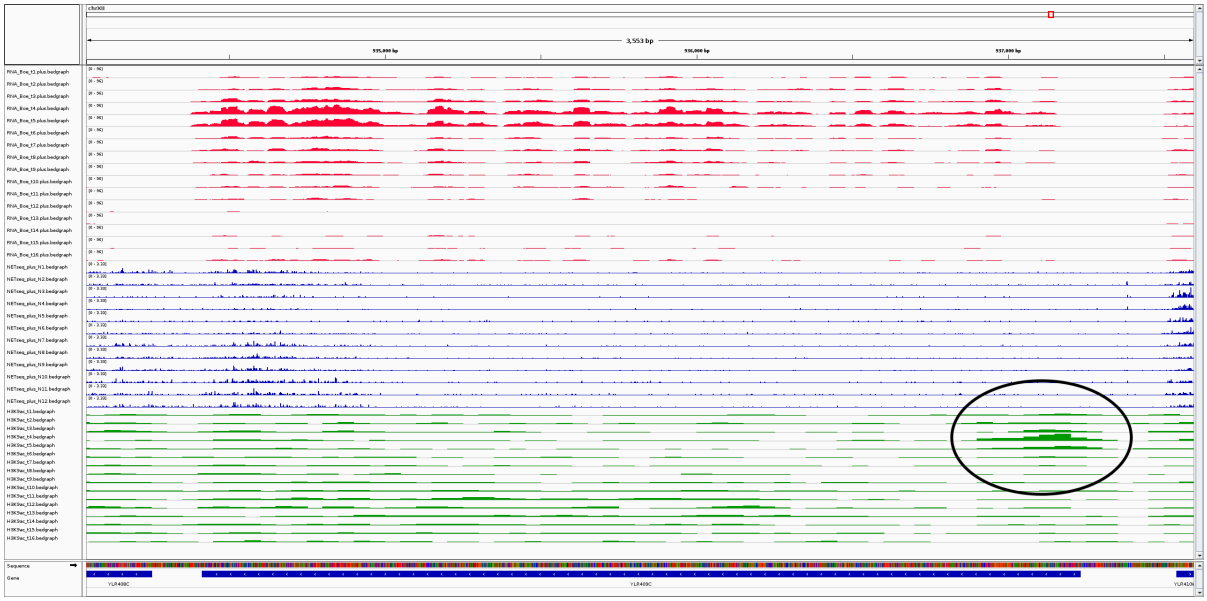


Figure 3.8: Raw sequencing data for gene YLR409C ($UTP21$) using the IGV display. $UTP21$ is the gene belonging to Cluster 8 in this analysis (see Fig. 3.6) that is most similar to the clusters' metagenes, as measured by the Pearson correlation coefficient.
x-axis: base pairs of gene YLR409C ($UTP21$) and its neighbouring genes in Chromosome XII. Total window span: 3553 bp. Blue boxes at the bottom mark the gene bodies; the narrow white arrows inside them mark the sense of transcription.
y-axis: **Red**: RNA-seq raw data, ordered from sample point t_1 (top) to t_{16} (bottom). At each base pair, the height of the red bar represents the amount of RNA expression measured in arbitrary units.
Blue: NET-seq data, ordered from sample point t_1 (top) to t_{12} (bottom).
Green: H3K9ac ChIP-seq data, ordered from sample point t_1 (top) to t_{16} (bottom). A similar H3K9ac *kick* -marked with a **black** oval and enlarged in Figure 3.9- to the one discussed in Fig. 3.7 is present at the TSS, since the gene lies on the $3' \leftarrow 5'$ Crick strand [610]. Similar results were observed for other H3 lysine acetylations (data not shown). YLR409C ($UTP21$) is a subunit of U3 containing 90S preribosome [611, 612] and SSU processome complexes [613] which are involved in production of 18S rRNA [614] and assembly of the small ribosomal subunit [615]. $UTP21$'s relationship with the YMC is briefly discussed in Murray *et. al.* [174, Supl. Mat.].

Note: This figure is too large to be included in one page while being readable; it is reproduced at double page in the offprint at the end of the thesis. The strong H3K9ac signal marked with a black oval is enlarged in Figure 3.9.

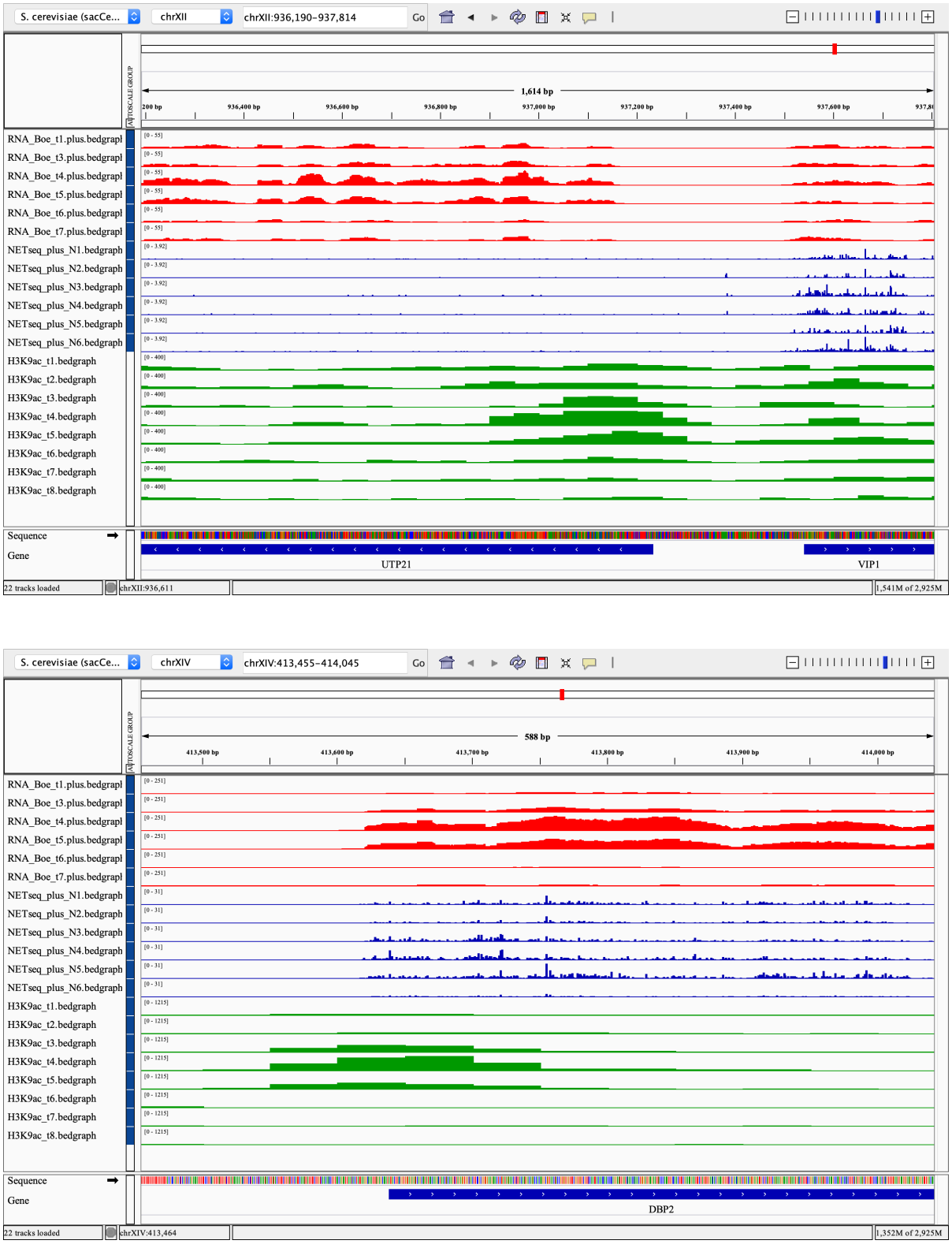


Figure 3.9: Enlarged view of the transitory signals in H3K9ac accumulation observed for genes $DBP2$ (Top) and $UTP21$ (Bottom) in the two previous Figures 3.8 and 3.7.

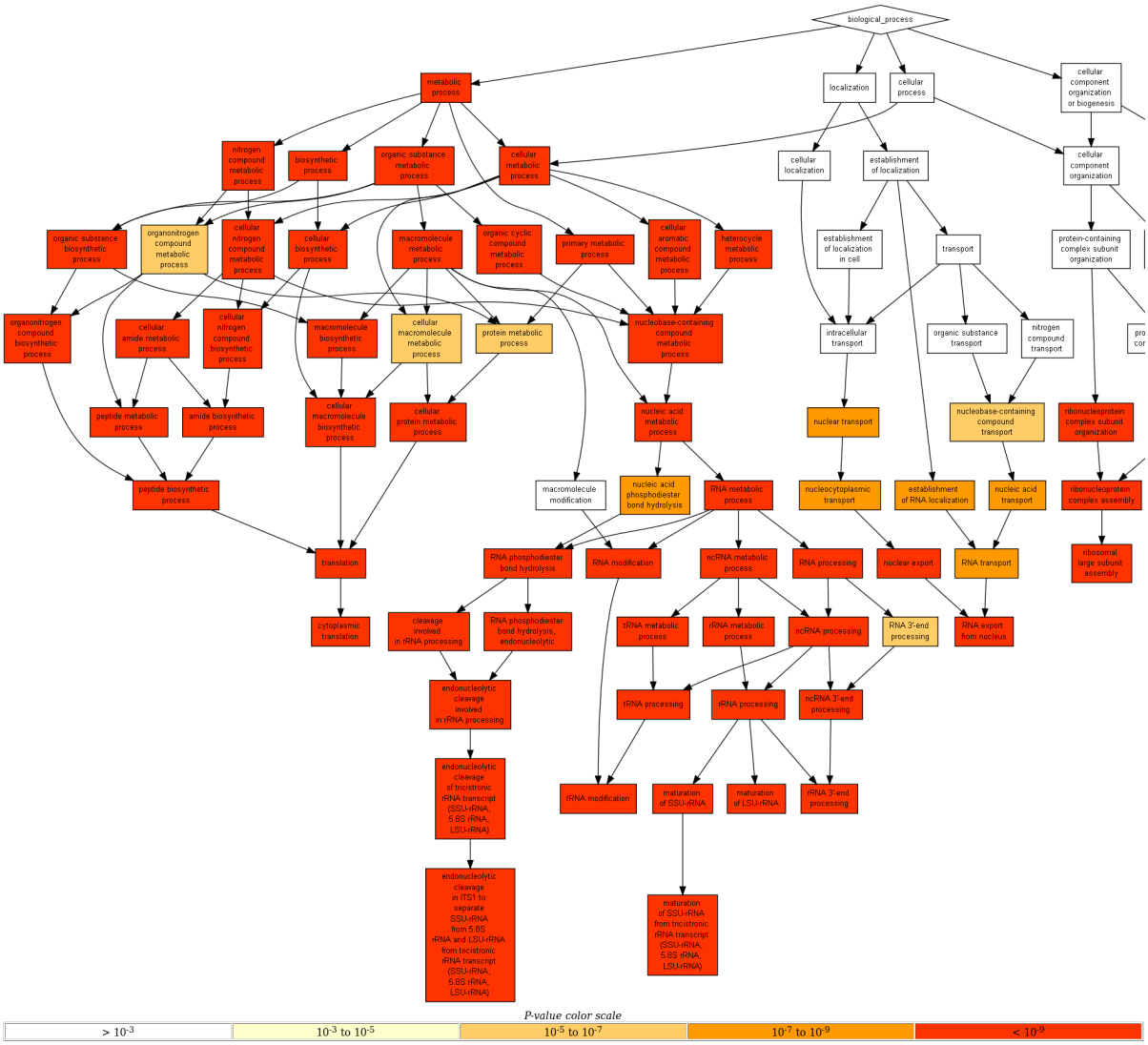


Figure 3.10: GO (GOrilla) tree stemming from the genes included in **Cluster 3** of Fig. 3.6. The low p-value ($p < 10^{-9}$, as discussed in the main text) processes in this tree have been associated with the YMC OX phase genes [127].
GO trees are DAG (directed acyclic graph) graphical representations of the enriched GO terms obtained, with color coding reflecting their degree of enrichment [378]. DAG nodes are the GO terms, and the arrows represent a transition to more specific processes. The first term (node) is always *biological process*, and terms in lower rows are specific differences of the more generic process on upper rows that lead to them in the directed graph.

Note: GO trees are very large; therefore, only the most relevant part of it is represented in order to make the image more readable; the whole figure is reproduced at double page in the offprint at the end of the thesis.

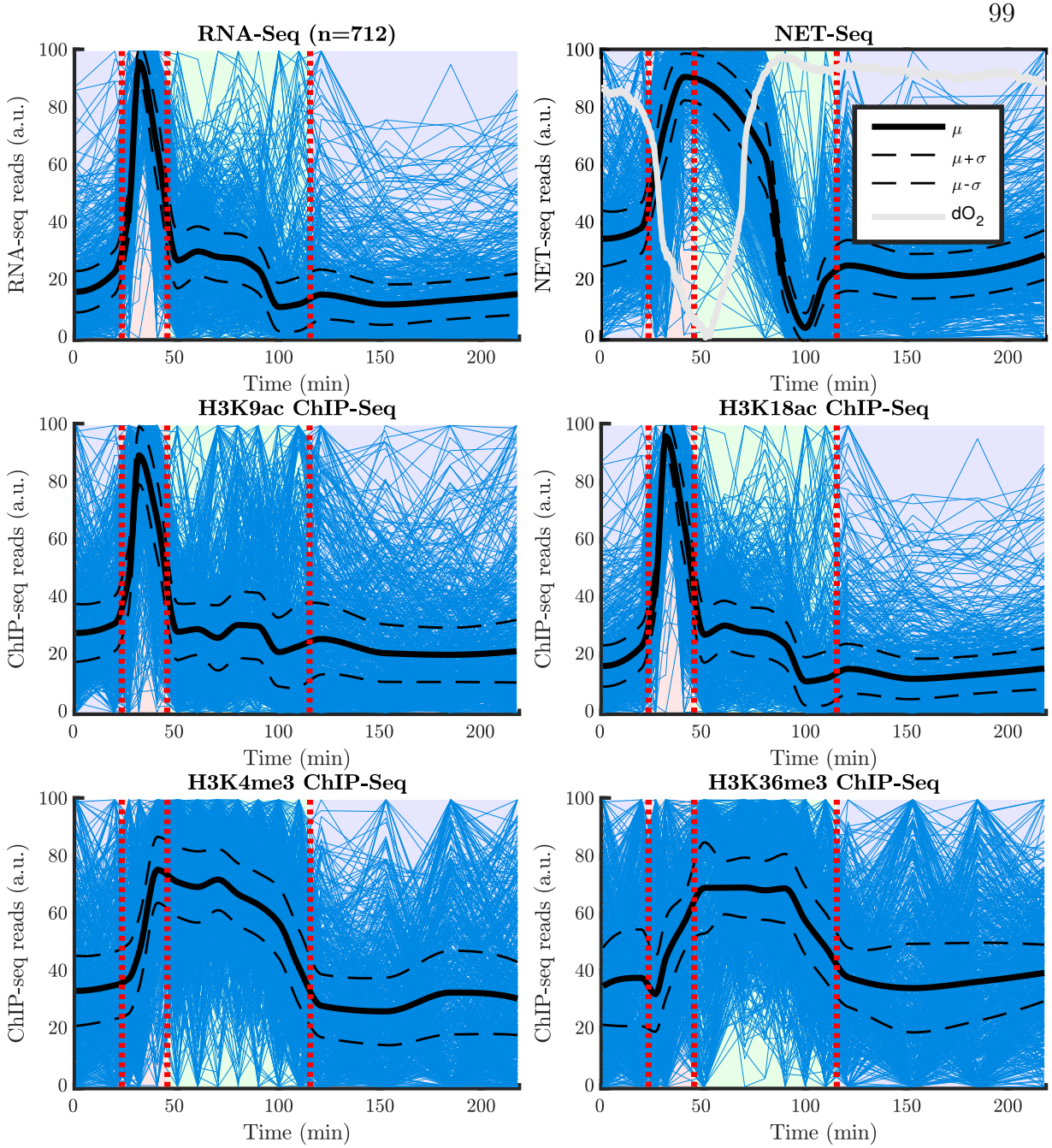


Figure 3.11: Metagenes associated with different HM data for genes belonging to Cluster 3 in Figure 3.6. Blue lines represent individual genes' dynamics. Solid black lines represent the averaged metagene μ; dashed black lines represent $\mu \pm \sigma$, they are a measure of inter-cluster dispersion. Dashed vertical red lines signal the YMC phase transition points in the order RC to OX to RB to RC. On the NET-seq subplot, the chemostat dissolved oxygen concentration dO_2 trace is plotted to indicate the yeast respiration during the YMC. Different HM accumulation dynamics help us to understand how they correlate with transcription. Early-tail H3 lysine acetylations H3K9ac and H3K18ac are highly dynamic, and peak with the appearance of transcripts. However, both $H3K4me_3$ and $H3K36me_3$ start to accumulate when RNA and the lysine acetylations peak, and they remain deposed for most of the RB phase, suggesting that these trimethylations are an epigenetic mark signalling that transcription has already taken place at that gene: they could be repressive to stop reinitiation of transcription during a permissive window of time at the YMC high O_2 consumption phase.

Table 3.3: (Left): Most relevant (in terms of $p-$value) Transcription Factors (YEASTRACT) associated with clusters 3 and 8 in this analysis (see Fig. 3.6). (Right): Gene Ontology terms (GOrilla) associated with such clusters. Both lists are ordered, the first element being the most relevant.

Cluster 3	**Cluster 8**
Met4p	Sfp1p
Yap1p	Cst6p
Sfp1p	Mga2p
Cst6p	Ace2p
Ace2p	Rap1p
Cup9p	Swi5p

Cluster 3	**Cluster 8**
ncRNA metabolic process	ncRNA metabolic process
ncRNA processing	rRNA metabolic process
cellular component biogen.	rRNA processing
ribonucleoprotein biogen.	ribonucleoprot. biogen.
rRNA processing	ncRNA processing
RNA processing	cellular component biogen.

3.5.3 Non-exhaustive clustering results

Exhaustive clustering algorithms are not always able to individuate biological processes regulated by only a handful of genes [616–618], as could be the case for the YMC. The aim of this Subsection is to focus the bioinformatics analysis not on the whole yeast genome, but only on those clusters of genes exhibiting strong simultaneous oscillations in their sequencing data.

The rationale behind non-exhaustive clustering algorithms is to individuate the sets of genes related to the biological process of interest -YMC regulation in this case- but without clustering *everything* [619, 620]. In this Subsection, the non-exhaustive oscillation-shape clustering algorithm presented in Subsection 3.2.2 is applied to several datasets simultaneously, exploring the possible HMs permutations and measuring the quality of the output in terms of the richness of high-significance GO terms (using GOrilla, as in the previous Subsection).

This different approach is justified by the danger of too large a type II error arising because of the requirement that all genes be clustered. To give a numerical example, imagine that a given biological process is regulated by 50 genes. During a cyclic process as the YMC, we would expect these 50 genes to follow rather similar oscillatory patterns in their sequencing data, and therefore being clustered together. But assume that we perform a genome exhaustive clustering into 20 clusters. Then, each cluster will contain ~ 290 genes, so even if this group of 50 genes were to be clustered together, it would only amount to 17% of the cluster population: the FDR for this process would be 83%; therefore, it would be very difficult -if not arbitrary- to discern the cluster-associated GO terms related to such biological process and the ones being a mere artefact introduced by the FDR. Avoiding this problem is precisely the rationale behind the introduction of non-exhaustive multiple-datasets clustering algorithms in our analysis.

Figure 3.12 shows the non-exhaustive oscillation-shape clustering algorithm results for an analysis including co-clustering of multiple datasets: RNA-seq, NET-seq, and H3 lysine acetylation (H3K9ac) ChIP-seq. Very similar results were obtained using H3K14ac or H3K18ac instead of H3K9ac (data not shown). The most striking feature of this figure is the mirrored time-evolution in mRNA, nascent transcription, and HM accumulation levels through the YMC within each of the clusters. By observing the temporal windows at which sequencing is highest for each of the clusters, these three clusters are readily identified with the three YMC phases; each of them peaks at a different and distinct YMC phase.

The next step in this analysis is to examine the common features of the genes belonging to each of these three clusters shown in Fig. 3.12. As in the previous Subsection, this is done by first, using the GO GOrilla tool to investigate the biological processes known to be regulated by genes belonging to each of the clusters, and then applying the YEASTRACT algorithm to discern which TFs are known to regulate a largest portion of genes within the cluster.

GOrilla Analysis For each of the three clusters obtained (see Fig. 3.12), a ranked list of genes was produced by ordering the genes within the cluster from highest to lowest Pearson's correlation coefficient with the cluster metagene. Then, GOrilla analysis for the Biological process ontology in yeast was performed considering a p-value threshold $p < 10^{-9}$ as the cutoff. GOrilla reports only GO terms with a p-value better than the selected threshold [378]; therefore, such a p-value has to be small enough to reflect the fact that, as calculated by GOrilla, it does not include the multiple hypothesis correction on the number of tested GO terms. The Bonferroni correction [621] adjusted with the Benjamini method [622,623] is therefore applied to take into account the combinations of the 5204 possible GOrilla GOID annotations, concluding that a GOrilla $p < 10^{-9}$ value threshold is required to have significantly non-random GO terms. A $p < 10^{-9}$ value is thus restrictive enough so that the FDR is less than 1% (calculation not shown) while, as the results of this Subsection evidence, it is not too restrictive, retrieving a significant number of biological processes known to be associated with the gene products.

Figure 3.13 shows part of the GO tree stemming from genes belonging to Cluster 1 (OX) of this non-exhaustive oscillation-shape clustering algorithm (see Fig. 3.12). GO trees description and how to interpret them was discussed in Subsection 3.5.2 (see Fig. 3.10). Table 3.4 displays the most significant (lowest p-value) GO terms obtained in this analysis: Cluster 2 (RB) produced mitochondrial translation, ATP synthesis, and protein transmembrane transport terms among others, while RC produced redox and catabolic processes, gluconeogenesis, and pyruvate transport terms among others. Results are analogous when other H3 early-tail lysine acetylations (H3K14ac or H3K18ac) are used instead of H3K9ac (data not shown).

YEASTRACT Analysis Table 3.5 displays the TFs associated with the three clusters in this analysis (see Fig. 3.12), obtained using the YEASTRACT algorithm discussed in Subsection 2.4.2. It is important to note that these TFs are unique for

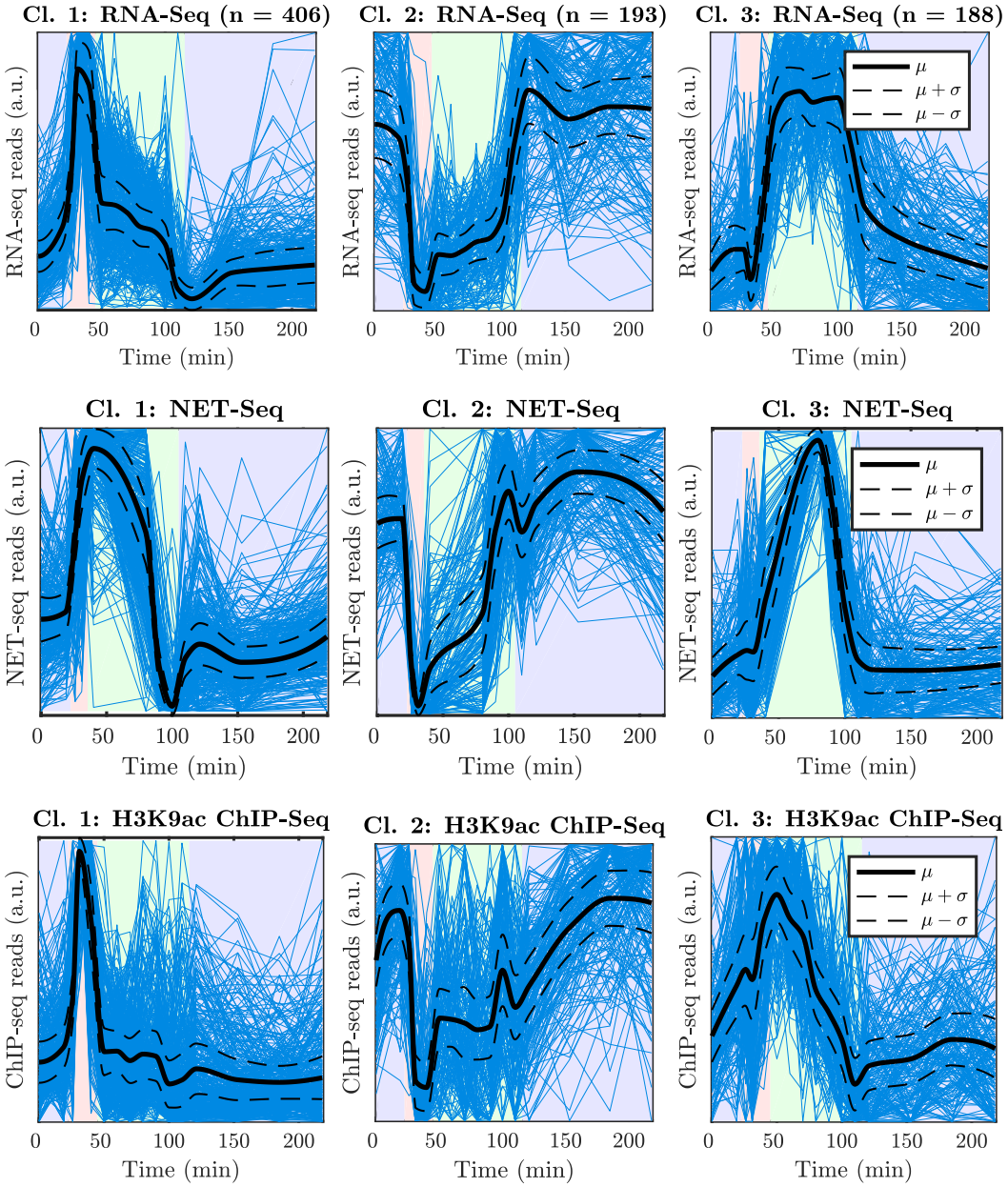


Figure 3.12: Multiple-omics non-exhaustive clustering results using the oscillation-shape clustering algorithm discussed in Subsection 3.2.2. This multiple-omics analysis uses data across time in mRNA, nascent transcription, and H3K9ac accumulation levels during the YMC. The first row plots the normalized RNA-seq expression levels, the second is NET-seq reads, and the third is H3K9ac ChIP-seq. Blue lines represent individual genes dynamics. Solid black lines represent the averaged metagene μ; dashed black lines represent $\mu \pm \sigma$, they are a measure of inter-cluster dispersion. Background colors signal the three YMC phases: RC (Blue), OX (red), and RB (green). Genes in Cluster 1 exhibit a strong peak in all three datasets during the OX YMC phase; similarly, Cluster 2 for RC and Cluster 3 for RB. It is interesting to note how multiple-omics clustering results are genuinely different from separate clustering: for example, Cluster 1 RNA-seq data (top left corner) includes ≈ 10 genes whose expression levels in the $t > 160$ min zone are surprisingly high, while all the other genes have very low expression levels at that time points. The reason for which these genes are included in Cluster 1, rather than Cluster 2 (genes active during RC), is because their NET-seq and H3K9ac ChIP-seq data were very similar to Cluster 1 metagenes.

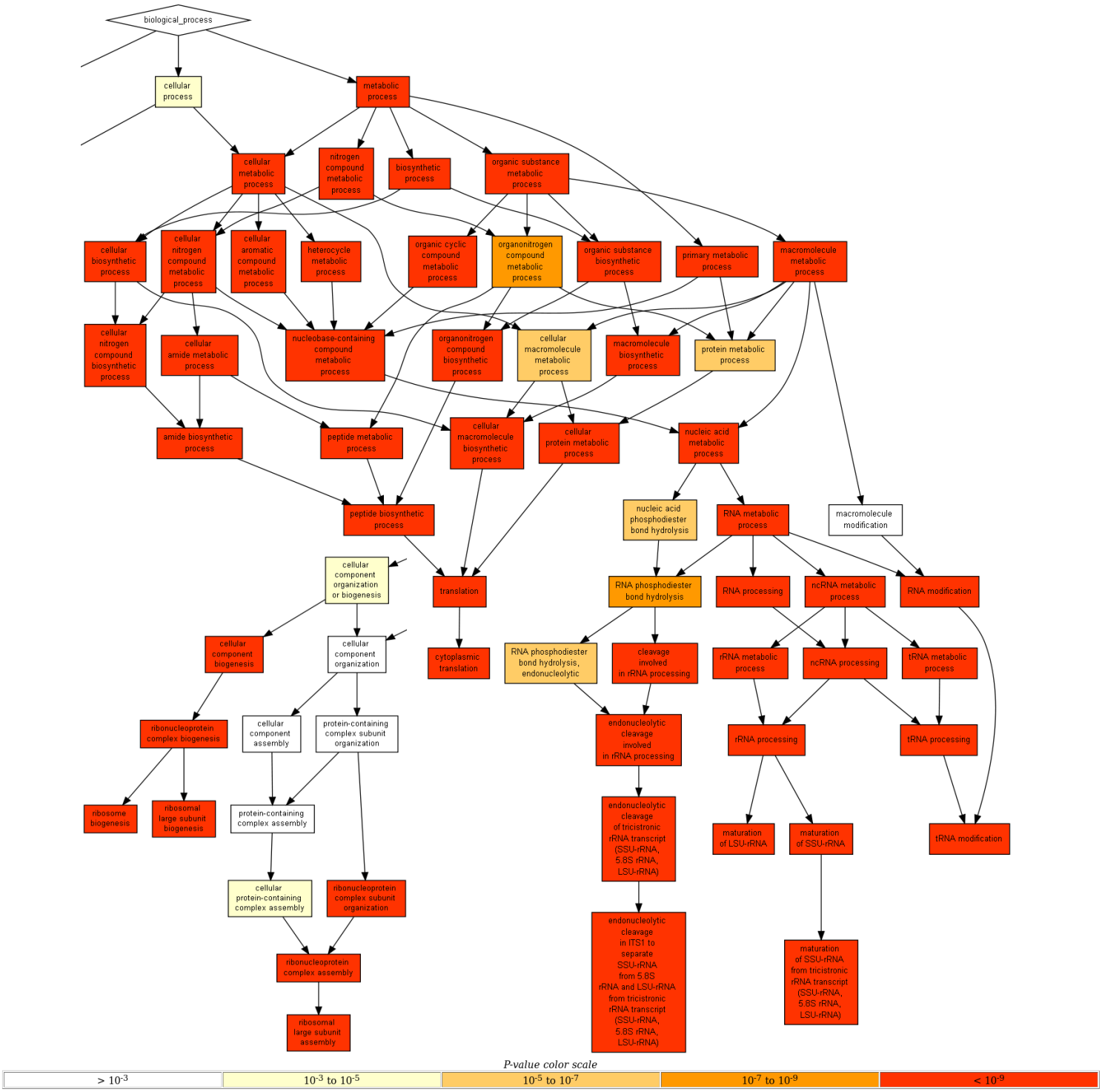


Figure 3.13: GO (GOrilla) tree derived from Cluster 1 (OX) in the non-exhaustive oscillation-shape clustering algorithm discussed in the main text (see Fig. 3.12). GO trees description and how to interpret them is discussed in Subsection 3.5.2 (see Fig. 3.10) Arrows correspond to biological process and cellular process relations. The tree shows processes related to the ribosomal protein genes, a set of genes exhibiting highly periodic oscillations during the YMC [3].

Note: GO trees can be very wide; therefore, the part most on the left of the tree has been placed on the bottom left corner; the whole figure with this part in its original position is reproduced at double page in the offprint at the end of the thesis.

Table 3.4: Gene Ontology (GOrilla) terms derived from the three clusters obtained using the non-exhaustive oscillation-shape clustering algorithm discussed in the main text (see Fig. 3.12). Lists are ordered, the first element being the biological process individuated with highest confidence by the GOrilla repertoire.

Cluster 1 – OX	**Cluster 2 – RC**	**Cluster 3 – RB**
ribonucleoprotein comp. biogen.	oxidation-reduction process	mitochondrial translation
cellular component biogenesis	cellular lipid catabolic process	energy coupled p^+ transport, down electrochem. grad.
ncRNA metabolic process	carbohydrate metabolic process	ATP synthesis coupled proton transport
ncRNA processing	lipid catabolic process	mitochondrial transmembrane transport
rRNA processing	antibiotic metabolic process	protein transmembrane import into intracell. organelle
ribosome biogenesis	monocarboxylic acid transport	protein localization to mitochondrion

Table 3.5: Transcription factors (YEASTRACT) derived from the three clusters obtained using the non-exhaustive oscillation-shape clustering algorithm discussed in the main text (see Fig. 3.12). Lists are ordered, the first element corresponding to the TF individuated with highest confidence by the YEASTRACT algorithm, which compares the fraction of genes in the cluster regulated by a given TF with the fraction of genes in the whole genome regulated by that TF.

Cluster 1 – OX	**Cluster 2 – RC**	**Cluster 3 – RB**
Met4p	Adr1p	Thi2p
Yap1p	Bas1p	Nrg2p
Sfp1p	Gis1p	Tye7p
Cst6p	Msn2p	Xbp1p
Ace2p	Sok2p	Hap2p
Cup9p	Msn4p	Hap4p

each cluster: this analysis does not individuate any TF regulating more than one cluster. The RC phase is regulated, amongst others, by the TFs **Msn2p** and **Msn4p**, which have been suggested as the main YMC regulation TFs [151, 624, 625]. Other retrieved TFs that have been individuated in the literature as key YMC regulators are **Met4p** [174, 626, 627], **Adr1p** [127, 174, 628], **Yap1p** [3, 56, 629], **Bas1p** [127, 472, 630], and the lesser-known **Thi2p** [630–632].

3.6 Conclusion

Throughout this Chapter it has been discussed how most of the yeast genome can be clustered using a multi-omics approach by their similar RNA-seq, NET-seq, and ChIP-seq temporal dynamics during the YMC. The results of this Chapter contribute to defend the validity of the YMC mathematical model that will be proposed in Chapter 4, since it shows that large sets of genes share common patterns not just regarding gene expression, but also of nascent transcription and deposition of covalent modifications in lysine residues at the histone H3 tail, providing good grounds for the hypothesis that these large sets of genes (or clusters) are co-regulated.

The different clustering algorithms built in Section 3.2 and compared in Section 3.4 share one common pattern: during the YMC, large ($n > 100$) groups of genes follow entirely similar expression dynamics, being active during precise periods of time. This finding confirms that the YMC is a highly compartimentalized system on which different yeast metabolic processes are segregated into different temporal moments, potentially to increase the efficiency of yeast metabolism [3, 633, 634]. Metabolic compartments combined with intercellular coordination have highly beneficial effects from an evolutionary biology point of view [635]: by coordinating the production and secretion of highly acidic metabolites [113][4] with ethanol secretion [138] during the OX phase, yeasts achieve an extraordinarily hostile environment on which very few other organisms can proliferate [636–639].

In Section 3.3 it was discussed how exploring the number of clusters of genes into which the YMC ought to be divided leads to the conclusion that there is not a unique (or even majoritarian) protocol in the literature to establish how many clusters stem from a particular dataset. This is relevant because sometimes scientific discussions tend to be centred about the number of clusters resulting from a particular experiment, rather than about the data themselves.

The measurement of clustering similarity performed in Section 3.4 revealed that the distance calculated between different clusterings of the Li and Tu datasets is consistently lower than the distance between clusterings of the Boeke dataset and either of such datasets, a result not dependent on the number of clusters selected (see Fig. 3.5). This finding reveals a considerable similitude between the two Microarray

[4]The chemostat buffer keeps the media acidity level constant at a pH=2.5 during the YMC. However, if no buffer was added, the intercellular media would become highly acidic (pH < 2) during the OX phase.

datasets compared to the RNA-seq dataset. Interpreting this result together with the experimental features of each of the corresponding experiments (see Table 3.1), it can be concluded that the sequencing technology employed (Microarray vs. RNA-seq) has a greater influence over the measurement results than other experimental differences, such as the yeast strain used or the addition of PZ after the first cycle of the experiment.

The multiple-omics clustering algorithms presented in Section 3.5 found a strong relationship between transcription and system progression through the YMC. This result provides good grounds for the hypothesis (to be tested in Chapters 4 and 5) that the YMC is, essentially, a transcriptional oscillator whose dynamics and regulation are achieved via biological processes that are internal to the cell, on which inter-cellular synchronization via metabolites concentration in the inter-cellular media plays only a relatively minor part.

A novel feature of the analysis performed in this Chapter is the introduction of transcription data, whereas the previous literature in the field was focused on transcript abundance analysis. In this light, Cluster 8 in Figure 3.6 (see Subsection 3.5.2) leads to new insights into the transcription/transcripts distinction, since I speculate that some genes with similar transcript dynamics are undergoing very different regulatory pathways, potentially via transcripts degradation control. The possible interplay of this mechanism with the transcript lifetimes calculations performed in Chapter 2 will be explored in Chapter 4.

To explore the transcriptional activity regulating this oscillator, the TFs associated with each of the three YMC phases were retrieved in Section 3.5.3 (see Table 3.5). These TFs are natural candidates for being the key YMC regulators, explaining thus their co-expression through the YMC. It is interesting to compare the individuated TFs (see Table 3.5) with other TFs-retrieving algorithms in the literature [150,151,174] to verify that some TFs are retrieved across different experiments, like Sfp1p [150,151,174,640], Cup9p [174,631,641] and Ace2p [150,151,318,642,643] for the OX phase, Msn2/4p [127, 150–152, 172, 174], Gis1p [150, 151, 174, 644, 645] and Adr1p [127,152,646,647] for RB or Xbp1p [151, 152, 174, 648, 649] for RC. The Msn2/4p influence in the YMC has been widely studied in the literature [150–152]. The $XBP1$::$3HA$ strain bears only a C-terminal tag, and shows defective YMC oscillations [150, Fig. 2C]. For the other TFs, it would be interesting to examine the corresponding mutant deletions to learn if they can cycle: for the key YMC regulation TFs listed in Table 3.5, I predict viable strains that, however, are unable to cycle.

Early-tail H3 lysine acetylations H3K9ac and H3K18ac are highly dynamic during the YMC, and peak with the appearance of transcripts, as discussed in Subsection 3.5.2 (see Fig. 3.11). By contrast, H3 lysine trimethylations H3K4me$_3$ and H3K36me$_3$ appear later and are maintained for approximately 40 minutes. This result is consistent with the conclusions of Chapter 2 (see Subsections 2.6.1 and 2.6.2), providing thus more evidence in support of the hypothesis that H3K4me3 and H3K36me3 are not instructive for transcription [650, 651]; indeed, they could be potentially repressive to stop reinitiation of transcription [334, 652] along a permissive window of time during the high O_2 consumption phase [279].

The metagene profiles obtained in this Chapter, such as the metagenes of the non-exhaustive time course clustering algorithm (see Fig. 3.12) are a good proxy for the overall multi-gene system dynamics, and will be the *building blocks* of the YMC mathematical model built in Chapters 4 and 5, being the target form of the oscillatory solutions of the systems of ODEs that will be developed.

Chapter 4

A Single-cell Transcriptional Oscillator

This Chapter aims to build a transcriptional oscillator able to explain and reproduce, at a cellular level, the dynamic patterns in gene expression stemming from the compartmentalization of metabolic activity in the YMC. The bioinformatics analysis in previous Chapters has provided metagene profiles for clusters of genes that describe their evolution through the YMC: the goal of this Chapter is to formulate a system of multiple Ordinary Differential Equations (ODEs) that replicates such dynamics.

A preliminary purpose of this thesis, that will be extended through this model, is to construct a set of mathematical expressions able to match, and hopefully explain, the multi-omics expression patterns retrieved in Chapter 3. The objective is to build a transcriptional oscillator that matches the YMC data from both the literature [3,108,127] and the unpublished work of my research group [158,373], reflecting the idea that biological time can be thought as a particular topological space [11] on which biochemical oscillators naturally arise [11,653,654].

Some early literature research on oscillators describing circadian clocks [655–658] sets us on route for more specific yeast metabolism models [160–162]. Most of these models use multiple differential equations, even as many as 80 [165], using up to 120 mutants to form a complex ODEs[1] system [659]. Some modelling techniques widely

[1] Ordinary Differential Equations are a special class of differential equations (functional equations including derivative terms) containing one or more functions of the independent variable and its derivatives. In partial differential equations (not used in this thesis), functional dependence occurs with respect to more than one independent variable.

used in the literature that will be employed in this Chapter include modelling biochemical oscillators using ODEs [660], Monte Carlo methods [661], and parameter sensitivity analysis [662].

Yeast glycolitic and glucose oscillations have been modelled by Rizzi *et. al.* in 1997 [663], in a paper concerned with modelling the Terminal Complement Complex (TCC) and the $ADP \leftrightarrow ATP$ exchange. This analysis is entirely different from ours, since his data predate the use of 21st century sequencing techniques to understand the YMC [3, 108].

This Chapter models the behaviour of an isolated yeast cell in chemostat conditions by constructing a transcriptional oscillator that matches and explains the RNA-seq, NET-seq, and ChIP-seq data, while making some predictions. Interactions with other cells or the extracellular media are not considered for the time being: they will be the matter of Chapter 5. The Chapter starts by showing how a transcriptional oscillator is a representation of yeast RNA and protein oscillatory dynamics described by ODEs (Section 4.1), detailing afterwards the methods used to solve them (Section 4.2) and exploring different possibilities for the inter-component feedback interactions (Section 4.3). The model equations are then formulated for both RNA and protein concentration dynamics (Section 4.4). The second part of the Chapter explores the numerical solutions of such equations: by calculating the best-fitting model parameter values (Section 4.5) and using numerical methods and simulation to solve these equations (Section 4.6), the complete model is established; and can be tested by examining the effect that single parameter changes have on the oscillations (Section 4.7), showing by comparison with more complex systems that three is the optimal number of components for a YMC-mirroring oscillator (Section 4.8). The conclusion of this Chapter is that the YMC can be reproduced as a robust three-phases transcriptional oscillator (Section 4.9). Computational details and program calls related to the calculations and simulations are discussed in Appendix A.6.

4.1 RNA and protein dynamics define the YMC

In order to replicate the RNA-seq metagene profiles that were produced by the multilevel clustering analysis of Chapter 3 with the information retrieved in Chapter 2, I aim to build a transcriptional oscillator able to replicate the RNA expression dynamics, therefore explaining the regulatory logic underlying the YMC.

In previous Chapters, the bioinformatics analysis was centred on the mRNA concentration (as measured by RNA-seq) evolution during the YMC; these results will act as the first of the 'building blocks' on which the mathematical models of this Chapter will rely. The first step towards the construction of the model consists of a quantitative fitting of the time evolution patterns retrieved in the previous analysis at Chapter 3, while obtaining a system that reflects other characteristics of the YMC, such as the oscillation period, the length of each of the three YMC phases, and the effect of single-parameter variations on these features. In the future, it would be interesting to explore also the behaviour of some relevant mutants.

The idea, sketched in Figure 4.1, is to model the previous time evolution patterns using a system of ODEs whose solution mirrors the metagene profiles shown in Figure 4.1.a. Observe how in Fig. 4.1.b the switch on/off times, and not just the shapes, match the OX genes metagene oscillations of Fig. 4.1.a in the $t > 55$ h region, which is when the yeast cells have entered the YMC. The model developed in this Chapter will replicate such switch on/off shapes and ratios, while Chapter 5 will be devoted to explain the transition from an unsynchronized yeast population to synchronized oscillations, a key characteristic of the YMC.

Following the central dogma of molecular biology [664], this model includes differential equations for both transcription and translation, as it is standard practice in the mathematical modelling of cellular dynamics literature [665–670], particularly in the case of regulatory networks identification via gene expression measurements in different systems [478, 671–674]. Unfortunately, proteomics experiments undertaken by our group did not produce a conclusive outcome, since no proteomics analysis can produce an exhaustive measurement of the proteome [675, 676]. This is particularly relevant in the case of TFs, which are key proteins in the genetic regulatory machinery [677–680], but have a very low concentration (few molecules per cell) [681, 682], which means they are not detectable by most proteomic analysis. Therefore, the protein concentration levels predicted by this mathematical model could not be experimentally validated; they will be eventually validated in the future.

Other information gathered in previous Chapters, like the HM time evolution, the yeast cells' oxygen consumption rate, or the mRNA degradation rates, will be used as part of this model. However, they will be used to perform a parametric fit of the RNA and protein ODEs and will not have an associated ODE by themselves; for example, mRNA degradation rates will be used to establish the values of the relevant constants of the RNA concentration ODEs in Section 4.5.

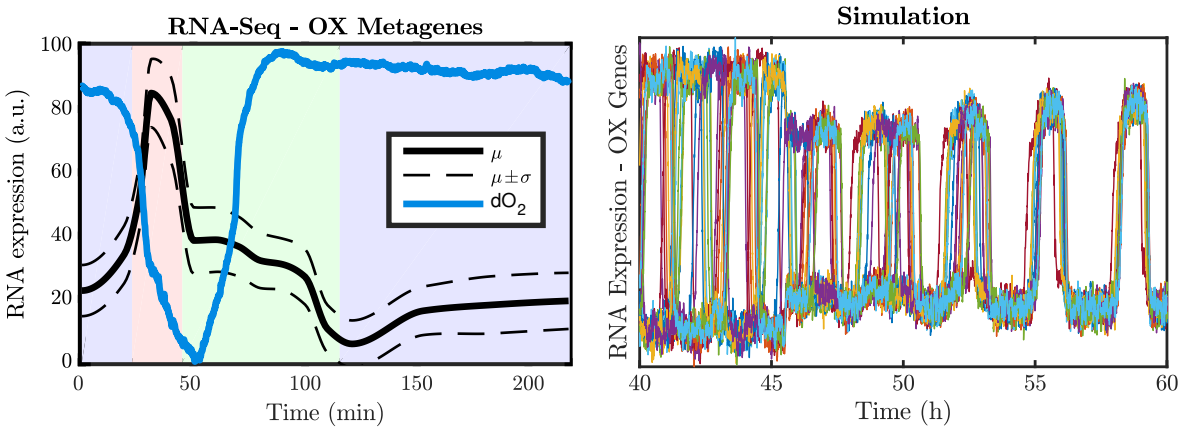


Figure 4.1: Sketch illustrating the correspondence between experimentally measured metagenes and ODE solutions in the YMC model. (Left) Oscillations in mRNA expression for OX cluster genes (see Chapter 3). Solid black line: metagene of the OX genes cluster retrieved from Fig. 3.6 metagene. Dashed black lines: standard deviation of gene expression from the metagene, they are a measure of inter-cluster dispersion. Background colors signal the three YMC phases: RC (Blue), OX (red), and RB (green). The individual genes expression dynamics is not shown. (Right) ODE model exhibiting similar oscillations in shape and period as the metagene on the Left. The cell enters the YMC in the $t > 55$ h region, and then oscillates regularly. Different lines correspond to different simulation realisations, which eventually converge.

Nomenclature Through the following Sections, genes will be named in italic upper case letters (*PER1*) and proteins in roman lower case letters (Per1p), consistent with the Species-Specific Convention's nomenclature for *S. cerevisiae* [683, 684]. Unless otherwise indicated, letters by themselves correspond to concentrations: A stands for $[A]$.

4.2 Ordinary differential equations form the basis of the model

The mathematical models presented in this Chapter have been formulated using non-linear ODEs, also called *rate equations*, in which the species' production and degradation rates are only a function of the concentration of that and other species in the

medium, a standard assumption in the yeast modelling literature [659, 660, 685–687]. The use of Partial Differential Equations [688–691] was also considered, but it did not seem to bear any biological meaning for this particular case: RNA and protein synthesis and degradation processes are regulated by the concentration of the species in the media [692–694] and not by their derivatives [695, 696].

ODEs are a widely used tool to model and explain time-dependent biological processes [697], since these systems do not exhibit a large *memory* [698] or hysteresis [699]: the system evolution does not depend on the past statuses of the system, but just on the present concentrations of the components.

Some efforts have been made to analytically solve nonlinear ODEs [700–702], in particular for systems of 6 coupled ODEs [703] like the transcriptional oscillator that will be built in this Chapter. The main difficulty with this approach is that, in general, these equations do not have an exact, or analytic, solution [704]: it is necessary to use numerical methods to find their solutions [705]. Nowadays, this is not a big problem in biology, since current computers are strong enough to make the difference in outcome from the numerical, or approximated, solution to its exact value so small that it lays several orders of magnitude below the standard error associated with laboratory measurements of the corresponding biological processes [706, 707], which makes it a negligible source of error for the purposes of this model.

The following two Sections explore the form that the ODEs describing RNA and protein concentrations must have, considering what we know about their synthesis and degradation and knowing the solutions must resemble the multi-omics data of the YMC.

Computational details As explained in Appendix A.6, where computational details and program calls are discussed at greater length, the numerical solution of the nonlinear ODEs presented in this Chapter was obtained using the MatLab SimBiology package [708, 709], a MatLab-integrated software allowing for simple modifications of conditions and for simulations of interacting cell systems (see Chapter 5). Throughout this Chapter, unless otherwise stated, equations have been implemented and then numerically solved using the SimBiology `sbiomodel` model object to input the species (RNA and protein), their constants, and their ruling ODEs. The model is then solved using the `sbiosimulate` package with the ODE numerical solution algorithm `ode23` [710–712], which is an implementation of an explicit Runge-Kutta method [713, 714]. To confirm the reliability of the analysis, in every model it was

checked that the maximum distance between any two consecutive time points in the simulation solution was small enough that it could not affect the simulation outcome.

4.3 Regulation mechanisms and feedback loops

In every biological system, the biochemical composition of the environment affects the concentration of RNA and protein associated with any given gene, a process known as gene regulation [715, 716]. The complete mechanisms guiding gene regulation at a given cell in any circumstance are poorly understood [717–719]; however, endless factors and mechanisms have been individuated for particular genes and cellular by-products, particularly in yeast [720–723].

These mechanisms in which the presence of a particular molecule, or complex of molecules, has an effect on the transcription or translation associated with a gene are known as regulation mechanisms, and their modelling and description is one of the main objectives of mathematical biology [39, 724–726]. Regulation mechanisms can be broadly divided into two main categories: positive regulation $A \to B$, on which system A amplifies system B, and negative regulation $A \longdashv B$, on which system A inhibits system B. Regulation mechanisms can be combined to form closed regulation paths, also known as feedback loops, on which system A regulates system B and vice versa. For example, $A \longdashv B \longdashv A$ is a feedback loop on which both A and B repress each other.

The interrelation of different feedback loops implies that cellular biology cannot be described by several, independent, RNA/protein oscillators, but rather by one multiple, coherent, system [727]. Not every combination of negative and positive feedback loops is capable of producing oscillations [660], therefore the aim of this Section is to individuate those combinations of loops that allow for oscillations that are sustained in time [728]. In particular, we are interested in retrieving, for the YMC, the functional dependence of RNA/protein concentration in one phase on the different species concentration in other phases.

Feedback loops may have endless functional forms [729]; therefore, this Section will be concerned with the most simple forms of negative and positive feedback loops: a more detailed study of the precise mathematical form of the loops in the YMC requires a biochemical discussion by its own, and is therefore deferred to Section 4.4.

Regulation elements can be identified with clusters of genes The clusters of genes associated with each of the three YMC phases could be forming a control circuit of key genes that regulate each of the clusters, creating a feedback loop responsible for the regulatory machinery underlying the YMC dynamics and hence driving the observed RNA and protein oscillations, an idea that will be tested in this Chapter.

4.3.1 Positive regulation mechanisms

Let C_1 and C_2 be two clusters of genes. Assume we know from a biological experiment that the cellular concentration of proteins coded by genes belonging to cluster C_1, which we call $[Prot_1]$, enhances the rate of transcription of genes in C_2, increasing therefore their RNA concentration, $[RNA_2]$, forming this way a positive regulation system $[Prot_1] \rightarrow [RNA_2]$. A priori, the functional form of the interaction is not known, but we know from the literature [478,730–732] that, for a wide number RNA synthesis processes, it can be mathematically approximated as

$$\frac{d[RNA_2]}{dt} = \frac{K_1 + [Prot_1]^p}{K_2 + [Prot_1]^p}; \qquad K_2 >> K_1, \quad p \in \mathbb{N}, \tag{4.1}$$

where K_2 and K_1 are specific to this particular interaction and the integer p is a non-linearity term widely present in the literature [658,733,734]. These interactions, even when combined with negative regulation mechanisms, are much less common than pure negative regulation mechanisms, which control most of the relevant biochemical mechanisms in yeast [735, 736].

4.3.2 Negative regulation mechanisms

Let C_1 and C_2 be two clusters of genes. Assume we know from a biological experiment that the cellular concentration of proteins coded by genes belonging to cluster C_1, which we call $[Prot_1]$, acts as a repressor for transcription of genes in C_2 forming this way a negative regulation system $[Prot_1] \longdashv [RNA_2]$.

Negative regulation can act via two main channels [737]: interactions where the signal molecule does not cause the dissociation of the transcription factor from the DNA [738, 739], and interactions where the molecule acts by sequestering transcription factors off the DNA [740, 741]. Negative regulation has been identified as the essential control mechanism for gene regulation [742–745]: its use will therefore be extensive throughout this Chapter. We know from the literature [736, 737, 745] that a frequent functional form for these interactions is

$$\frac{d[RNA_2]}{dt} = \frac{K_1}{K_2 + [Prot_1]^p}, \quad p \in \mathbb{N}. \tag{4.2}$$

The non-linearity term p range of values $3 \leq p \leq 6$ is explained by the cooperativeness of the relevant TFs [746, 747]. Higher values of p will widen the parameter space on which oscillations are observed [660], although it could be argued that $p > 6$ values do not bear a biochemical significance, since the TFs cooperativeness only contributes up to a $p \approx 6$ polynomial term [748].

In the case of protein synthesis, it is also frequent to encounter 'eigen' negative feedback loops [749] $[Prot_1] \longdashv [Prot_1]$, on which the abundance of the own protein concentration dampens new translation [750–752].

4.3.3 Different functional forms for the regulation mechanisms

In general, an arbitrary interaction regulating mRNA dynamics via protein concentration can be represented by its functional form

$$\frac{d[RNA_1]}{dt} = f([Prot_1], \ldots, [Prot_n]) - \tau_1 \cdot [RNA_1], \tag{4.3}$$

where $[Prot_1], \ldots, [Prot_n]$ are the protein concentrations in the n clusters regulating $[RNA_1]$ and τ_1 is the degradation rate of RNA_1. In the case of a positive regulation mechanism $[Prot_i] \rightarrow [RNA_1]$, f will be a monotonic increasing function in its ith component, while on a $[Prot_1] \longdashv [Prot_1]$ mechanism it will be monotonically decreasing. In a first approximation, there are two effects regulating the C_1 cluster gene expression dynamics: how much protein $[Prot_1]$ there is in the cell, and how much of the other proteins there is. Therefore, f can be divided as

$$f = f_1([Prot_1]) + f_2([Prot_2], \ldots, [Prot_n]).$$

There must be an upper bound for f_1 maximum possible value, so it can be approximated as,

$$f_1 = \begin{cases} 0, & \text{if there is no loop } Prot_1 \longdashv RNA_1 \\ \frac{K_{max}}{e^{(Prot_1 - Prot_1^{opt})/k_F}}, & \text{otherwise} \end{cases} \tag{4.4}$$

A second possible form of Eqn. (4.4) is the Fermi-Dirac statistics [753, 754]

in which an optimal protein $Prot_1$ concentration level, $Prot_1^{opt}$ is assumed [755]; synthesis is $\approx K_{max}$ when $[Prot_1] < Prot_1^{opt}$ and ≈ 0 when $[Prot_1] > Prot_1^{opt}$.

4.3.4 Combination of interactions to form loops

Having established the functional form of individual interactions, these different regulation mechanisms can be combined in loops to form a transcriptional oscillator [756]. The most simple biochemical oscillator there can be includes transcription and translation via the control circuit in a two-component system [660], since only one component is not enough to produce indefinite oscillations [757, 758].

The goal of this Subsection is to build a simple, two-component biochemical oscillator that serves both as an illustration of the concepts expressed in previous Sections, and as a base for more sophisticated mathematical models. By showing that the RNA–protein and synthesis–degradation terms separation of the ODE system can produce indefinite oscillations, and by exploring the adequate regulatory logic, it should allow for a YMC model that identifies its more relevant regulatory features.

We start with a simple oscillator given by the interaction of two mRNA/protein components, whose dynamics is sketched in Figure 4.2. Similar systems, focused on different regulatory pathways and cellular components, have been proposed in the literature [660, 759, 760], also to model yeast cells [161, 761]. In a first approach, this system should be as adaptable as possible, in order to obtain a multi-parametric, highly general, ODE formulation that can form the basis of the forthcoming models. Let us consider an RNA – Protein interaction system as described in Fig. 4.2 exhibiting the –very generic– functional form

$$\begin{array}{l} \frac{d[RNA_1]}{dt} = \frac{T_{\max}}{1+K_1\cdot[Prot_1]^p+K_1'\cdot[Prot_2]^p} - K_{D_1}\cdot[RNA_1] \\ \frac{d[Prot_1]}{dt} = S_1\cdot[RNA_1] - \frac{Z_1}{K_M+[Prot_1]}\cdot[Prot_1] \\ \frac{d[RNA_2]}{dt} = \frac{T_{\max}}{1+K_2\cdot[Prot_1]^p+K_2'\cdot[Prot_2]^p} - K_{D_2}\cdot[RNA_2] \\ \frac{d[Prot_2]}{dt} = S_2\cdot[RNA_2] - \frac{Z_2}{K_M+[Prot_2]}\cdot[Prot_2] \end{array} \tag{4.5}$$

where $T_{\max}$ is the maximum transcription rate and is given by Pol II abundance, S_i are the protein synthesis rates, K_{D_i} are the mRNA degradation rates and the other constants are system-dependent and have to be experimentally measured. The computational implementation of this model using SimBiology ODEs is presented in Appendix A.6.1.

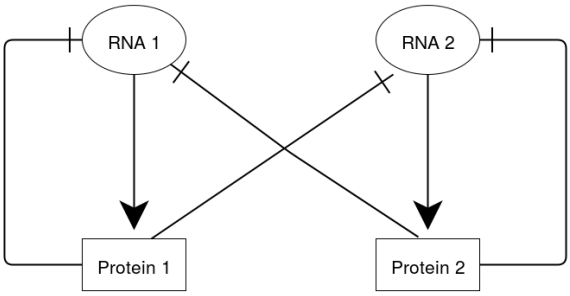


Figure 4.2: A regulation model involving a two element control circuit. Each cluster of genes codes their own proteins, and both proteins inhibit each of the RNA production processes, a system that could be described by the generic form of Eqns. (4.3). In Appendix A.6.1 is presented a Matlab SimBiology computational implementation of this control circuit.

ODE systems of this form will always have a solution within the biochemical boundary conditions [762], given by requirements such as continuous processes and all of the magnitudes being positive at all times. When protein concentration is low, $[RNA_i]' \approx 1 - K_{D_i} \cdot [RNA_i]$, and it is impossible to have $[RNA_i] < 0$, since when $[RNA_i]$ tends to zero, $[RNA_i]'$ tends to 1. Conversely, when RNA concentration is low, $[Prot_i]' \approx \frac{Z_i}{K_M + [Prot_i]} \cdot [Prot_i] > 0$.

Therefore, this system is self-maintained within the biochemically relevant boundary conditions. Similarly, it is easily proved that all the single reactants' concentrations have an upper bound. In this case, there are two possible solutions: either a steady-state or an oscillatory solution. We turn next to the task of examining which parameter combinations result in oscillatory solutions, since the YMC is an oscillation, not a steady-state system.

A further requirement is that the oscillatory solutions to the ODE system set by Eqns. (4.5) for RNA concentration must be similar to the RNA expression dynamics through the YMC, as represented by the time courses that were produced in Chapter 3.4. The simulations' inputs are the different values of the constants involved in Eqns.(4.5), and ansatz values can be retrieved using random scattering in parameter space followed by an adapted Simulated Annealing algorithm [763, 764], which is discussed at greater length in Subsection 4.6.1.

The result, displayed in Figure 4.3, retrieved oscillations for the two-component model with negative feedback loops. However, no adequate solutions were found

when replacing one of the two negative-feedback loops sketched in Fig. 4.2 with one of the positive regulation equations discussed in Subsection 4.3.1. It has been shown [660] that it is much less frequent to obtain biochemical oscillators when positive feedback is present.

The two-component oscillator models discussed in this Section do not reproduce the YMC dynamics obtained in Chapter 3. Therefore, a $n > 2$ components oscillator is required to model the YMC, as will be explored in the next Section, on which three-phases oscillators are simulated.

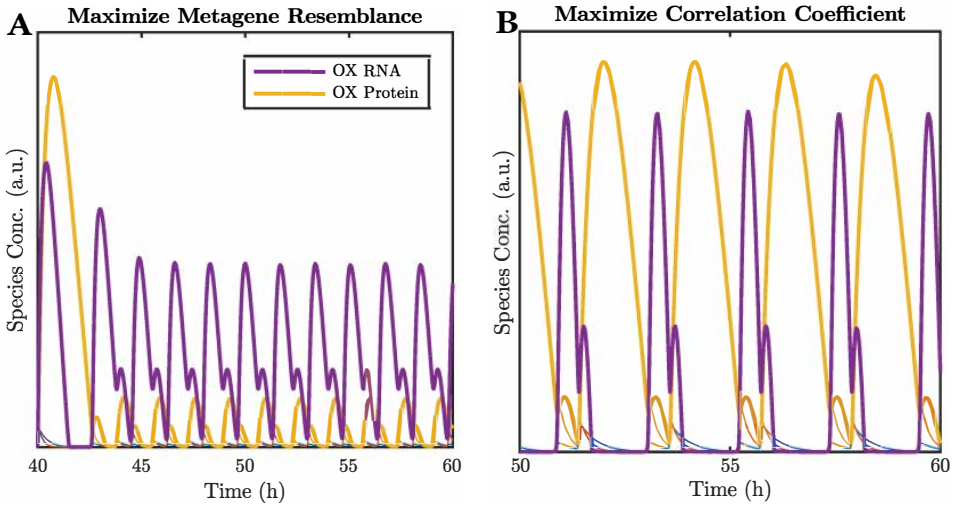


Figure 4.3: Different simulations of the two components model. It is not straightforward to decide which of them is the most accurate YMC representation. OX mRNA is purple, and OX protein is yellow. (a) This simulation gives for the OX mRNA (purple) the largest resemblance to the RNA_{OX} metagene. (b) This simulation is the result that correlates best with the overall clusters' metagenes.

4.3.5 Computational algorithms for numerical solutions

The goal of this Subsection is to explore the possible oscillatory solutions of a three components system with different feedback loops to individuate the parameter space region that yields solutions that can be used as YMC models. By comparing the different forms and shapes of the positive and negative feedback loops, we learn how they compare to the bioinformatics metagenes retrieved in Section 3.

In Appendix A.6.1 is presented a simplified Matlab SimBiology computational implementation for circuits exploring combinations of negative and positive interactions [765]. For example, an $A \dashv B \dashv C \dashv A$ feedback loop resulted in a fraction (approximately 1 in 10) of the simulations starting with a random vector of constants exhibiting oscillatory solutions which could correspond to YMC simulations, as in Figure 4.4.a.

Loops in the generic form $A \to B \dashv C \dashv A$ were found to be less robust, with only approximately 1 in 2000 simulations producing oscillations, most likely because of the $A \to B$ interaction [766]. Furthermore, as we can see in Fig. 4.4.b, these few solutions could not be associated with the YMC metagenes. A similar result was retrieved when exploring repression loops in the forms $A \dashv B \dashv C \dashv A \dashv C \dashv B \dashv A$ and $A \dashv B \dashv C \dashv B \dashv A$, but none of these produced any oscillations in the first 10^5 simulations.

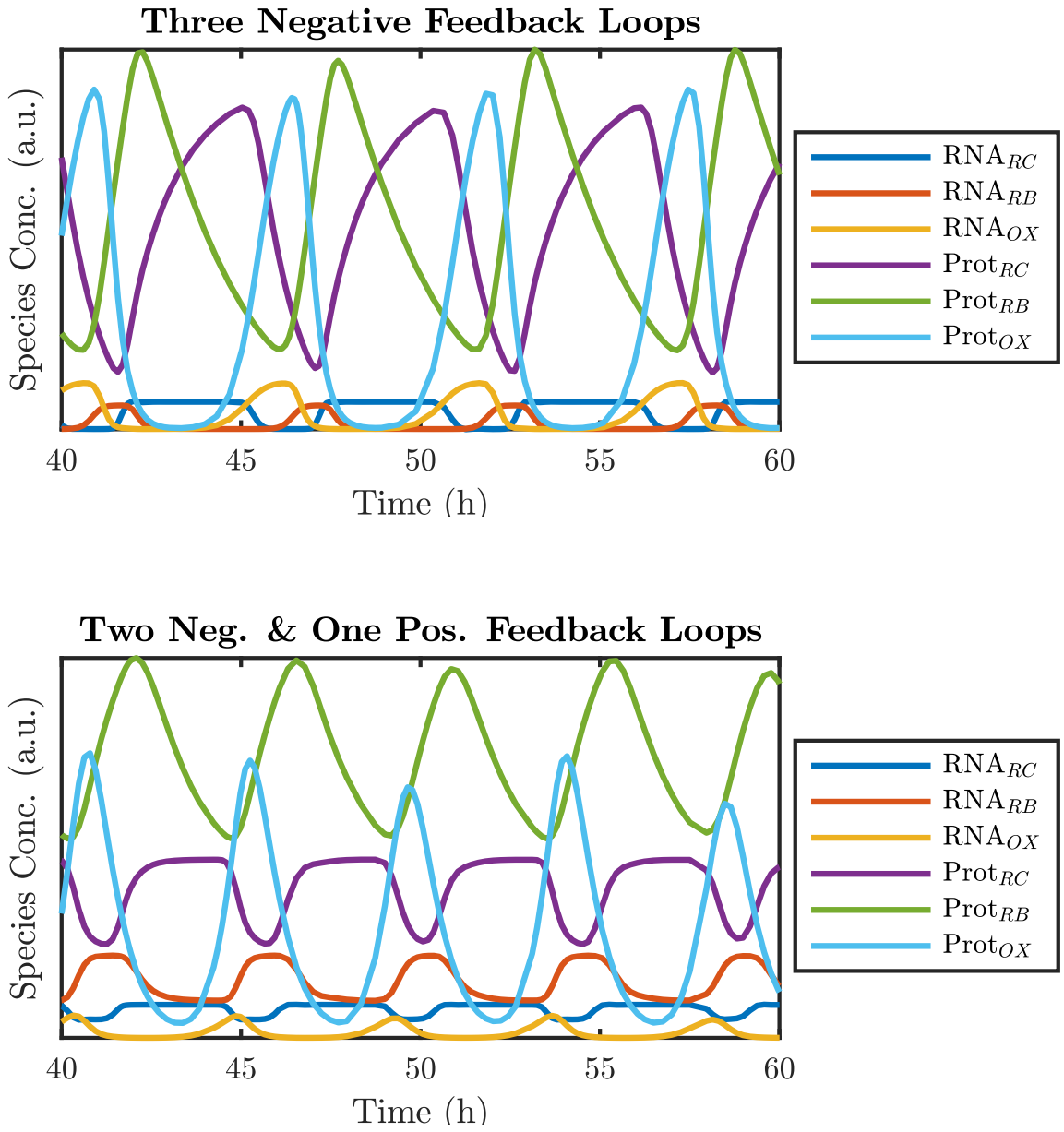


Figure 4.4: (Top) Negative and (Bottom) Positive plus negative regulation mechanisms simulation results. For a wide variety of initial conditions, the loop composed entirely of negative interactions by itself produces oscillations that match more closely the metagene profiles retrieved in Chapter 3.4 (see Fig. 3.11), becoming a better candidate for the mathematical model to be developed in the following Sections.

4.4 RNA and protein equations

This Section is devoted to formulating the ODE system that describes RNA and protein concentration evolution through the YMC. It is pivotal to the work of this Chapter, since it establishes the functional form of these equations: the rest of the Chapter is mainly devoted to finding the numerical values of the constants introduced by these equations.

Figure 4.5 displays a scheme representing the regulatory interactions of the candidate model through the YMC, in which each gene codes for its own protein, and TFs associated with one cluster repress the expression of the genes of the previous phase. In this Section, we will retrieve the mathematical expressions describing each of the regulation relationships shown in the figure.

RNA synthesis and degradation processes can be separated and described by two distinct terms in a differential equation [767–769]. The same is true for protein synthesis and proteolysis [770–772], since both processes are regulated by different mechanisms [773, 774]. RNA degradation is known to follow an exponential decay [775–780], while RNA synthesis does not have a consensus on the analytical form in the literature [775,778,781,782]. In the case of proteins, the situation is the opposite, as we will see: protein synthesis is easily modelled, while multiple models have been proposed for proteolysis in different biological systems [260, 685, 783–785].

4.4.1 RNA synthesis

RNA synthesis is a highly complex process that has been described by multiple functional forms in the literature, like the Michaelis–Menten kinetics [786–789], regulation by negative feedback loops in which enzymes repress RNA synthesis [733, 742, 745, 790, 791], stochastic models [792–796], Bayesian networks [797–799], or postulating the mediation of interacting factors like Hes1 [800, 801], Pif4 [802], or Ap1-3 [803].

The YMC is a synchronized oscillation of yeast populations, therefore it is unlikely to be best explained via stochastic models or Bayesian networks. The Michaelis–Menten kinetics does not usually lead to oscillatory solutions and the mediation of unknown interacting factors is mathematically equivalent to regulation by negative feedback loops. Therefore, RNA synthesis equations take a form similar to [745], assuming the existence of a negative interaction $[Prot_{RB}] \longdashv [RNA_{OX}]$, in which the presence

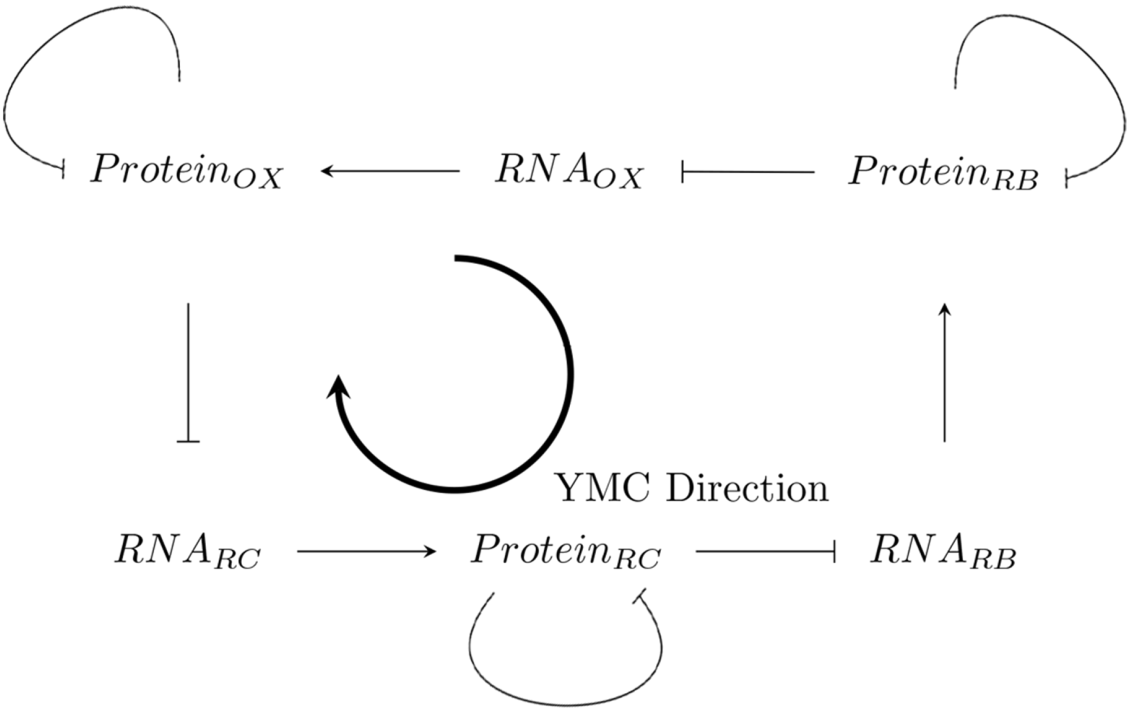


Figure 4.5: The candidate model of positive and negative regulation mechanisms describing the logic of the YMC interactions. Each class of genes codes for their own protein ($\text{RNA}_A \rightarrow \text{Protein}_A$), and transcription factors associated with one cluster repress both the expression of protein at the same phase ($\text{Protein}_B \dashv \text{Protein}_B$) and RNA at the previous phase ($\text{Protein}_B \dashv \text{RNA}_A$). This system has in total nine interactions (six negative and three positive), which is sufficient to produce indefinite oscillations [660].

of protein synthesized by RB genes stops transcription of the OX genes, and similar dynamics follow for the other two phases. Experimental evidence reveals that the functional form of transcription equations is best described by Hill curves [804, 805]; the RNA synthesis rate of OX cluster genes SR_{OX} is then given by

$$SR_{OX} = \frac{P_{OX}^{\max}}{1 + [Prot_{RB}]^p / \mu_{OX}^p}, \tag{4.6}$$

where constant $P_{OX}^{\max}$ is the maximal transcriptional activity achievable during the OX phase (in the absence of RB proteins repression), and it is limited by both the number of genes to be transcribed and Polymerase II abundance, which is approximately constant [806–808]; $[Prot_{RB}]$ is the protein concentration of RB genes; μ_{OX} is a standard Fermi-Dirac like normalizing constant representing the $[Prot_{RB}]$ threshold from which transcriptional activity for OX genes is reduced to a half of its maximal value, $P_{OX}^{\max}/2$, because of the $[Prot_{RB}] \longdashv [RNA_{OX}]$ negative feedback loop; and $p \in \mathbb{N}$ is a non-linearity term representing the rapidity of the ON/OFF transition which is widely present in the literature [658, 733, 734], and it is necessary to achieve oscillatory solutions [660]. $p > 2$ values are explained through the cooperativeness of TFs [809–811], which is sometimes nucleosome-mediated [812].

In summary, the ON/OFF transition represented by Eqn. (4.6) is modelled using three parameters: $P_{OX}^{\max}$, corresponding to the synthesis rate during the ON part of the oscillation; μ_{OX}, which is the halfway point in the transition from ON to OFF; and p, which represents the abruptness or gradualness of the transition.

4.4.2 RNA degradation

mRNA is not stable [813–815]: it degrades spontaneously. mRNA stability and decay rates have been measured in batch yeast [305–308], and there is consensus in the literature that this process follows an exponential decay. Using the results obtained in Section 2.4, mRNA degradation during each of the YMC phases is approximately a constant τ_i whose numerical value was estimated in Table 2.4. There is wide evidence of mRNA degradation following an exponential decay in both batch yeast [307, 404–408] and other species [409–415]. In other words, all mRNA molecules have the same chance of being degraded, independently of when they were synthesized. Hence, for genes belonging to the OX cluster, their RNA degradation dynamics DR_{OX} is expressed by

$$DR_{OX} = \tau_{OX} \cdot [RNA_{OX}]. \tag{4.7}$$

The variation in RNA_{OX} concentration is given by the difference between synthesis SR_{OX} and degradation DR_{OX} rates, so the final equation describing $[RNA_{OX}]$ dynamics is

$$\frac{d[RNA_{OX}]}{dt} = \frac{P_{OX}^{\max}}{1 + [Prot_{RB}]^p / \mu_{OX}^p} - \tau_{OX} \cdot [RNA_{OX}]. \tag{4.8}$$

And analogous equations describe $[RNA_{RB}]'$ and $[RNA_{RC}]'$.

For the discussion of the protein dynamics equations, consider the process we mentioned before [770–772] that protein dynamics can also be described as the difference between the Synthesis rate of the Protein SP and the Degradation rate of the protein DP, where SP is a function of the corresponding RNA abundance, and DP is an enzymatic process depending on both the enzymatic activity and protein concentration.

4.4.3 Protein synthesis

Ribosomes are easily available inside cells [816–818]; therefore, translation is mainly regulated by the abundance of the relevant mRNA molecules [742, 752, 819, 820]. Protein synthesis dynamics during the OX phase can then be approximated as

$$SP_{OX} = \Lambda_{OX} \cdot [RNA_{OX}], \tag{4.9}$$

where constant Λ_{OX} represents the translation rate during OX, since the translation rate is very similar for all genes belonging to the same group within the control circuit [821]. Identical equations follow for RB and RC.

4.4.4 Protein degradation

As opposed to mRNA, protein molecules are stable *per se* [822–825]; they can, however, be decomposed -a process known as proteolysis- by the action of some enzymes [826–829]. Proteolysis can also occur due to non-enzymatic causes, such as extreme pH [830] or high temperature [831], but this should not be the case in the YMC: we assume that non-enzymatic proteolysis is negligible compared to enzymatic proteolysis.

Enzymatic protein degradation is a highly complex process, which is mainly dependent on both the protein and relevant enzyme concentrations [821, 832]. Protein

degradation is well described by the Michaelis-Menten kinetics [42, 833].

Including an enzymatic term into my equations would by far exceed the purposes of this chapter; I therefore assume that the protein degradation rate D can be expressed as

$$DP_{OX} = Z_{OX} \cdot \frac{[Prot_{OX}]}{K_M + [Prot_{OX}]} \tag{4.10}$$

where Z_{OX} is a measure of the enzymatic activity for proteins synthesized from genes within the control circuit and K_M is the Michaelis-Menten constant [834] of the system. Taking the synthesis and degradation rates together yields the equation governing OX protein dynamics:

$$\frac{d[Prot_{OX}]}{dt} = \Lambda_{OX} \cdot [RNA_{OX}] - Z_{OX} \cdot \frac{[Prot_{OX}]}{K_M + [Prot_{OX}]}. \tag{4.11}$$

And identical equations follow for RB and RC protein dynamics. Grouping all of these equations together, we propose that the YMC dynamics is well described by the following set of equations:

The YMC as a transcriptional oscillator: ruling equations

$$\frac{d[RNA_{OX}]}{dt} = \frac{P_{OX}^{\max}}{1 + [Prot_{RB}]^p/\mu_{OX}^p} - \tau_{OX} \cdot [RNA_{OX}] \tag{4.12a}$$

$$\frac{d[Prot_{OX}]}{dt} = \Lambda_{OX} \cdot [RNA_{OX}] - Z_{OX} \cdot \frac{[Prot_{OX}]}{K_M + [Prot_{OX}]} \tag{4.12b}$$

$$\frac{d[RNA_{RB}]}{dt} = \frac{P_{RB}^{\max}}{1 + [Prot_{RC}]^p/\mu_{RB}^p} - \tau_{RB} \cdot [RNA_{RB}] \tag{4.12c}$$

$$\frac{d[Prot_{RB}]}{dt} = \Lambda_{RB} \cdot [RNA_{RB}] - Z_{RB} \cdot \frac{[Prot_{RB}]}{K_M + [Prot_{RB}]} \tag{4.12d}$$

$$\frac{d[RNA_{RC}]}{dt} = \frac{P_{RC}^{\max}}{1 + [Prot_{OX}]^p/\mu_{RC}^p} - \tau_{RC} \cdot [RNA_{RC}] \tag{4.12e}$$

$$\frac{d[Prot_{RC}]}{dt} = \Lambda_{RC} \cdot [RNA_{RC}] - Z_{RC} \cdot \frac{[Prot_{RC}]}{K_M + [Prot_{RC}]} \tag{4.12f}$$

The ruling equations described by (4.12) have 17 constants; since the RNA degradation rates τ_i were calculated in Section 2.4 (see Table 2.4), this system contains 14 degrees of freedom (DoFs). The aim of the following Sections is to explore the

possible values of these 14 unknown constants, using both our biochemical knowledge of the YMC and its cellular dynamics, and computational methods to ensure the solution of these ODEs mirrors the measured YMC features.

4.5 Biochemical estimation of constants

The transcriptional oscillator obtained in the previous Section, and summarized by Equations (4.12), requires 14 constants whose values are, a priori, not known. Even if we knew the order of magnitude of these constants (plausible using biochemical estimates), and assuming each constant can have three values (low, medium, and high), this would still result in $3^{14} \approx 4.8 \cdot 10^6$ possible combinations. Most of these would be non-oscillatory solutions, but still would make it almost impossible to discern what the precise values of the constants are.

The aim of this Section is to use known facts from the yeast biochemistry literature to narrow down parameter space, either by estimating the values of some constants, or by expressing them as functions of other constants.

4.5.1 Synthesis rates

The first group of constants to be estimated will be the six RNA and protein synthesis rates $P_i^{\max}$ and Λ_i respectively. In particular, we start by examining the ratio between the maximum Polymerase II RNA synthesis rate, $P_i^{\max}$ (Units: M$\cdot$ s^{-1}) and the maximum protein synthesis rate $\Lambda_i \cdot [RNA_i]_{\max}$ (same units) at each of the three YMC phases, as a first step for calculating their numerical values.

RNA metagene data have been obtained from RNA-seq (see Appendix A.1.1), and this technique cannot reliably provide absolute [mRNAs] values unless the yeast cells size is accurately measured by other means [572, 835]. Therefore, it can be assumed that the maximum RNA concentration is roughly the same for all three YMC phases.

Transcription rate is limited by RNA Polymerase II (Pol II) abundance in the cell [836, 837]. In yeast batch culture, Pol II has a maximum transcription speed ranging from 40 to 60 nucleotides per second [838–840]. This figure however could however be different for the YMC, since as discovered in Section 3, ribosome-related processes are compartmentalized during the OX phase of the YMC, and therefore this speed could potentially be different at the different YMC phases.

The translation rate in batch yeast ranges from 10 to 30 nucleotides (3 to 10 amino acids) per ribosome per second [841–843]. Assuming the upper range for both transcription and translation, I can make the approximation

$$\frac{P_i^{\max}}{\Lambda_i} \approx \frac{60}{30} = 2. \tag{4.13}$$

The next step in this estimation is to find the ratios $P_i^{\max}/P_j^{\max}$ between the maximal transcription rates achieved at two different YMC phases i and j. It is reasonable to assume that the yeast transcriptional efficiency remains unchanged during the YMC, as hinted by the literature [136]. Cycling genes are selectively expressed during the YMC [127], *i.e.* they are expressed (switched ON) during one of the YMC phases and not expressed (switched OFF) during the other two phases. There are 1546 genes which are expressed during the OX YMC phase, 951 during RB, and 1448 during RC [127]. It is also important to note that during the YMC, far less than 20% of the yeast cells are budding [359]; therefore, the total amount of Pol II inside an individual yeast cell can be taken to be constant during the YMC, and I can make the approximation:

$$\frac{P_{OX}^{\max}}{P_{RB}^{\max}} = \frac{\text{Number of RB genes}}{\text{Number of OX genes}} = \frac{951}{1546} = 0.62, \tag{4.14}$$

since the more genes are being actively transcribed at a given phase, the fewer Pol II per complex of genes there are. A similar reasoning could have been done with the Λ_i values instead, but since the $P_i^{\max}$ and Λ_i values are in a constant ratio -as per Eqn. (4.13)-, the result would have been the same.

The final step in this Subsection is to estimate $P_{OX}^{\max}$, since Eqns. (4.13) and (4.14) will then reveal the other five constants. A normal haploid yeast cell contains on average from 473 to 1 mRNA copies per cell and per gene [407, Table 1]. OX genes are in general highly expressed [127], so I can make the within order of magnitude approximation of 100 mRNA copies/cell. The average haploid cell has a diameter $d = 4\mu$ m [844], which makes a cellular volume $V_{cell} \approx 34\mu\text{m}^3$. Then, a within order of magnitude estimate for $[RNA_{OX}]_{\max}$ scale range is given by

$$[RNA_{OX}]_{\max} = \frac{\text{Copies per cell}}{N_A \cdot V_{cell}} = 5\mu\text{M}, \tag{4.15}$$

where $N_A = 6.022 \cdot 10^{23}\text{mol}^{-1}$ is the Avogadro constant [845, p.134].

I can now make the approximation

$$\left.\frac{d[RNA_{OX}]}{dt}\right|_{\max} \approx P_{OX}^{\max}, \tag{4.16}$$

since in the fastest-transcription part of the YMC mRNA degradation is small. The OX phase is 30 minutes long, and during this time $[RNA_{OX}]$ passes from ≈ 0 to $\approx 5\mu$M, so I can write

$$P_{OX}^{\max} \approx \left.\frac{d[RNA_{OX}]}{dt}\right|_{\max} \approx \frac{5\mu\text{M}}{30\text{ min}} \times 2 = 20\mu\text{M} \cdot \text{h}^{-1}, \tag{4.17}$$

where I have used the fact that the ON/OFF transition for OX genes is sharp (as discussed in Section 4.6), and an *ad hoc* $\times 2$ factor has been included to correct for both mRNA degradation and the fact that from the expression function valley to its peak the average derivative of the function is somewhat smaller than the maximum derivative value. Similar values (within the same order of magnitude) have been obtained by other studies [846–848].

4.5.2 mRNA degradation rates

The average mRNA half-lives for genes belonging to the OX, RB, and RC clusters are 12.3, 16.5, and 14.6 minutes respectively, as discussed in Section 2.4.2 (see Table 2.4). I can use these values to estimate the τ_{OX}, τ_{RB}, and τ_{RC} values appearing in Eqns. (4.12), since

$$\tau_{OX} = \frac{\ln 2}{\text{HL}_{OX}} = 3.4\text{h}^{-1}, \tag{4.18}$$

where ln is the natural logarithm and HL_{OX} is the average half-life of genes in the OX cluster, as per Table 2.4. Analogous equations follow for $\tau_{RB} = 2.5\text{h}^{-1}$ and $\tau_{RC} = 2.8\text{h}^{-1}$.

4.6 Numerical simulations to obtain the remaining constants

In the previous Section, we reduced the DoFs in our system from 14 to 8. The goals of this Section are first, to individuate the region on parameter space capable of producing oscillatory solutions using computational methods and then, to use knowledge

about 8 features of the YMC, like length of its phases or metagene profiles, to obtain a unique solution for the YMC as a transcriptional oscillator represented by Eqns. (4.12).

As we saw in previous Sections, very different system features, or values of the constants in Eqns. (4.12), result in a plethora of solutions to the ODE system. Starting with a study of the different parametric configurations in Eqns. (4.12) that result in expression levels which separate cycling into three phases, we will then discuss the different values of the constants in such equations and how their different values produce (or not) different oscillations.

For example, we saw in Fig. 4.4.a that each of the three mRNAs phases has a switch ON/OFF-like behaviour, in the sense that gene expression levels are readily identified with two states, one in which transcription takes place (ON) and another one on which it does not (OFF), the switch from one to the other being moderately fast. Similarly, each of the protein concentrations peaks at a different time point, with the peaks being wide enough to divide the cycle into three parts, each encompassing the time at which each kind of protein is more abundant; these dynamics approximately mirror, with the expected time delay [849–851], the phases at which each of the three clusters of genes is maximally expressed (ON).

In this Section, the different possibilities for the model parameters will be explored in order to obtain oscillatory solutions where RNAs and proteins have matching partitions of phase space by using heuristic algorithms [852–854]. The DoF reduction is achieved in two steps: first, by adjusting the mRNA ON times to the length of each of the YMC phases (Subsection 4.6.1) and then, by refining this reduced parameter space with $[RNA_i]$ solutions that mirror the metagenes obtained in Chapter 3 (Subsection 4.6.2). Afterwards, a brief discussion on the possible values of the non-linearity term p and Michaelis-Menten constant K_M is presented using both biochemical and computational information (Subsection 4.6.3) to finally obtain the numerical values of the constants involved in this transcriptional oscillator (Subsection 4.7.2).

4.6.1 Matching phase lengths with their experimental values

The YMC has three phases, OX, RB, and RC, which in the experimental set up are maintained for approximately 25 min (OX), 87 min (RB), and 107 min (RC), making

in total a $T = 218$ min oscillation. Consider the ratios between the duration of these phases $R_1 = t(RB)/t(OX) = 3.5$, where $t(RB) = 87$ min indicates the duration of the RB phase, and $R_2 = t(RC)/t(OX) = 4.3$.

Let $\boldsymbol{k} = \{p, \mu_{OX}, \mu_{RB}, \ldots, Z_{RC}, K_M\}$ be a vector whose components k_i are the constants used in Equations (4.12) whose numerical values we do not know yet. We define the configuration space of the oscillatory solutions $\mathcal{R} = \{\boldsymbol{k}_{osc}\}$, where $\boldsymbol{k}_{osc}$ are all the possible combinations of constants such that, when such parameters are used, the system of equations (4.12) results in oscillatory solutions.

We are interested in finding vectors $\boldsymbol{k} \in \mathcal{R}$ such that these oscillatory solutions result in ratios between phases as close as possible to $R_1 = 3.5$ and $R_2 = 4.3$, as stated above. The idea is to use simulated annealing [763,764,855] followed by mono-dimensional local maxima optimization [856–858]. The function ϕ to be maximized by a simulated annealing algorithm is defined as

$$\begin{aligned} \phi\colon \mathcal{R} &\to \mathbb{R}^+ \\ \boldsymbol{k} &\mapsto |R_1^{sim} - R_1| + |R_2^{sim} - R_2| \end{aligned} \tag{4.19}$$

where $R_1 = 3.5$ and $R_2 = 4.3$ are the ratios defined above, and the R_i^{sim} are the corresponding ratios obtained from the numerical solution of Eqns. (4.12) that uses the configuration vector $\boldsymbol{k}$. It is clear that

$$\lim_{R_{1,2}^{sim} \to R_{1,2}} \phi(\boldsymbol{k}) = 0,$$

so we set the simulated annealing simulation to halt when it results in a $\phi < 0.001$ value. The computational details of this simulation are explained in Appendix A.6.2.

The next step is mono-dimensional local maxima optimization, a method widely used after simulated annealing to retrieve relative minima of the function ϕ being optimized [859, 860]. A local maximum along a single (mono) dimension $\boldsymbol{k}_0 \in \mathcal{R}$ is defined to be a point such that

$$\frac{\partial \phi(\boldsymbol{k}_0)}{\partial k_i} < 0 \;\; \forall k_i,$$

where $\boldsymbol{k}_0 = \{k_i\}$ is the vector of constants as defined before. This is not to be confused

with a simple local maximum, since it could be the case that ϕ increases along a derivative direction different to one of the $\boldsymbol{k}_0$ components. The computational details of this simulation are also explained in Appendix A.6.2. Figure 4.6 represents the computational result obtained from running this process and retrieving a parameters vector with the adequate components.

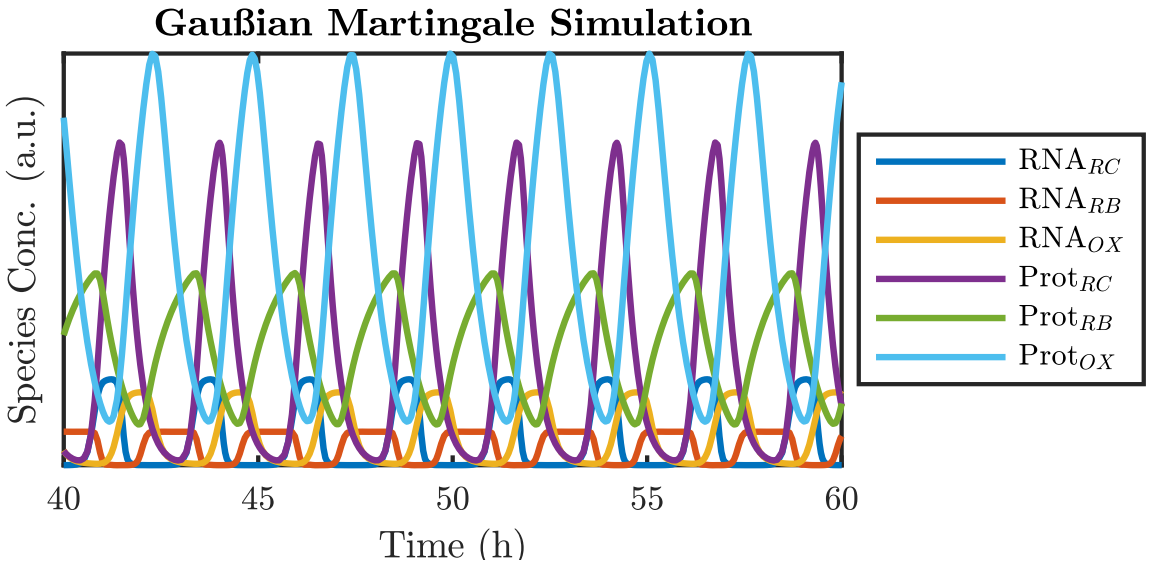


Figure 4.6: Gaußian Martingale walk algorithm simulation results obtained from the algorithm used to solve the mathematical problem presented in Subsection 4.6.1 and fully discussed in Appendix A.6.2. x-axis: time; y-axis: species concentration in arbitrary units. The oscillations in RNA concentration reveal a sharp ON/OFF transition dynamics, which is less abrupt for protein concentration, a result consistent with the empirical evidence of the YMC.

4.6.2 Metagenes shape adjustment

The goal of this Subsection is to construct a function similar to ϕ, as was defined in Eqn. (4.19), able to retrieve $\boldsymbol{k}$ values resulting in Eqns. (4.12) ODE solutions resembling as much as possible the metagenes obtained in Chapter 3.

A good metric to measure the likeliness of two curves is provided by the Pearson correlation coefficient between the vectors formed by the values which these two curves take at discrete times [861, 862]. In this case, the discrete times were the metagenes values taken at integer minutes. The function ϕ' to be maximized in this

Subsection is the sum of the Pearson correlation coefficients for each of the three ODEs solutions with their corresponding metagene from the clustering algorithm explained in Chapter 3:

$$\phi' = r_{OX} + r_{RB} + r_{RC}, \tag{4.20}$$

where r_i is the Pearson correlation coefficient between the metagene associated with cluster i (OX, RB, or RC) and the simulated curve. The computational details of this simulation are entirely analogous to previous Subsection.

4.6.3 Other constants

In this Subsection we discuss the compatibility of the results obtained by simulated annealing in the previous Subsections of this Section with the biochemical knowledge from the literature to investigate the values of the constants that were not assessed in Section 4.5. Such constants are the non-linearity factor p, protein degradation terms Z_i, and the Michaelis-Menten constant K_M.

Non-linearity term p

The dimensionless p term is associated with oscillatory double-negative feedback loops, and $p > 3$ values are in general required to obtain oscillations [660], as discussed in Section 4.4. There is a wide range of values this parameter could take, since $p > 6$ values make the RNA synthesis rate become an almost instantaneous ON/OFF switch, and different p values lead to almost identical solutions. Furthermore, this binary system can always achieve metabolic oscillations, since the ON/OFF switch configuration is unstable [863, 864], and can only hold for at most one of the three YMC phases.

As we discussed in Section 4.4, where the introduction of the term was justified in terms of the relevant literature, $p > 2$ values are explained through the cooperativeness of TFs [809–811]. The cooperativeness of TFs seems to suggest values between 3 and 4 are the most appropriate [865], since $p > 4$ values would be explained only via models collecting a simultaneous regulation of more than 4 TFs [866], which we have not encountered in the yeast literature. Therefore, we set the lowest value able to produce oscillations, $p = 3$.

Protein degradation Z_{OX}

As discussed in Subsection 4.4.4, protein degradation is well described by the Michaelis-Menten kinetics [42, 833]. Protein degradation rates in *E. coli* follow such kinetics, exhibiting Z_i values in the range of $2 - 8\mu\text{M} \cdot \text{min}^{-1}$ [867, Fig. 4B] that can lead to oscillatory solutions. Assuming that similar values hold for yeast, we set $Z_i \approx 5\mu\text{M} \cdot \text{min}^{-1}$; the precise values are tuned in a 2-fold range to allow for YMC-matching oscillations. Such values are consistent with other protein degradation studies in the yeast literature [868–870].

Michaelis-Menten constant K_M

Some enzymatic reactions of protein degradation in *E. coli* models similar to ours retrieved values of the Michaelis-Menten constant K_M in the range of $500 - 1000\text{nM}$ [871–874]. The K_M value has not been measured in the literature for the YMC, and even if protein degradation was known to be a first-order process, the right order of magnitude for K_M would be hard to ascertain. I then set $K_M = 750\text{nM}$, being aware of the wide error range present in this estimation.

4.7 Parameter elasticity and range

The purpose of this Section is to examine the transcriptional oscillator that has been built in this Chapter to assess two of its crucial features: parameter elasticity, or how changes in the values of the constants in Equation (4.12) translate into the solution of the systems, and parameter range, or which intervals of the different constants obtained in the previous Sections allow for oscillatory solutions of the system. Analysis to retrieve limit cycles for parameter values [660, 868, 875–877], as well as discussions on the different behaviours of the system when varying parameter values [685,785,878], are standard practice in the modelling literature to assess a models' validity and limitations, and are a standard measure of its robustness [165, 879, 880].

The first part of this Section explores how the system magnitudes, like period or amplitude of the oscillations, are affected by changes in one parameter, focusing particularly on whether these new configurations lead to oscillatory solutions or not (Subsection 4.7.1). Afterwards, we summarize the numerical values of the constants involved in this transcriptional oscillator, including the ranges on which individual parameters lead to oscillatory solutions of the system (Subsection 4.7.2).

4.7.1 Parameter elasticity

The most frequent method for testing the robustness of a multi-parametric system like Eqns. (4.12) consists of calculating the tolerance range for each of the constants in the ODE system before oscillations are lost [162, 660, 726, 881].

Consider the point $\boldsymbol{k}^{opt} \in \mathcal{R}$ in the possible configurations vector space such that it optimizes the algorithm discussed in Section 4.6, by being the solution to the system of ODEs (4.12) best matching both the experimental data and the biochemical insights discussed before. For each component of the vector k_i^{opt}, the new solution of (4.12) is calculated with k_i^{opt} values ranging up to a 10-fold increase and decrease from the initial value, and the difference in mRNA concentration levels between the original and the new state is used to quantify such a parameter change.

In the case of RNA and protein production and degradation rates, modifications of the initial parameters by up to a 2-fold change were in general able to cycle, and in a number of cases 10-fold changes were also able to maintain the cycles, as shown in Figure 4.7. The full oscillatory range of each of the parameters imputed in Eqns. (4.12) is listed in Table 4.1. This information is relevant to assess the model validity since the maximum fold-change for each parameter provides us with a reliable approximation of the system robustness [165]. It is observed that the parameters' ranges are well above the $\pm 20\%$ benchmark set by the literature [660].

The next logical step is to learn how the system period T and amplitude A evolve under perturbations $\delta \boldsymbol{k}_{i,j}^{opt}$ in two parameters, k_i and k_j, at a time. It would be expected that some of these perturbations would change the form of the modelled YMC oscillations in the same way as some of the yeast mutants do. In the future, it would be interesting to compare the simulation results with YMC oscillation data for a number of mutants [156], since different gene deletions affect the YMC by altering both its period -usually by altering the RC phase length- and the ratios R_1, R_2 between phase lengths that were discussed in Subsection 4.6.1 [156].

Results

Figure 4.7 shows the system dependence (measured by changes in both period and amplitude) when changes in two parameters are introduced at the same time, scanning for 10-fold up and down perturbations (rest of the data not shown). In general, both period and amplitude are monotonous functions of the pair of parameters for most of the range (as shown for K_M, τ_{OX}), with deviations from this trend usually in-

dicating that oscillations are about to be lost. Behaviour is completely unpredictable for the RC/OX ratio (data not shown).

Some simulation models in the literature have taken similar approaches to model, in fission yeast, the action of relevant mutants to explain both DNA replication control [161] and cell cycle quantized times [162], providing explanations for the role of the different constants in the model in terms of the effects caused by infinitesimal perturbations.

Figure 4.8 provides an example of the computational method used to determine the maximum individual parameters' elasticity able to produce oscillatory solutions. Further refinement of the parameter swap displayed in Fig. 4.8 gives an OX RNA degradation threshold value $\tau_{OX}^{Lim} = 12.9$ after which oscillations are lost. Since τ_{OX} value in the model is $\tau_{OX}^{0} = 3.4$, the upper bound elasticity is $\epsilon_{up} = 12.9/3.4 = 3.8$. Similar calculations were performed to individuate the lower bound elasticity, and for all of the parameters in the system, the results being detailed in Table 4.1.

4.7.2 Numerical values of the constants in the transcriptional oscillator

Final Values The numerical results for the transcriptional oscillator that have been obtained from Section 4.4 until here are summarized in Table 4.1 below:

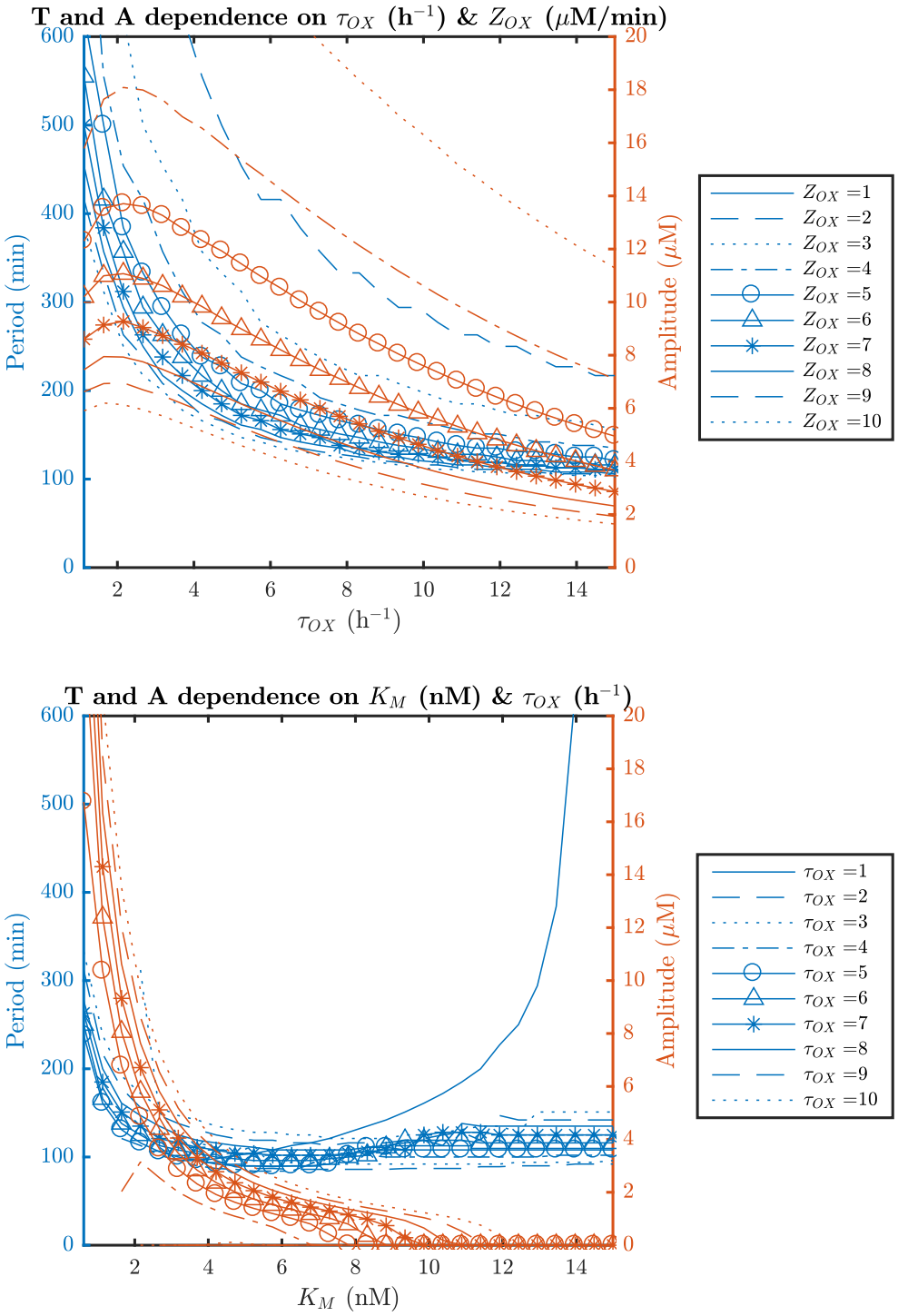


Figure 4.7: Transcriptional oscillator changes in period (T) and amplitude (A) of the oscillations when two parameters of the system are changed simultaneously, as discussed in the main text. Period and amplitude of the oscillations is obtained when the system of ODEs defined by Eqns. (4.12) is evaluated using different (Top) τ_{OX} & Z_{OX} and (Bottom) K_M & τ_{OX} pairs of parameters' values. Each of the lines in the figures can be reduced to a 1-D variational change. Amplitude is measured as the maximum oscillation observed in the most oscillatory species in the simulation. When period or amplitude start to diverge, it is an indication that oscillations will soon disappear.

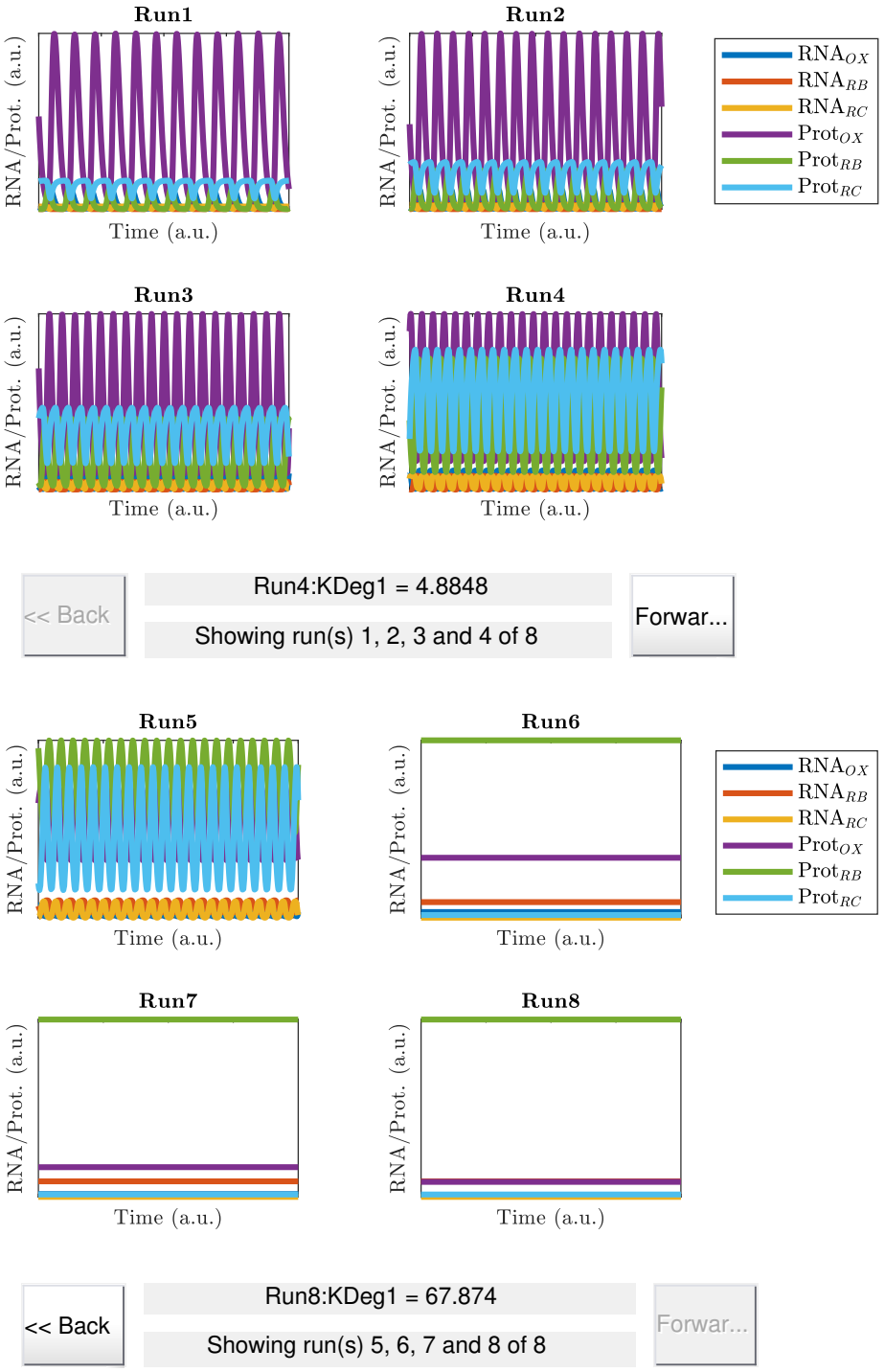


Figure 4.8: Example of a SimBiology parameter swap robustness simulation result for the transcriptional oscillator described by the system of equations in Eqns. (4.12). A swap on the OX RNA degradation rate τ_{OX} values in Eqns. (4.12) reveals the threshold at which oscillations are lost, at some point between the τ_{OX} values in Runs 5 and 6. Further refinement (data not shown) reveals that this limit is reached when $\tau_{OX}^{Lim} = 12.9$.

Table 4.1: Numerical values of the constants in the three-component YMC model developed in this Chapter. Range represents the minimum and maximum fold-change in that individual constant retaining oscillatory solutions of the system.

Constant	**Value**	**Units**	**Range (fold)**
p	3	–	$[1, \infty)$
$P_{OX}^{\max}$	12	$\mu\text{M} \cdot \text{h}^{-1}$	(0.5, 1.2)
$P_{RB}^{\max}$	6.0	$\mu\text{M} \cdot \text{h}^{-1}$	(0.8, 1.7)
$P_{RC}^{\max}$	4.8	$\mu\text{M} \cdot \text{h}^{-1}$	(0.4, 20)
μ_{OX}	1.2	μM	(0.8, 1.6
μ_{RB}	0.4	μM	(0.6, 1.3)
μ_{RC}	0.2	μM	(0.6, 1.7)
τ_{OX}	3.4	h^{-1}	(0.5, 3.8)
τ_{RB}	2.5	h^{-1}	(0.7, 1.5)
τ_{RC}	2.8	h^{-1}	(0.5, 2.3)
K_M	750	nM	(0.5, 2.3)
Λ_{OX}	10	h^{-1}	(0.4, 4.8)
Λ_{RB}	16	h^{-1}	(0.9, 1.3)
Λ_{RC}	11	h^{-1}	(0.4, 2.3)
Z_{OX}	2.5	$\mu\text{M} \cdot \text{min}^{-1}$	(0.7, 2.9)
Z_{RB}	5.1	$\mu\text{M} \cdot \text{min}^{-1}$	(0.8, 1.5)
Z_{RC}	7.4	$\mu\text{M} \cdot \text{min}^{-1}$	(0.1, 3.0)

4.8 A five-component oscillator

The three-phases transcriptional oscillator that has been created in this Chapter does mirror the main three-phases dynamics observed in the YMC [3,127]. However, the clustering results of Chapter 3 open the door to the possibility that there are five or more main clusters of genes with similar possible behaviour in the YMC. My aim for this Section is to learn if a five-components system as the one presented in this Section is more robust (stable) than the three-components model presented in previous Section 4.5 and whether it can fit the data. This Section explores two simple models, the first being governed by the set of equations

$$\frac{d[RNA_n]}{dt} = \frac{1}{1 + [Prot_{n-1}]^p} - K_{Deg} \cdot [RNA_n] \tag{4.21}$$

and

$$\frac{d[Prot_n]}{dt} = K_s \cdot [RNA_n] - \frac{K_{dp} \cdot [Prot_n]}{K_M + [Prot_n]}, \tag{4.22}$$

where it is assumed that the K_{Deg}, K_s, K_{dp}, p, and K_M constants are the same in all five phases, reducing thus the number of DoFs in the system.

The second model has the same Eqn. (4.22) for protein time-dependence, but RNA concentration evolution is described by

$$\frac{d[RNA_n]}{dt} = \frac{1}{1 + [Prot_{n-1}]^p + a \cdot [Prot_{n-2}]^p} - K_{Deg} \cdot [RNA_n], \tag{4.23}$$

where constant a is the second-neighbour interaction force term, widely used in condensed matter physics models [882–884].

Eqn. (4.21) includes repression from the previous term, but since in such model each of the phases is a relatively short duration, it makes more sense from a biochemical point of view to assume that protein levels of phases $n-2$ can also affect the system evolution; therefore, the strength of the $n-2$ *phase* interaction term can be set as $a = 1/2$ in Eqn. (4.23). The result of these simulations is shown in Figure 4.9.

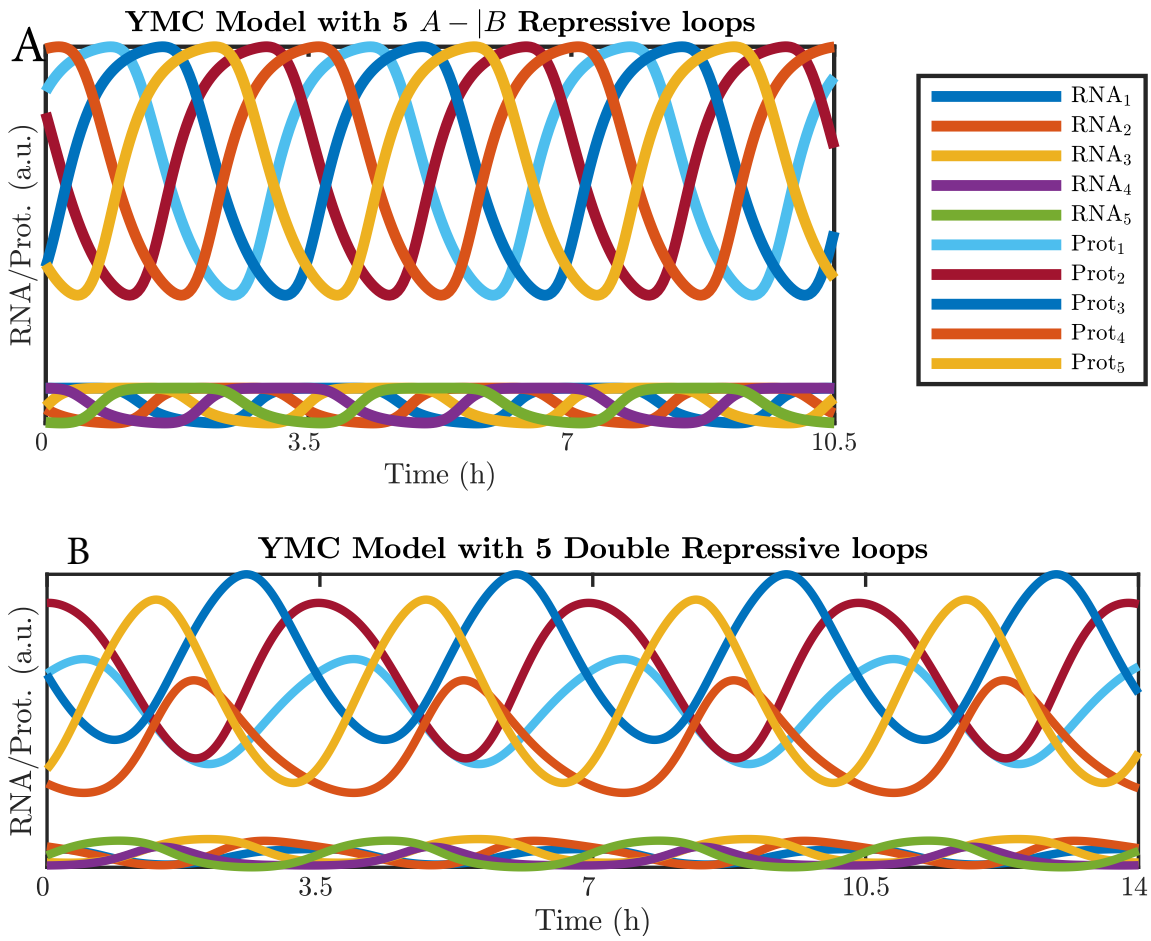


Figure 4.9: A five-component model of an oscillator replicating the YMC dynamics. SimBiology simulation result for the systems described by (a) Eqns. (4.21) & (4.22) and (b) Eqns. (4.23) & (4.22). x-axis: time; y-axis: concentration of the different species in arbitrary units. On both simulations it is observed a sharp ON/OFF transition of RNA concentration which is consistent with the empirical evidence.

4.8.1 Comparison with the model described in Section 4.5

Results revealed that the 5-component model was consistently less robust than the 3-component model to maintain oscillations under 2-fold parameter variations. Performing limit calculations analogous to the 3-component oscillator (see Fig. 4.8), it was found that the amplitudes for the parametric swap intervals were significantly lower in the 5-loops case.

Another question assessed is on which system is it easier to obtain *sharp* transitions, or solutions that minimize the overlap metric introduced and discussed in Section 4.6. A first-approach answer is obtained by visual inspection of Figs. 4.6 and 4.9.a; in the latter (5-interactions model), we can observe that in general, no single phase is switched on merely by itself at any point, revealing a very significant overlap in every parametric configuration. Similar results hold for the 5-interactions double repressive loop simulation, as illustrated in Figure 4.9.b. In this case, we can notice how some waves peak twice per cycle, which is a result not consistent with the metagene profiles obtained in Chapter 3.

Both of these results provide enough grounds to defend that a robust oscillatory system like the YMC is most likely not regulated by a complex, not-robust, 5-components mechanism, but by a simpler, 3-components process, which is coherent with the results obtained in Chapter 3.

4.9 Conclusion: The YMC can be reproduced by a robust three-phase transcriptional oscillator

The goal of this Chapter was to build a transcription-controlled oscillator capable of reproducing the YMC dynamics observed in Chapters 2 and 3, which is reproduced in Figure 4.10. In this Figure, we can notice how each of the proteins peaks slightly after the mRNA expression peak and how the three clusters exhibit a relatively fast ON/OFF transition.

Establishing such a transcriptional oscillator implies that there are solid grounds for maintaining that the YMC regulation is transcriptionally controlled [3, 133, 150, 885], and it is so through a three-component oscillator, since it was concluded in Section 4.3 that two components are not enough, oscillators with an even number greater than two of components tend to be unstable [660], and Section 4.8 concluded that

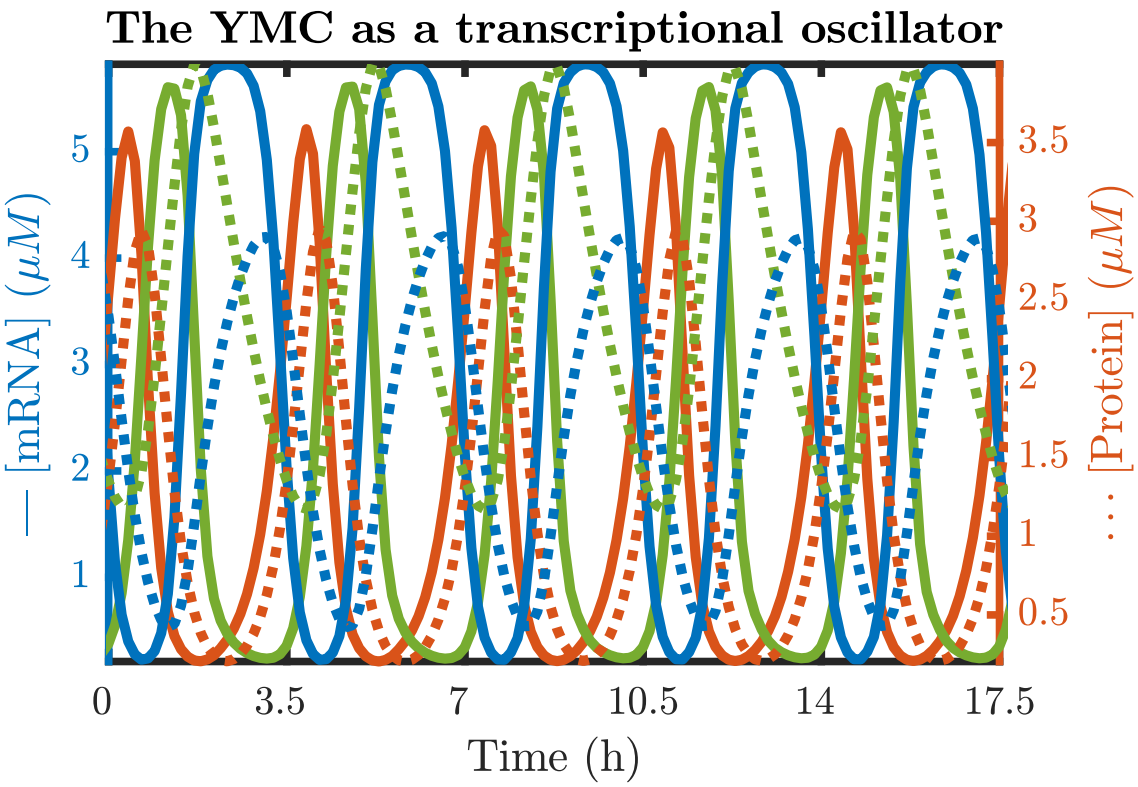


Figure 4.10: Final form of the YMC transcriptional oscillator, obtained when the system of ordinary differential equations described by Eqns. (4.12) is solved using the constants values in Table 4.1. Solid lines (—) represent mRNA concentration; dashed lines ($\cdots$) represent protein concentration. Red color describes the OX phase, green is for RB, and blue is for RC. A sketch of the systems' regulatory machinery is shown in Fig. 4.5.

five components or more make a less robust transcriptional oscillator. It should be noted as well that the YMC is remarkably stable, since once the oscillations are set, the system is kept on an oscillatory state state for an indefinitely long time [3, 108].

The oscillator presented in Fig. 4.10 is also consistent with the protein behaviour discussed in Appendix A.7 (see Fig. A.7), where it is shown that for most genes, their coded protein concentration peaks slightly later than the mRNA concentration.

A single-cell transcriptional oscillator model has been built. But let us keep in mind that the YMC involves coordination between the cells in the chemostat, so it will not be explained by merely describing the cellular internal machinery. In the next Chapter, I will explore different models of intercellular interactions, aiming to explain why yeast populations set and maintain these synchronized oscillations.

Chapter 5

Intercellular Synchronisation Model

The aim of this Chapter is to understand how the yeast cells synchronize and exhibit respiratory bursts, obtaining a complete mathematical model able to explain the most relevant features of the YMC that are observed in the literature, and to predicting the outcome of some validating experiments in the form of precise research questions. This follows the results of Chapter 4, where I modelled the yeast internal regulatory machinery at the single-cell level. Individual yeast cells exhibit regular metabolic oscillations [134, 154, 886], but the YMC is a population-wide system in which the yeast cells synchronize their metabolic state.

This Chapter is devoted to combining a number of the individual oscillators introduced in Chapter 4 to make a full description of the YMC as a synchronised activity of a yeast population, starting by modelling cellular synchronization mechanisms (Section 5.1), the YMC initial perturbation (Section 5.2), and some stochastic gene expression dynamics (Section 5.3). Afterwards, I provide separate models predicting acetate-mediated OX phase induction (Section 5.4) and the co-occurrence of cellular oxygen consumption and some proteins' abundance (Section 5.5). Finally, I ideate an experiment to test these models using a co-culture of two different yeast strains (Section 5.6), to conclude that the yeast internal transcriptional oscillator is stronger than the synchronization with the other cells (Section 5.7).

The YMC is highly sensitive to changes in the inter-cellular media composition; it can be altered for example by spiking acetate or another suitable Carbon

source [116–118], altering the pH [629][1], or changing the feeding media composition [138]. During the YMC, the abundance of NADH [887] and hundreds of other yeast metabolites is observed to oscillate in the intercellular media [112–115, 157, 174, 888], suggesting that these abrupt changes in concentration may be components of an intercellular synchronization system. The overall aim of this Chapter is to learn which process has a larger influence on establishing and maintaining the YMC synchronisation: the yeast internal, transcriptionally regulated, periodic oscillations in gene expression and protein concentration that were modelled in Chapter 4, or the cell-to-cell metabolic phases synchronization mechanism via the intercellular media's chemical composition that will be discussed in this Chapter.

It is important to stress that the YMC takes place in a closed, stable environment. There are not periodic signals from outside of the chemostat. This is probably what makes the YMC so peculiar compared to other biological rhythms models: in the YMC, there is no entrainment from external factors like sunlight [253, 254] or temperature [255] in plants circadian clocks.

5.1 Cells are synchronized during the YMC

The multi-cellular synchronization models built in this Chapter are based on the transcriptional oscillator described by Eqns. (4.12) and the constants of Table 4.1, whose solution is graphically represented in Fig. 4.10. The first goal of this Section is to construct a cellular synchrony model able to explain the chemostat dO_2 oscillations displayed in Figure 5.1.a: the most regular part of the YMC achieved a few cycles after the oscillations first appear in Figure 5.1.b. During the pre-YMC yeast starvation phase, when the culture is being prepared, there is no synchronisation between the cells in the chemostat [129, 140]; each of them is at a different, randomly distributed, phase of their biological processes. The 8 hours or more glucose starvation and subsequent glucose re-feeding [3] breaks this uniform distribution and brings yeast cells to synchrony by making their metabolic processes converge. This convergence is not absolute, since microscope imaging [359] reveals that approximately 30% of the cells inside the chemostat do not follow the YMC, and as we can see in Fig. 5.1.b, the first YMC oscillations generated are of poor quality, suggesting

[1] In the chemostat, acidity is kept constant with a pH= 3.5 buffer. If this buffer was not present, the media would become highly acidic (pH< 2) during the OX phase, but the increased basic buffer usage at the beginning of each OX phase prevents that. An interpretation of this phenomenon, which is matched by increased ethanol secretion, would be the natural selection advantage of having a highly acidic, antiseptic, media on which competitors cannot survive.

that intercellular synchronisation is not immediate, but the results of a convergent synchronization mechanism. Since these first YMC oscillations when $t > 32$ hours are not perfectly regular, it takes a number of hours before the system is regular. In this Chapter I explore the processes behind the transition from a non-synchronized to a synchronized culture.

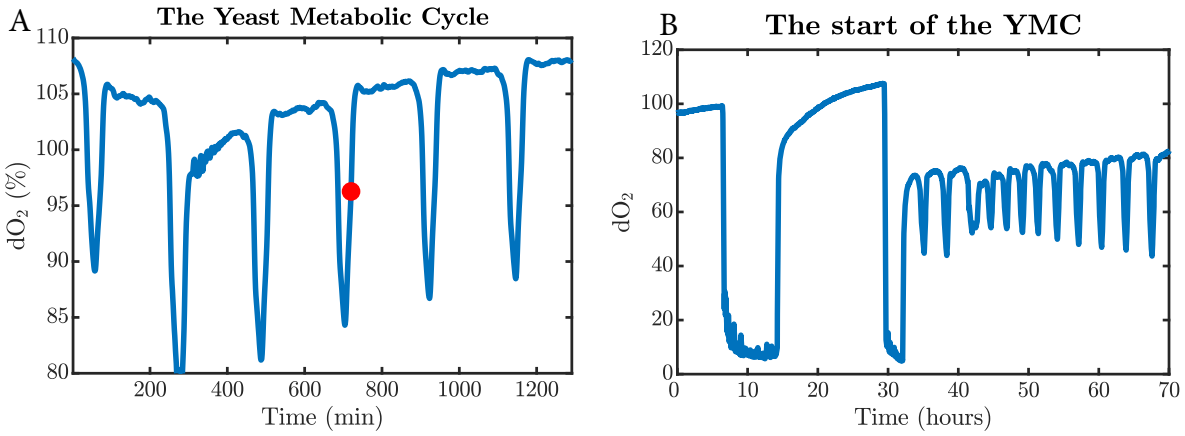


Figure 5.1: (a) Respiratory bursts of the yeast cells are observed in the form of the chemostat regular dO_2 oscillations. x-axis: time; y-axis: dissolved oxygen in the chemostat, measured as % of the saturation level. The red dot marks the point at which dO_2 increases fastest. (b) The start of the YMC. From $t = 14$ h to $t = 28$ h glucose starvation takes place. From $t = 32$ h to $t = 55$ h there are respiratory bursts, but these are not very regular oscillations. In the $t > 55$ region, the system is maintained in an oscillatory state for an arbitrary long time [108, 112].

The inter-cellular synchronisation model I propose assumes that cell-to-media interactions can be modelled via a *mean field* synchronization term that synchronises the metabolic activity of single cells with the intercellular media composition. The biochemistry behind a mean field Hamiltonian term [889, 890] can be justified either in terms of the hundreds of metabolites oscillating during the YMC [112–115, 157, 174, 888] or by changes in the cell wall chemical composition[2], as it is the case in the CDC [891, 892]. The strength of the interaction is taken to be proportional to the phase difference between such a cell and the average of the cells in the chemostat (average media composition). Hence, in a media containing n yeast cells, we have

[2]Because of the permanent stirring, the chemostat is a turbulent system on which yeast cells are constantly clashing into each other. These collisions could be a form of inter-cellular communication capable of producing the cellular synchrony.

that the synchrony term for an individual yeast cell i as a function of time $\xi_i(t)$ can be expressed as

$$\xi_i(t) = k_\alpha \cdot \left(\sum_{j=1}^{n} \frac{\varphi_j(t)}{n} - \varphi_i(t) \right), \tag{5.1}$$

where k_α is the strength of the interaction, or media composition influence on the individual cells metabolism, $\sum_{j=1}^{n} \varphi_j(t)/n$ is the average phase of the n cells in the chemostat (modulo 2π), and $\varphi_i(t)$ is the phase of the individual cell i, as measured by its OX genes RNA concentration $[RNA_{OX}]_i$: the temporal moment t_{OX}^{max} at which $[RNA_{OX}]_i$ reaches its peak in a metabolic oscillation has its phase defined to be $\varphi_i(t_{OX}^{max}) \equiv 0$, and successive time-evolution is given by $\varphi_i(t + t_{OX}^{max}) = 2\pi \cdot t/T_{osc}$, modulo 2π, T_{osc} being the period of the metabolic oscillations obtained from the numerical solution of Eqns. (4.12). The rationale behind selecting OX genes' RNA concentration as a proxy for the cellular metabolic phase stems from the evidence retrieved in Chapter 3 that OX is the sharpest YMC phase, and its start is most easily measured, since it corresponds with the sudden decrease in the chemostat dO_2 concentration. YMC phase division calculations are very frequent in the literature [69, 108, 119, 120, 157, 172–174, 228, 627, 758, 887].

The mean field Hamiltonian described by Eqn. (5.1) modifies the individual cells' OX genes RNA production. OX phase genes seem to be the most actively involved in YMC-specific metabolism coordination [127,426]. The 'kick in metabolism' produced by this Hamiltonian term modifies the Chapter 4 ODE describing OX genes RNA production given by Eqn. (4.12a), which is now

$$\frac{d[RNA_{OX}]_i}{dt} = \frac{P_{OX}^{\max}}{1 + [Prot_{RB}]_i^p / K_{OX}} - \tau_{OX} \cdot [RNA_{OX}]_i + k_\alpha \cdot \left(\sum_{j=1}^{n} \frac{\varphi_j(t)}{n} - \varphi_i(t) \right), \tag{5.2}$$

where the last term on the RHS of the equation incorporates the ξ_i synchronisation term of Eqn. (5.1). The other five RNA and protein concentration equations describing the system, Eqns. (4.12b-f), remain unaltered.

5.1.1 k_α measurement

The strength of the constant k_α in Eqn. (5.1) can be estimated by matching the simulation data to the values observed in Fig. 5.1.b for the synchrony time t_S, *i.e.* the time elapsed from glucose starvation to the beginning of the first cycle. Usually

a $t_S \approx 8-14$h is required for wild type (WT) strains [3,126], but some yeast mutants like *sic*6Δ or *bem*2Δ exhibit much larger t_S values [156].

Experimental results have shown a strong negative correlation $r = -0.86$ between t_S and the cellular density in the chemostat [222], as measured by their optical density (OD_{600}). Such an observation is consistent with the synchrony term represented by Eqn. (5.1), since the larger the cell density, the sharper the fluctuations in the carbon-based metabolites concentrations [112,113], and the stronger the inter-cellular signalling mechanism and its measure of the k_α value.

To estimate k_α, I find the numerical solutions of Eqns. (4.12), with Eqn. (4.12a) replaced by Eqn. (5.2), and consider the convergence time in these simulations as a function of k_α. I then perform a parametric scan over k_α until obtaining a simulation convergence time of $t_S \approx 10$ hours (approximately 3 YMC periods) that yeast cells take to start cycling [3, 127, 222]. This occurs when $\boldsymbol{k_\alpha = 0.1h^{-1}}$. Once k_α has been estimated for WT yeast in standard chemostat conditions, I am a step closer to model different mutants [156, 158] by making the approximation $k_\alpha \propto \frac{1}{t_S}$.

5.1.2 Multiple synchronization points during the YMC?

A priori, there could be several possible functional forms to model the cellular phase convergence term ξ_i in Eqns. (5.1) and (5.2) [893, 894]. I have chosen the mean field theory approach because of its versatility in non-equilibrium systems [895]. Being a first-order convergent process, mean field theory lets us conveniently model the effect of an arbitrary, unknown, coherent force, whose magnitude can be determined by comparison to the experimental data of the dO_2 trace.

I compared the effect of the ξ_i term in Eqn. (5.2) -where it only modifies the RNA_{OX} production dynamics- to the effect of three smaller, $k_\alpha/3$ terms, one at each of the three YMC phases. I believe that it is unlikely that this functional form can represent the YMC, since as we saw in Sections 1.4, and 1.5.1, YMC phase synchrony is chiefly mediated by Acetyl-CoA [117] and therefore it mainly takes place during the OX phase. Translated into mathematical language, this means that the synchronising force must be relatively weak, and for this kind of oscillators, the linear one discussed in Section 5.1 is the most versatile [109, 896]. The conclusion I gather from these considerations is that there exists only one YMC synchronization point, located in the OX phase, and acting as a gauge of the cellular metabolic state [117]. Such an OX checkpoint is consistent with the size-related control checkpoint observed

right before the *Start* CDC phase (the G_1/S transition) [897], albeit it is true that more regulatory mechanisms and check points post *Start* are thought to exist in the CDC [898]. CDC control is achieved through the G_1 cyclin Cln3 [899] and its associated cyclin-dependent kinase, Cdc28 [900], via the Swi6 TF [901]. Finally, the YMC OX phase TFs and cyclin-dependent kinases machinery have been identified with the first third of the CDC G_1 phase [128]. As we have seen, the current state of the science regarding the CDC is in line with the biochemical reality underlying the functional form of Eqn. (5.2): there is chiefly one checkpoint control point during the YMC, and it is located somewhere in the OX phase.

5.2 Initial perturbation

The YMC is preceded by a starvation phase on which the yeast consume all of the glucose available in the media, reducing their cellular mitochondrial activity [902, 903] and oxygen consumption [222] to a minimum. During this time, cells in the chemostat do not exhibit a synchronized metabolic activity,[3] which means that before the beginning of the YMC the different cells phases φ_j follow a continuous distribution

$$\varphi_j(t) \sim \mathcal{U}(-\pi, \pi), \tag{5.3}$$

where $\mathcal{U}(-\pi, \pi)$ is a uniform distribution across the $(-\pi, \pi)$ interval, as sketched in Figure 5.2.a.

The YMC is started at a time t_0 when starvation is interrupted by the reintroduction of glucose feeding (Fig. 5.2.b), which modifies the uniform distribution in phase space given by Eqn. (5.3) and we assume causes the different cells phases φ_j to follow a distribution

$$\varphi_j(t_0) \sim 2\pi \cdot \mathcal{N}(0, \sigma), \tag{5.4}$$

where $\mathcal{N}(0, \sigma)$ is a normal distribution across phase space (modulo 2π) with large σ and $\mu = 0$, since during the starvation phase RNA_{OX} production is at its minimum [117, 157]. The biochemical justification for the transition from Eqn. (5.3) to Eqn. (5.4) is that most cells in the culture will, when starved, adopt a roughly similar metabolic activity, reducing the processes associated with the CDC and a fraction of them entering quiescence [904], so even when the perturbation is removed, cells have already exited the uniform distribution outlined before and, even if they will not show perfect synchrony yet, they will have a non-uniform distribution with large

[3] However, cells are known to exhibit individual unsynchronized oscillations [134, 136, 154, 851].

standard deviation (Fig. 5.2.c) that will get more and more coherent with time, until after ≈ 3 oscillations almost all the cells are in a coherent state (Fig. 5.2.d).

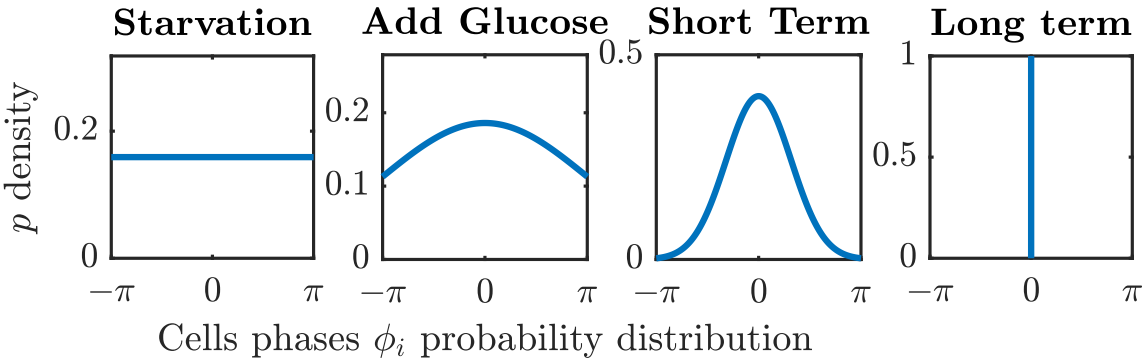


Figure 5.2: Overall system phase distributions at different stages of the YMC. (a) During the yeast starvation phase, cells are not synchronized, so there is not a *privileged* phase onto which all cells can converge, and cellular metabolic phases follow a uniform distribution $\varphi_j(t) \sim \mathcal{U}(-\pi, \pi)$. (b) Glucose addition acts as a Gaußian kick. When the system enters into the YMC, there is a "preferred phase": the $\varphi = 0$ region has a larger number of cells in a semi-synchronised metabolic state. Since this is not a stable configuration, (c) during the first few cycles more and more cells join the *trend*, narrowing the phases distribution standard deviation. (d) In the long term, after a few YMC oscillations, all cells have synchronized their metabolic phase, effectively exhibiting a 'single cell' behaviour; their distribution in the $t \to \infty$ limit is a Green's function $\delta_0(\varphi)$ [905]. The pace at which the transition from the configuration on the Left to the configuration on the Right takes place is given by the strength of the intercellular synchronisation term k_α in Eqn. (5.1). Recall that phases φ_i are measured modulo 2π.

5.3 Noise

Biochemical processes like gene expression [792, 906–908] and signal transduction [909–912] in a single cell have a strong stochastic component; furthermore, the YMC takes place in a turbulent fluid, which means that modelling the different yeast cells interactions requires stochastic equations [913–915]. Mathematical models of stochastic gene expression [760, 916–918] usually incorporate this fact by introducing noise terms of different functional forms into the ruling equations describing the dy-

namics of the system. In particular, in the case of differential equations describing gene expression, stochastic models in the literature usually incorporate at least one Wiener process [919–922] in the form of a summand in the gene expression equations.

In this model, noise associated with gene expression and signal transduction is incorporated into the ODEs describing the YMC by adding a *noise* term to Eqn. (5.2) in the form of a Wiener process, which is the integral of the Gaussian white noise

$$\eta_i(t) = k_\beta \cdot W_i(t), \tag{5.5}$$

where k_β is a measure of the quantity of noise in the system, and $W_i(t)$ is the integral of a Gaußian white noise originated from the distribution $\mathcal{N}(0,1)$. The properties of Wiener process $W_i(t)$ are [923, 924]:

1. $W_i(0) = 0$
2. W_i has independent increments: $\text{cov}[W_i(t+u) - W_i(t), W_i(s)] = 0 \; \forall t, u > 0$
3. W_i has Gaußian increments: $W_i(t+u) - W_i(t) \sim \mathcal{N}(0, u)$
4. W_i is continuous in t

The final form of the equations is then given by

$$\frac{d[RNA_{OX}]_i}{dt} = \frac{P_{OX}^{\max}}{1 + [Prot_{RB}]_i^p / K_{OX}} - \tau_{OX} \cdot [RNA_{OX}]_i + k_\alpha \cdot \left(\sum_{j=1}^{n} \frac{\varphi_j(t)}{n} - \varphi_i(t) \right) + k_\beta \cdot W_i(t). \tag{5.6}$$

The processes described by Eqn. (5.6) follow what statisticians call an ARIMA (AutoRegressive Integrated Moving Average) model [925–927], which is particularly suitable to avoid the oddities derived from the infinitesimal behaviour of the ξ and η terms introduced by Eqns. (5.1) and (5.5) [928, 929]. The numerical methods solution of Eqn. (5.6) takes advantage of the fact that the `ode23` ODE solution algorithm [710, 712] incorporates a discrete number of points at which the ARIMA's ξ and η components can be imputed into the simulation.

Figure 5.3.a shows a solution to the deterministic system of ODEs represented by Eqn. (5.2), i.e. including the ξ_i term of Eqn. (5.1) for a system formed by $n = 1000$ cells. Figure 5.3.b shows the numerical simulation of Eqn. (5.6), obtained when stochastic gene expression is taken into account, and the Wiener process described by Eqn. (5.5) -η_i term- is incorporated. In the future, it would be interesting to

explore a radically new approach in which the system is directly assumed to follow an arbitrary ARIMA [927] or ANOVA [930] process.

Noise propagation Because of the average zero value of Wiener processes, this system can bear high k_β values able to induce strong phase shifts on the individual cells before the oscillations are destroyed. The noise term η_i introduced in Eqn. (5.5) results in chemostat cells exhibiting a higher standard deviation in φ_i -compared to the no noise case-, setting them in a Brownian motion in phase space [931, 932]. But the overall effect on the system's global phase $\Phi(t) = \sum_{j=1}^{n} \frac{\varphi_j(t)}{n}$ is almost negligible, since the overall noise contribution to Φ, $\sum_j \eta_j/n$, is an average of normal distributions, and the $\sum_i \eta_i$ noise is only propagated with $\sqrt{n}$ [933]. In the chemostat $n \approx 10^8$, therefore the overall system phase is almost not affected, even when very large k_β values are introduced. This is the reason for which I do not provide a numerical value for the strength of the environmental noise k_β constant, since this system could include values much higher than noise estimates in the yeast literature [934–936] that would not destroy the YMC oscillations. Note that it is almost impossible to measure the noise of the system: sequencing techniques [574, 937, 938] and proteomics experiments [471, 939, 940] for single yeast cells are extremely noisy; even if single-cell analysis was performed, it could not be discerned which part of the detected noise is caused by the cell's intrinsic stochastic metabolism and which part stems from the use of noisy single-cell sequencing techniques.

Therefore, the model presented in this Section, and summarized by its ruling equations (5.6) can incorporate the stochastic gene expression and protein production observed in single yeast cells [134, 471, 941], while explaining how such is noise diluted for multicellular systems.

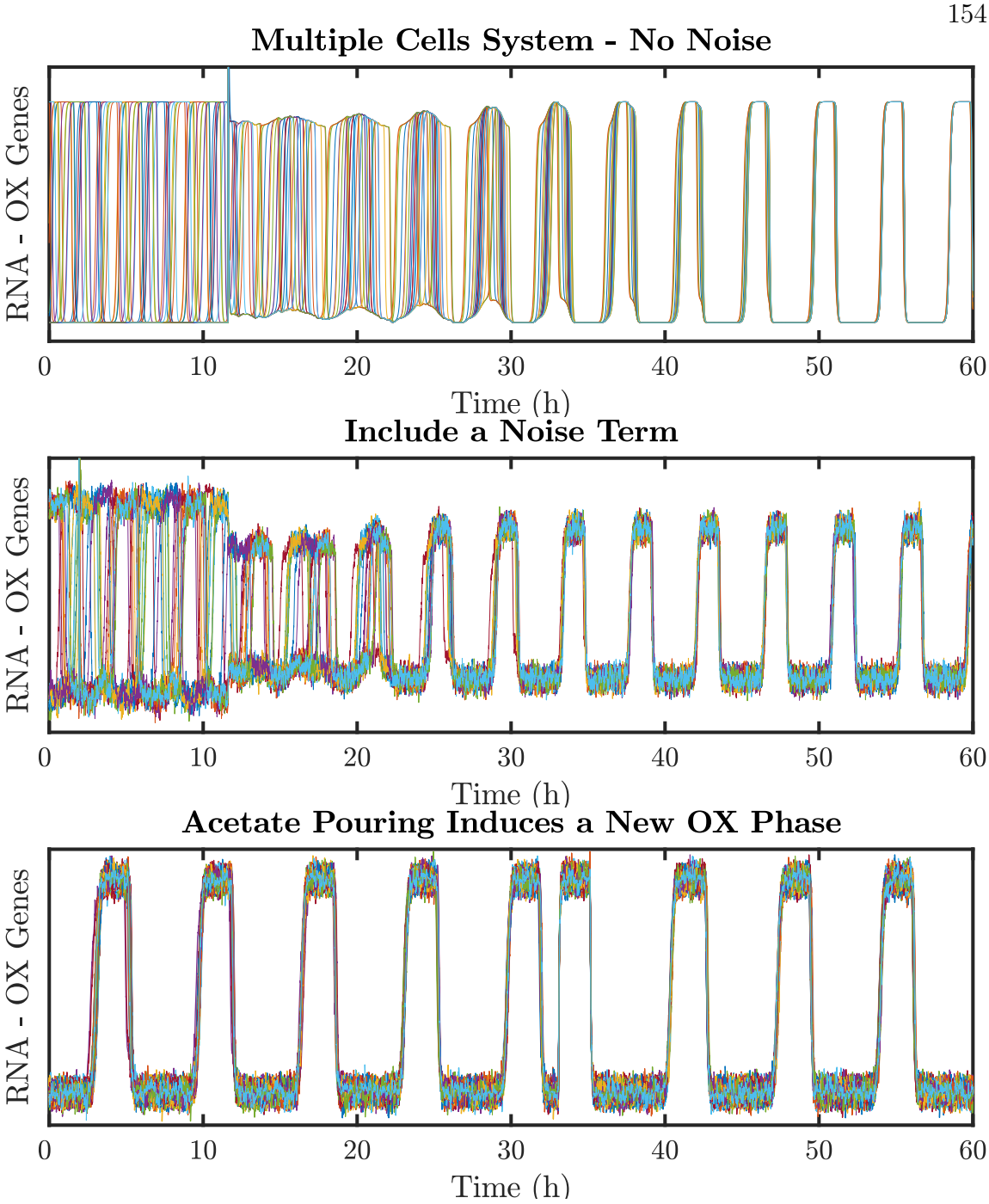


Figure 5.3: Multicellular system simulation results: average $[RNA_{OX}]$ time-evolution under different assumptions. (Top) Simulation of the deterministic system of ODEs given by Eqn. (5.2), where there is a synchronisation term ξ. The end of the glucose starvation takes place at $t = 12$ h (marked with a light blue spike), and the system needs approximately 3-4 oscillations to gain synchrony [3, 127, 222]. (Middle) Same simulation, with the inclusion of the white noise term η of Eqn. (5.5). This is the numerical solution for the system whose dynamics is described by the final multicellular model equations, Eqn. (5.6), incorporating both ξ and η. (Bottom) When acetate is added to the system, it is suddenly driven into the OX phase [118], as explained in Section 5.4. This is equivalent to a sudden, unexpected, overall $[RNA_{OX}]$ increase, and it is reflected at $t = 33$ h, when a sudden rise in $[RNA_{OX}]$ takes place before a full reductive phase -low $[RNA_{OX}]$- has taken place. All three simulations were done with a system of $n = 1000$ cells (for legibility purposes, only 20 of them are plotted).

5.4 Cells are driven into OX via a carbon source

In Sections 5.2 and 5.3, I have studied two mechanisms by which cells can be perturbed from their otherwise periodic $\varphi_i(t)$ rhythm driven by Eqn. (4.12). In this Section I will study a system perturbation capable of producing sudden shifts in the chemostat's overall phase $\Phi(t) = \sum_{j=1}^{n} \frac{\varphi_j(t)}{n}$ value: the addition to the intercellular media of some suitable carbon sources. When acetyl-CoA, acetate, ethanol, acetaldehyde, or lactate[4] are poured into the chemostat, even at low concentrations,[5] cells are suddenly driven into a fresh OX phase, the process being only 11 minutes long [118]. This is also the case for mutants like $msn2\Delta$ and $msn4\Delta$ exhibiting longer RC phases [151, Fig. 2A], suggesting that acetyl-CoA can be a gauge of cellular metabolic state [117] by driving the transcriptional growth program in yeast [116].

I simulate spiking a carbon source into the chemostat by suddenly changing the mean field term Φ in Eqn. (5.6) to $\Phi(t_{Red}) = 0$, when the system is at some point t_{Red} within the YMC reductive phases (either RB or RC). In other words, by setting the intercellular synchronisation mean field term Φ equal to zero we are simulating -see Eqn. (5.1)- that for each individual cell i, all of the other cells in the chemostat have just entered the OX phase. By performing this simulation at different temporal moments of the reductive phases, phase resets to $\Phi = 0$ are observed to occur within very few minutes, approximately 10 minutes for most points in these reductive phases. The result of this simulation is displayed in Figure 5.3.c, where a sudden increase in overall $[RNA_{OX}]$ - *i.e.* an OX phase simulated induction- is observed to occur within very few minutes of the acetate spiking event at $t = 33$ h.

Figure 5.4.a shows that the model built so far makes a novel prediction about OX phase induction: if acetate spiking takes place at some point during the OX phase, then the system needs to exit OX before it can be induced again into OX. Figure 5.4.b shows how this model can reproduce the consecutive induction of two OX phases one immediately after the other, by two successive acetate kicks, observed in the literature [151, Fig. 3A].

[4]Note that all of them are characteristic carbon sources of the yeast metabolism [942, 943], particularly in nutrient limitation environments [944] and in chemostats [945].

[5]As little as 1 mM of $[1,2{-}^{13}C_2]$ acetate is required. It is possible that this effect was discovered by accident, since vaporizing cleaning ethanol over the chemostat is enough to induce the OX phase.

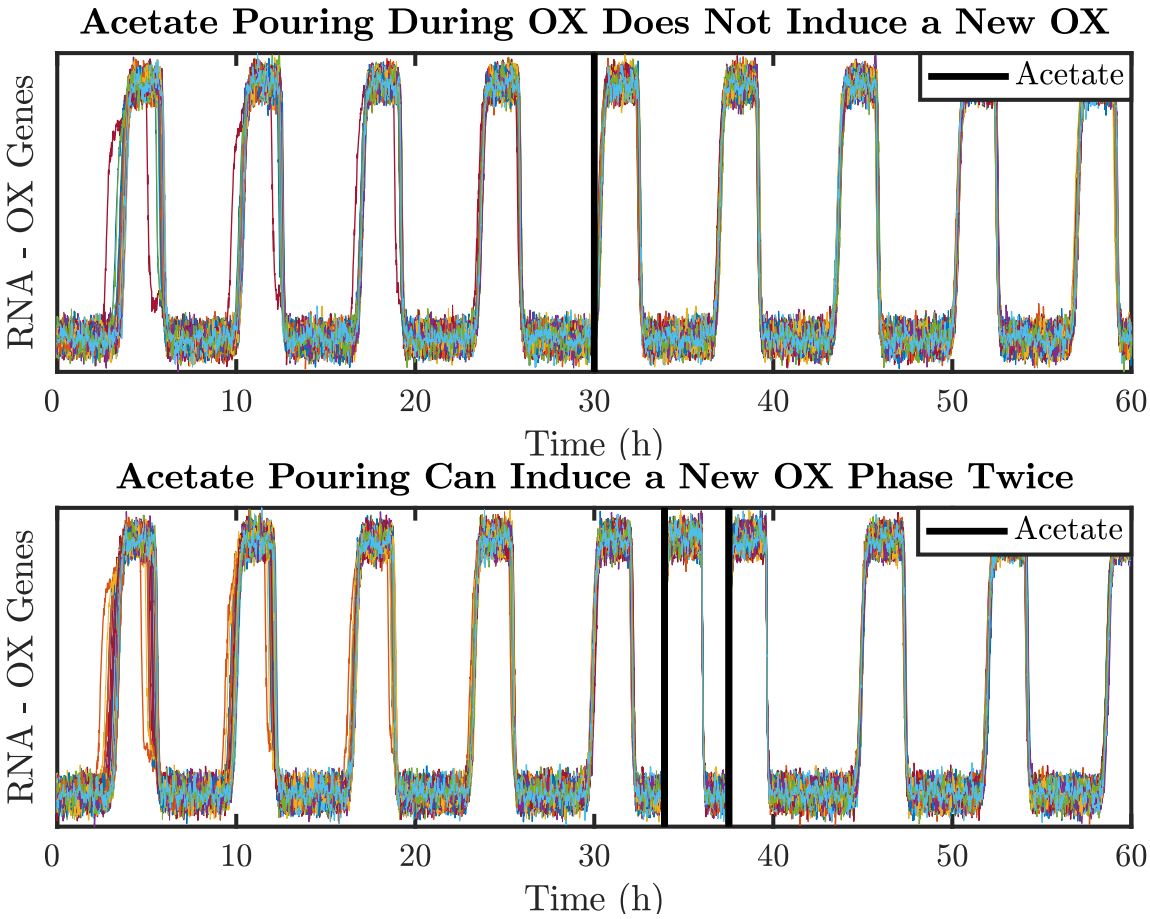


Figure 5.4: (Top) A novel prediction of the model: acetate pouring during OX phase does not induce a new OX phase. (Bottom) A literature result reproduced by the model: two OX phases can be induced one immediately after the other by two successive acetate kicks [151, Fig. 3A]. x-axis: time; y-axis: predicted level (in arbitrary units) of RNA concentration for genes belonging to the OX cluster. Black lines mark the acetate spiking times.

5.4.1 A threshold for acetate spiking

The model simulations (see Fig. 5.3.c) reveal that strong perturbations induced by acetate spiking ($\Phi \equiv 0$ shifts) do induce OX phases, but the literature has not investigated the YMC behaviour when spiking is below the 1mM concentration threshold on which this effect is observed. This Subsection predicts the system simulation outcomes under different acetate spiking concentrations.

So far, the model of the acetate spiking presented in this Section is consistent with the two main features of phase shifting: one, shifts take place because of sudden changes in acetate concentration, absolute concentration levels being less relevant [116][6]; two, phase induction is a matter of minutes [117, 118]. The acetate signal must be strong enough so that it be perceptible; 1 mM of $[1,2{-}^{13}C_2]$ acetate is enough to produce a full OX phase induction [118]. But what will happen when the acetate kick is of lower concentration than the level required for a full induction? There are two possibilities: one, nothing will happen, since there exists an induction threshold, and it has not been reached; two, OX phase is only induced in some cells, this partial induction resulting in an incomplete dip in the dO_2 trace.

Figure 5.5 shows the OX phase induction predictions for different kicks of acetate concentration: low concentration (Fig. 5.5.a) should yield a small, partial, OX phase induction. This phase induction gets stronger the higher the acetate concentration (Figs. 5.5.b-c), eventually reaching the full induction threshold at 1mM concentration. In other words, there is no threshold for phase induction, low concentration spiking provides a partial entry into the OX phase.

What happens if, rather than an instantaneous kick, the increase in acetate concentration is maintained over a long period of time? We know from the literature that yeast cells are very efficient in consuming the media acetate [946, 947], therefore a boost of the CDC-related processes while relegating the YMC to a background could be expected. Figure 5.6 shows the predictions of this model: unless the concentration is very weak [0.1mM] (Fig. 5.6.a), continuous acetate feeding destroys the YMC oscillations (Figs. 5.6.b-c).

The biochemical interpretation of this model prediction is that, by sustaining an acetate concentration that is much higher than what it would be if the system was just fed with the normal media preparation, the yeast feeding rate is effectively

[6] *I.e.* if there is a high media acetate concentration, a fresh spiking still provokes a phase shift.

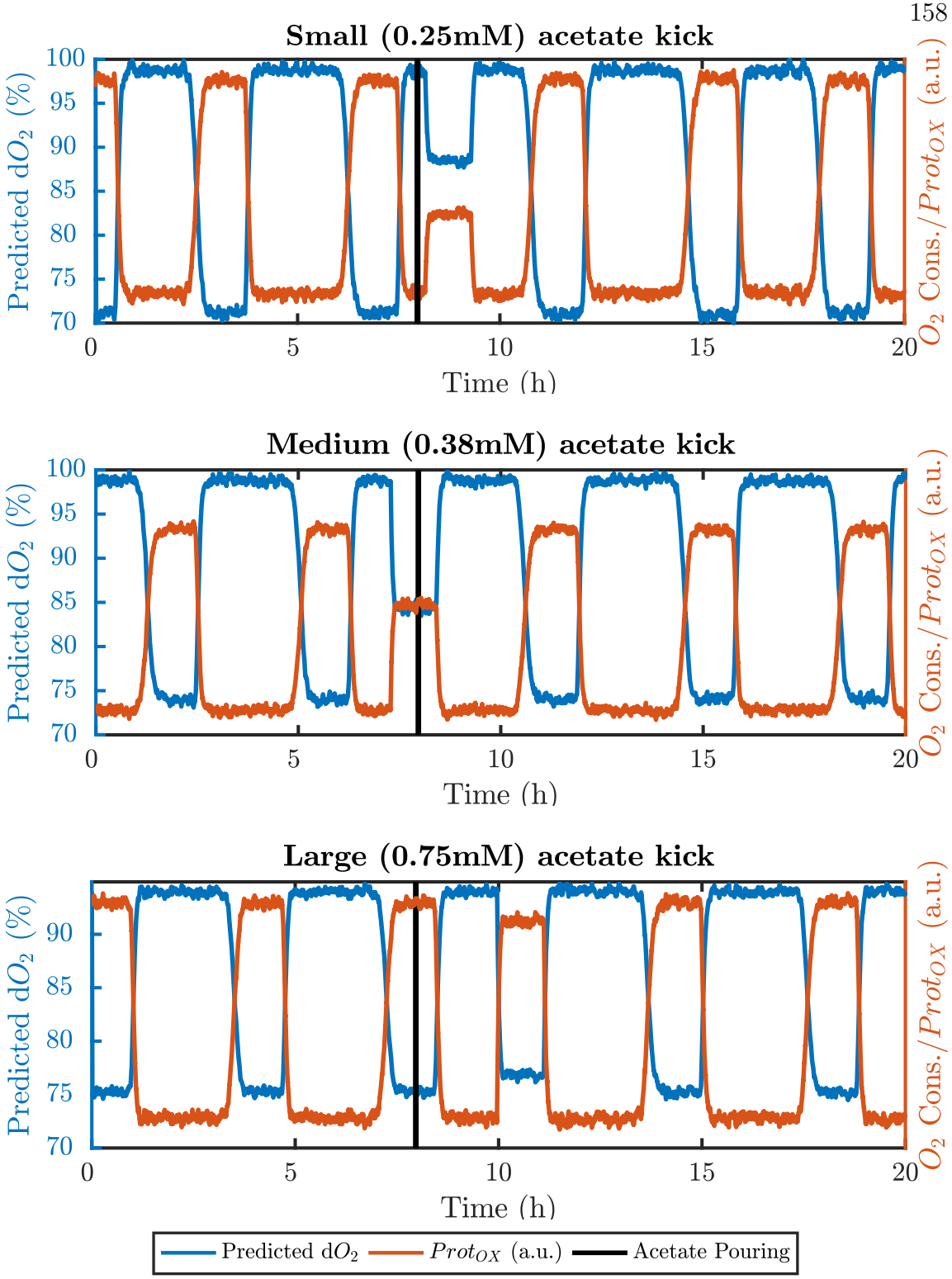


Figure 5.5: Predicted dO_2 trace when different acetate concentrations are poured into the system during the YMC reductive phase. (Top) A small amount of acetate or equivalent carbon source is poured into the chemostat [0.25 mM]. (Middle) a medium [0.38 mM] pulse. (Bottom) A large kick [0.75 mM]. Notice how the dO_2 dip increases in amplitude until it gets approximately constant at the normal YMC amplitude level. Similar results were obtained when acetate was poured at other moments of the Reductive (RB/C) phase, OX being induced within approximately 10 minutes after the signal.

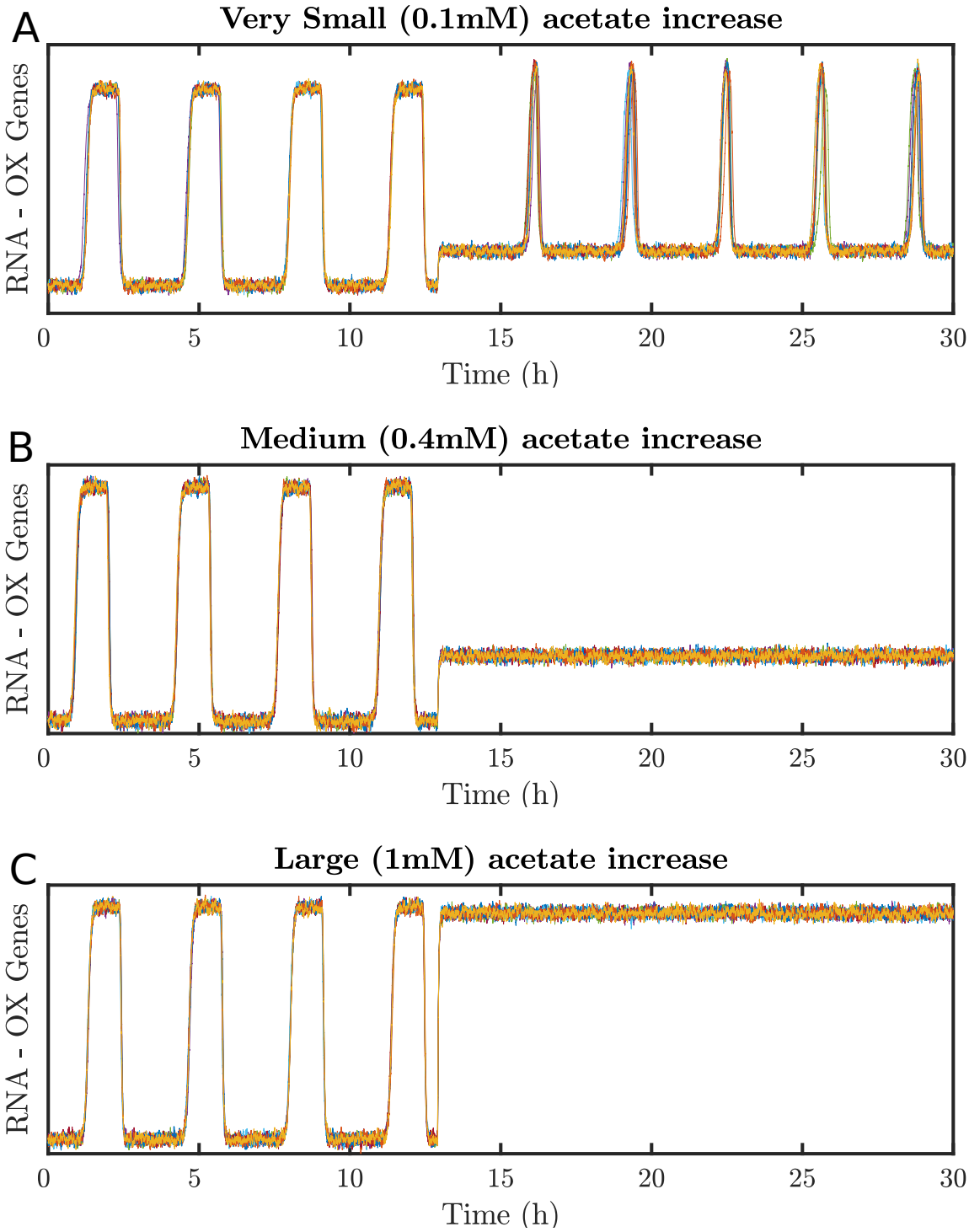


Figure 5.6: Predicted [RNA_{OX}] dynamics when different sustained acetate concentrations are present in the media. Acetate is spiked when $t = 12.5$ h. (a) A very small [0.1 mM] increase. (b) A medium [0.4 mM] sustained increase is enough to prevent cells from cycling, since they change their metabolic pathway from glucose to acetate. (c) A [1 mM] sustained increase, which is enough to induce a permanent OX phase [118].

being increased, and therefore the yeasts' preferred acetyl-CoA metabolic pathway is changed from glycolysis [948] to acetate conversion [118], which is more efficient in terms of oxygen consumption. This means that the YMC is not necessary any more, since yeast have exited the nutrient limitation phase. In the future, it would be interesting to relate this point with glycolysis being one of the fundamental components of quiescence state exit [904, 949, 950].

5.4.2 Cells are not driven into OX phase merely by a decrease in the media O_2 concentration

We saw in Subsection 1.5.1 that some organisms like *E. coli* can sense the environmental dissolved oxygen (dO_2) levels and plan their metabolic activity accordingly. This fact provides good grounds for the hypothesis that chemostat dO_2 levels are not a YMC consequence, but the very reason for which cells synchronise their entrance into the OX phase. We also saw in Subsection 1.5.1 how depriving cells of oxygen by replacing the chemostat air feeding with Argon feeding for two minutes does not induce a fresh OX phase [157]. Such a result is consistent with this model for two reasons: one, the overall system phase Φ depends only on the aggregate φ_i values, which in turn depend only on the individual $[RNA_{OX}]_i$ values, which are not *per se* affected by a drop in the chemostat dO_2; two, it was discussed in Subsection 5.4.1 how the system takes approximately 10 minutes to respond to external perturbations. Unfortunately, because of budget constraints, this experiment was only performed with a 2 minutes long perturbation, which means its results are somewhat inconclusive. However, this evidence is a common conclusion of both the model and the Argon pumping experiment [157]: dissolved O_2 in the media is not one of the compounds driving, or synchronizing, the YMC.

In the future, it would be interesting to replicate this experiment[7] increasing the asphyxiation phase from 2 minutes to over 10 minutes; something like ≈ 20 minutes should allow for meaningful conclusions to be drawn. My prediction is that the result will be the same: lower dO_2 levels at the OX phase are merely a consequence of the YMC, and just by replicating that OX phase feature, no OX induction can take place. Note that in this case the OX phase cannot be identified with dO_2 drops, but rather with RNA-seq: well-known OX genes should be highly expressed on this dO_2 drop, but they are not. Similar phase identification via RNA-seq is the approach

[7] Maybe using N_2 rather than Ar, which is also an inert gas, and considerably cheaper. Compared to oxygen, both gases have an almost negligible media solubility [231], hence N_2 should be as good as Ar.

that should be followed if additional experiments were performed with mutants of different dO_2 traces, on which different curve shapes meant YMC phases could not be readily identified (data not shown).

5.5 O_2 consumption is identified with OX protein concentration

The chemostat oxygen probe dO_2 trace is the YMC experimental result most easily obtainable and accurate, since sequencing techniques require a considerable use of resources and are inevitably noisy. By expressing the predictions made in this Chapter in the form of predictions of the chemostat dO_2 trace under different conditions, I hope these hypotheses will be straightforward to test in an eventual experiment. In Section 1.5 I discussed the functional relationship between cellular oxygen consumption and dO_2 in the media. The goal of this Section is to link that cellular oxygen consumption model with the multicellular system model developed in the previous Sections of this Chapter in order to explain the form of the observed dO_2 oscillations, aiming to predict dO_2 trace variations at different experimental conditions and in mutant strains.

In Section 3.5 I calculated which GO terms are associated with genes maximally expressed at each of the three YMC phases, obtaining the clusters of genes presented in Fig. 3.6. In Table 3.3 I explained the main GO terms associated with each of these clusters, chiefly ribonucleoprotein biogenesis, and rRNA, ncRNA, and RNA processing and metabolic processes. Ribosomal processes are highly associated with cellular oxygen uptake [951–954]. Mitochondria are the only source of cellular oxygen consumption [645, 955–958]; the clusters of genes in Fig. 3.6 include, for genes expressed during the RB phase, several mitochondrial activity-related GO terms. Since high O_2 consumption levels start at the beginning of the OX phase, and are maintained during part of the RB phase, the model shown in Fig.4.10 suggests that Prot_{OX} concentration is a good proxy for the cellular oxygen consumption Q_{O_2}. Another possible path could have been extrapolating the results obtained in Subsection 1.5.2, Fig. 1.12, where the YMC high oxygen consumption phase is associated with OX-RB. It is interesting to note that the same result can be independently obtained by observing which cellular processes are consuming more oxygen providing a further validation of the models presented in this Chapter.

By making the approximation that Prot_{OX} concentration $[\text{Prot}_{OX}] \propto Q_{O_2}$, the cellular O_2 consumption, and merging Eqns. (1.3) & (1.4) with the $k = 0.1\ \text{h}^{-1}$ value calculated in Subsection 5.1.1, we obtain the differential equation

$$\frac{d(\text{dO}_2)}{dt} = k \cdot \left(1 - \frac{\text{dO}_2}{\max(\text{dO}_2)}\right) - \text{Prot}_{OX}(t), \tag{5.7}$$

whose boundary conditions to obtain a unique solution are that the system start point $t = 0$ is a maximum in $\text{dO}_2(t)$, hence

$$\text{dO}_2(t = 0) = \max(\text{dO}_2); \qquad \frac{d(\text{dO}_2)}{dt}\Big|_{t=0} = 0. \tag{5.8}$$

Figure 5.7 represents the numerical solution of Eqn. (5.7), confirming that the modelled Prot_{OX} concentration level (which was calculated with the system of ODEs given by Eqn. (5.2) and whose simulation results are in Fig. 5.3.a) exhibits a similar behaviour to the cellular oxygen uptake that was independently calculated in Subsection 1.5.2, simulation results in Fig. 1.12.a.

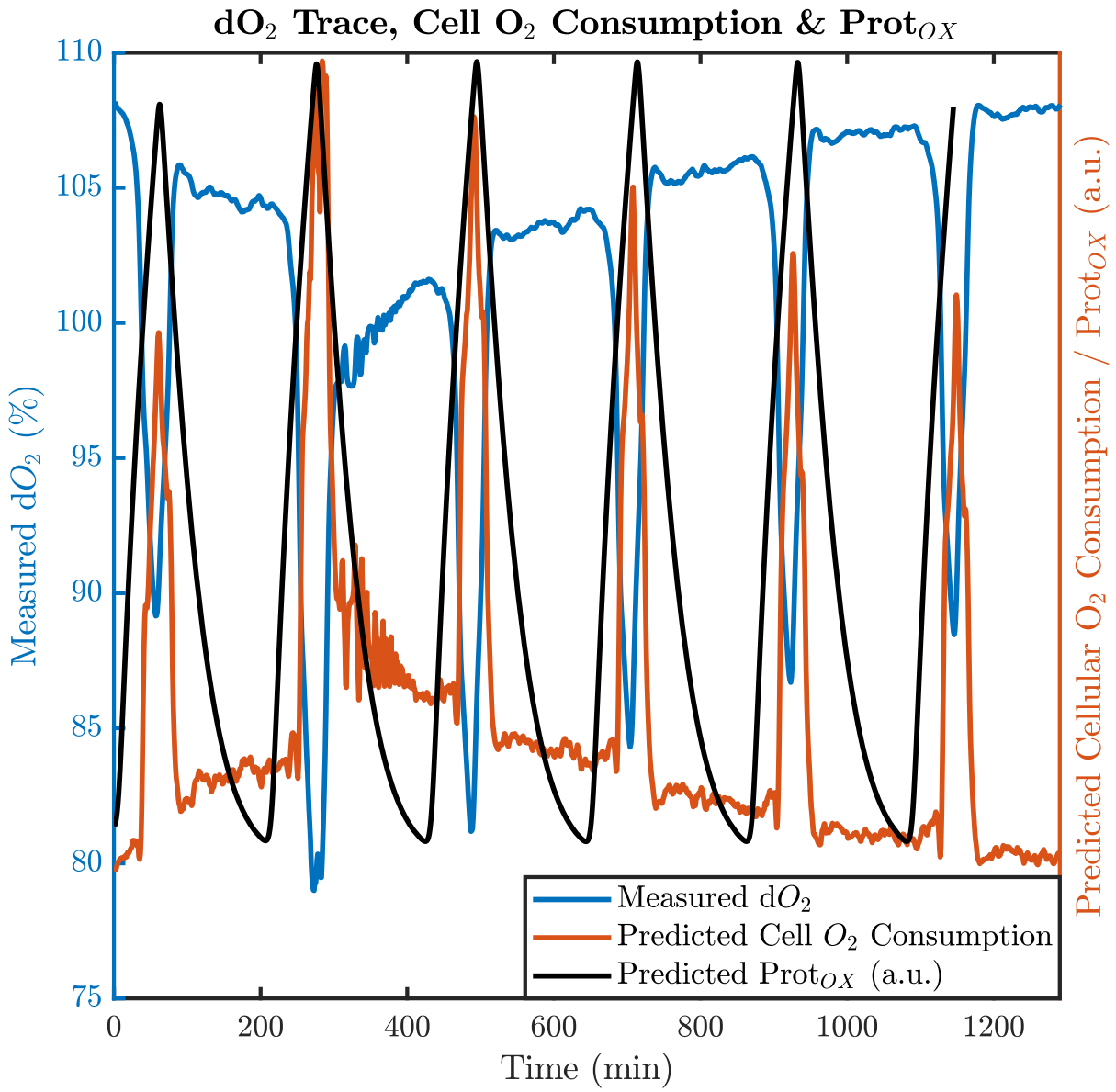


Figure 5.7: Dissolved oxygen trace and protein concentration model revealing the overlap of the experimentally measured YMC dissolved oxygen trace dO_2 (blue), the cellular oxygen consumption model calculated in Section 1.5 from such trace (red), and the $Prot_{OX}$ predicted concentration from the independent models of this Chapter (black). Note that both cellular oxygen consumption and $Prot_{OX}$ have arbitrary units, which means the goodness of fit is only concerned with the shape of the oscillations, the levels on the y-axis having been scaled accordingly.

5.6 Different period strains will not coordinate their oscillations

The YMC mathematical model constructed so far is composed of an internal oscillator, which is the transcriptional regulatory mechanism built in Chapter 4, and an external intercellular synchronisation mechanism, as discussed in Section 5.1, where the internal oscillator has a greater contribution to the YMC maintenance. This Section aims to produce a thought experiment that will test the relative strength of these two mechanisms in the model: the intercellular synchronisation mechanism whose strength is given by the $k_\alpha = 0.1\ \mathrm{h}^{-1}$ value calculated in Subsection 5.1.1, and the internal oscillator described by Eqns. (4.12) with the constants values given in Table 4.1.

Under the same experimental conditions, colonies of different WT yeast strains exhibit large differences in their YMC period [172]; for example, the CEN.PK strain has a period of 218 minutes [3], while DBY12007 has a period of 109 minutes [138]. Different WT yeast strains also have very different forms for their YMC dO_2 traces [138, Fig. 1B]. Finally, it is possible to merge two out of phase batches of the same yeast culture, and the one-batch oscillations observed in NADH fluorescence are eventually recovered [959, Fig.1], [101, Fig.1]; this model replicates the glycolytic oscillations recovery observed after a few phases.

The experimental design proposed in this Section merges these two ideas by predicting what would happen if the *out of phase batches merging* experiment [101, 959] was to be repeated, but having a different WT yeast strain in each of the batches. If the yeast intracellular transcriptional oscillator is stronger than the intercellular synchronisation mechanism, then there will be no coordination within the two strains, and each of them will cycle independently at their own intrinsic period. But if the external synchronisation mechanism is stronger, then they will eventually coordinate their oscillations.

CEN.PK and IFO 0233 are the two WT yeast strains to be used in this proposed experiment. CEN.PK [446] is a very popular strain in YMC studies [3, 112, 115, 118, 127, 138, 139, 151, 156, 157, 228, 267, 899, 960, and many others]. The IFO 0233 strain can be tuned via glucose feeding to have a similar ≈ 8 hours doubling time[8] [105]

[8] It is crucial to select two strains with doubling times as similar as possible: otherwise, by the moment they are mixed in the same fermenter, the fastest-growing culture will have outnumbered the other one, and will dominate any future form of YMC oscillations.

and has been used in some early YMC experiments [105,106,108,109,120,261][9]. The relevant point is that both strains exhibit very different YMC periods: CEN.PK has a $T_{CEN} = 218$ minutes period [3, 127], while IFO 0233 has $T_{IFO} \approx 40$ minutes [108, 109]. These experiments were performed under different conditions[10], therefore when brought under the same conditions their oscillatory periods may slightly differ. Because $T_{CEN}/T_{IFO} = 5.5$ and both experimental conditions are not too different, I expect them to have significantly different periods when cultivated under the same experimental conditions.

The literature claims that the only fundamental variation between the YMC cycles of the different yeasts is the highly strain-specific oscillatory period [138]; the underlying transcriptional and metabolic machineries are essentially not affected [228], except for sulfur assimilation, glucose uptake, and ethanol secretion [961]. Different experimental conditions other than glucose concentration in the feeding media are thought to be of little relevance [172]. Therefore, it can be assumed that the CEN.PK and IFO 0233 YMCs in fact possess the same metabolic process, but on a different time scale; the internal transcriptional oscillator is assumed to be regulated by a very similar clusters of genes in both strains, differing in YMC period only, and the intercellular machinery is driven by the same metabolites in both cases, as the literature seems to suggest [138, 172, 228, 961].

As mentioned before, the proposed experiment consists in inducing separately YMC oscillations and then mixing n_1 CEN.PK strain cells with n_2 IFO 0233 strain cells into the same chemostat. Mathematically, the resulting system is described by its corresponding ODEs, in which the cells ruling equations are divided into two groups: n_1 of them are governed by Eqn. (5.6) with the constants of Table 4.1, as we have done in the previous Sections, resulting in a period of $T_{CEN} = 218$ min, and the n_2 IFO 0233 strain cells have their constants proportionally scaled in the way proposed in Section 4.5 so that they oscillate with the same shape but a different period $T_{IFO} = 40$ min.

Figure 5.8.b represents the numerical solution of this system of ODEs, predicting no synchronisation between both strains: each of the two populations continues with its own characteristic oscillatory frequency, that is unaffected by the mixing. The physical explanation beneath this result is that the low k_α value of the intercellular coordination term cannot compel the transcriptional oscillator to change its char-

[9] I have not identified any experiment in the YMC field that uses IFO 0233 after 2006.

[10] Crucially, different glucose concentrations: 10 g/L for CEN.PK and 20 g/L for IFO 0233.

acteristic periods T_{CEN} and T_{IFO}, even if k_α is strong enough to make cells change their individual YMC phases φ_i into a synchronised state Φ. A coordination of populations n_1 and n_2 could not be observed unless a $k_\alpha = 3$ h^{-1} value was reached, which is 30 times the k_α value estimated in Subsection 5.1.1.

Figure 5.8.a reveals a second result of this model: the behaviour of an $n_1 = 1$ and $n_2 = 0$ single CEN.PK cell system, which is able to capture the yeast single-cell oscillations observed in the literature [134, 154, 886]. When $n = 1$, there is no synchronisation term in the form of Eqn. (5.1), since $\Phi = \varphi_1$, but noise is still considered by the model because of the η term introduced by Eqn. (5.5). The strong transcriptional oscillator of this system replicates these noisy oscillations shown in Fig. 5.8.a. In the high-temporal definition noise limit, RNA expression would follow a Weierstraß function [962, 963] [964, p. 7]. Note that these high noise levels are similar to the levels found in single-cell oscillations [134, Figs. 2C, 4A, 4G, 6B].

I explored the system dynamics under different ranges of n_1/n_2 and T_1/T_2 ratios aiming to find some form of synchronisation between both populations without modifying the k_α value, but even 10:1 $n_1 : n_2$ ratios did not retrieve a coupled oscillator (data not shown).

Therefore, I predict that when the experiment is performed, we will observe two distinct oscillations in the chemostat dO_2 trace. A simple Fourier transform will reveal that these dO_2 trace oscillations occur at exactly the same frequencies as the individual batches, the ratio of the Fourier peaks amplitudes A_1/A_2 being equal to the mixing ratio n_1/n_2.

5.6.1 dO_2 trace prediction

Using the O_2 consumption calculations of Section 5.5 it is possible to predict the dO_2 trace that will result from this experiment, setting $T_{IFO}/T_{CEN} = 40/218 = 0.18$ and $n_{IFO}/n_{CEN} = 0.5$. Figure 5.9.a reveals the numerical simulation results prediction for the dO_2 trace. As we can see, the co-culture experiment dO_2 trace form results in two combined, not coordinated, oscillations. Combined oscillations in the form $A_1 \sin(\omega_1 t) + A_2 \sin(\omega_2 t)$ are hard to deconvolve by eye [967–969]; to obtain the precise frequencies and amplitudes that compose the oscillation observed in Fig. 5.9.a it is necessary to calculate the Fourier transform of the predicted dO_2 trace, which is shown in Figure 5.9.b. This Fast Fourier Transform of the dO_2 trace exhibits two

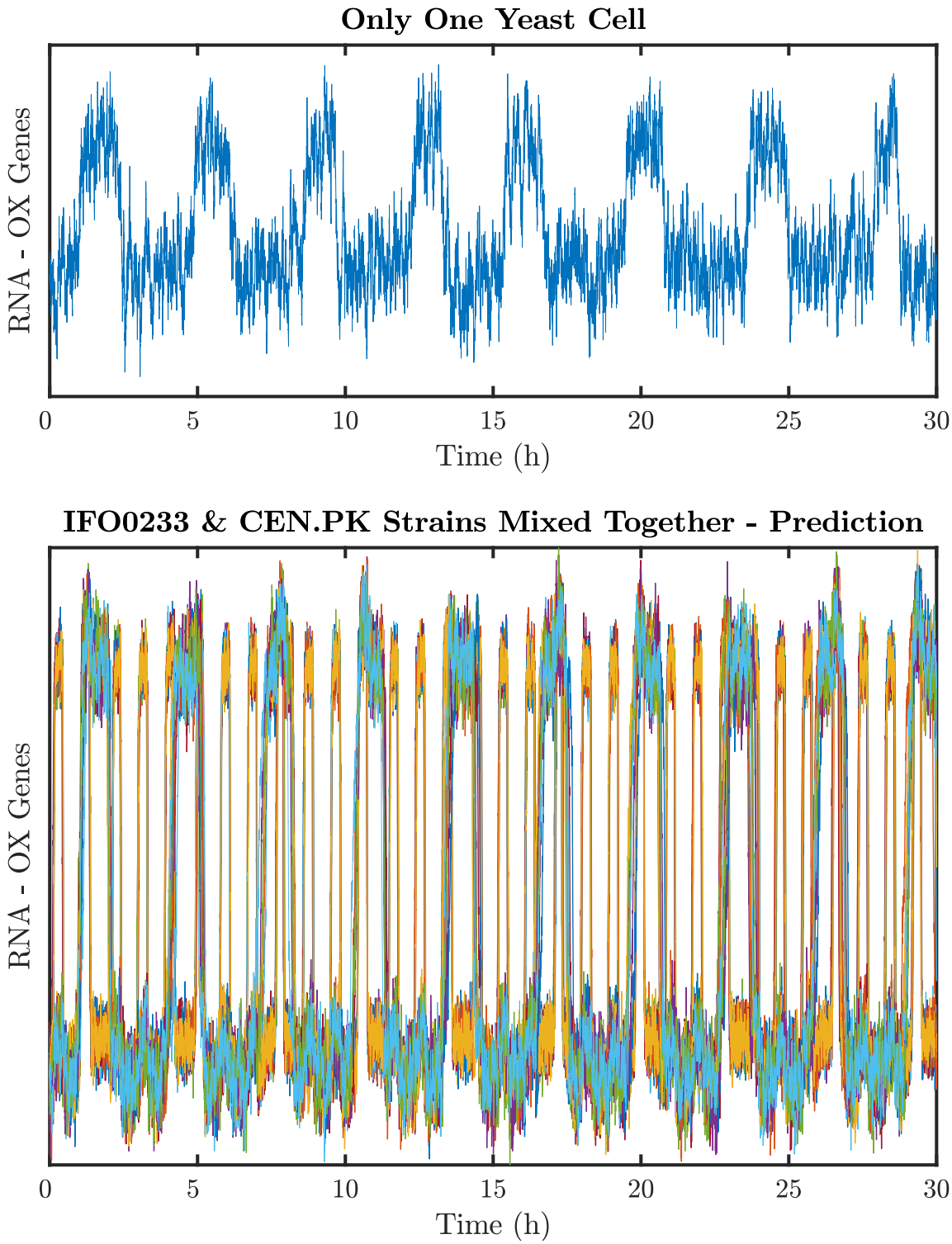


Figure 5.8: (Top) Numerical solution of the system described by Eqn. (5.6) for the single-cell in the chemostat $n = 1$ case. This model replicates the observed single-cell oscillations for unsynchronized cultures [154]; RNA expression follows a form of Weierstraß function [965, 966]. For high values of the noise term k_β -see Eqn. (5.5)-, regular oscillations are still detectable by a Fourier transform. (Bottom) Numerical solution of the co-culture system discussed in Section 5.6, with the ratios $T_1/T_2 = 0.3$, and $n_1/n_2 = 2$. A lack of coordination between both populations is observed: each strain cycles at its own characteristic frequency, as confirmed by Fourier analysis (see Fig. 5.9.b).

main peaks at frequencies $\nu_1 = 13$ and $\nu_2 = 2$, verifying that $\nu_2/\nu_1 = 2/13 = 0.15 \approx 0.18 = T_{IFO}/T_{CEN}$; the amplitudes of these two main peaks have a ratio $A_1/A_2 = 25/60 = 0.42 \approx 1/2 = n_{IFO}/n_{CEN}$. This approximation is even more accurate once the multiple-frequency peaks have been integrated together [970,971], in this case the $\nu_3 = 25$ peak observed in Fig. 5.9.b. The first peak in Fig. 5.9.b ($\nu = 2, A = 25$) corresponds to the CEN.PK strain oscillation and the second peak ($\nu = 13, A = 60$) corresponds to IFO 0233. The third peak arises because the MatLab `fft` algorithm [482, 483, 972] outputs only integer frequencies, so for a given non-integer frequency ν_0, smaller peaks appear at frequencies $n \cdot \nu_0, n \in \mathbb{N}$. Finally, note that because of the FFT algorithm, frequency in the figure x-axis is not measured in Hz (or mHz in this case), but in arbitrary units that are related to the total number of sample points as in Fig. 5.9.a.

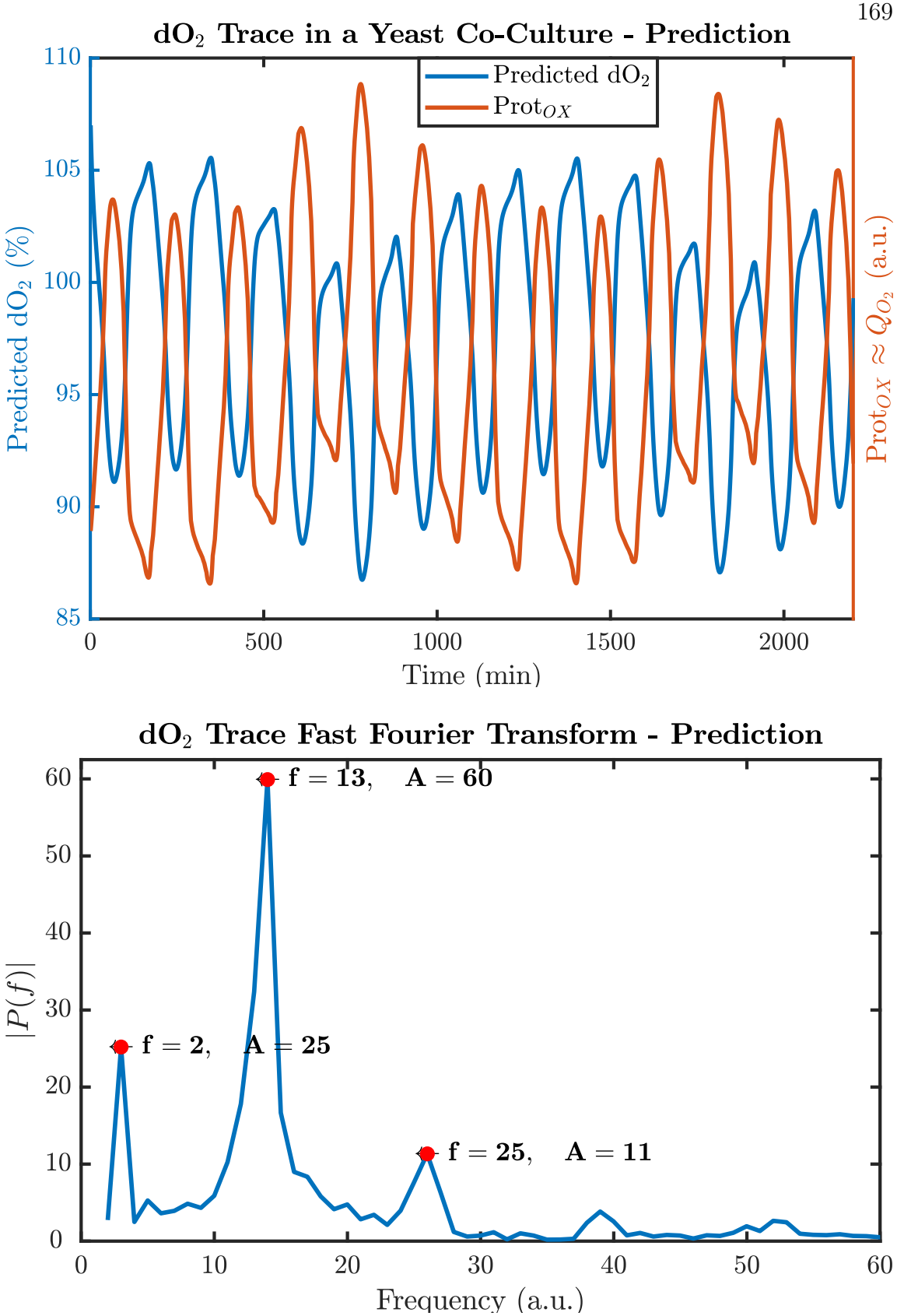


Figure 5.9: (Top): Predicted dO_2 trace for the CEN.PK and IFO 0233 strains co-culture experiment discussed in Subsection 5.6.1. (Bottom): Corresponding Fourier Transformation of the dO_2 trace on top using FFT [971]. The red dots mark the main peaks of the FFT function, $|P(f)|$. Only frequencies up to the Nyquist frequency $\nu_N = 0.5 f_S$ are plotted [973, 974].

5.6.2 Discussion and perspectives

The model developed in this Section predicts no coordination at all when yeast strains of different intrinsic YMC periods are mixed together, as seen in the dO_2 trace of Fig. 5.9.a. This Section provides a simple method for calculating the n_1/n_2 ratio between the number of cells belonging to each strain, via measuring the A_1/A_2 ratio of the corresponding Fourier peaks amplitudes, as evidenced in the analysis of the dO_2 trace in Fig. 5.9.b. Such a plot has three main peaks and two smaller peaks in the frequency domain, as opposed to the eight harmonic peaks found by experiments in short-cycle respiratory oscillations merged out of phase [261]. I believe this divergence is due to the fact that high-frequency Fourier peaks are extremely noise-sensitive [975], and the overlap in harmonics that is characteristic of the FFT algorithm is more crucially revealed under noisy conditions.

This analysis is completed with the anticipation of some of the experimental problems arising in the two strains co-culturing experiment, while providing some design guidelines for the initial experimental verification. First of all, it is crucial that, after some investigation on the co-culture's most appropriate growth conditions, each of the strains is let to cycle on its own to verify that:

1. Their periods under conditions allowing for simultaneous cycling are not too different from the individual cycles conditions [3, 108].

2. Under these conditions, both strains retain similar doubling times. If these times are more than a 10% different, it is also necessary to calculate the form of the $n_1/n_2(t)$ function while in co-culture, adding a considerable error source to the results. The accuracy of this calculation can be tested via a PCR analysis on a co-culture sample after the oscillations have been measured to verify that one strain has not overgrown the other. It is crucial to measure this ratio as accurately as possible to make sure it matches the predicted A_1/A_2 Fourier peaks amplitude ratio.

If growth rates are too different, it may not be possible to perform the experiment, since in the time from glucose starvation to glucose re-feeding and YMC start, the fastest-growing strain will have outnumbered the other strain by a fraction similar to the dO_2 probe sensitivity and the experiment will not be able to provide a definitive conclusion.

Table 5.1 shows the main differences between the IFO 0233 (short cycle) and CEN.PK (long cycle) strains in the main examples of their experimental studies. The

other media ingredients and rest of the experimental setup were analogous as far as I am aware. The difference between the media composition nutrients concentration in both experiments lies within the same order of magnitude, so it should not be hard to find a compromise feeding media composition on which both strains can cycle individually. It would be interesting to explore how tagging by radioactive labelling or by other means could help to accurately identify both populations, and maybe in the future allow for differentiated sequencing analysis in case the growth rates are just too different.

Table 5.1: Differences between CEN.PK and IFO 0233 strains YMC period, media composition, and culture parameters. The other media ingredients and rest of the experimental setup where analogous.

Strain	**CEN.PK**	**IFO 0233**
YMC Period	218 min	40 min
Main Experiment	[3]	[108]
Other experiments	See Section 5.6	[105, 106, 109, 120, 261]
Media Composition		
H_2SO_4	0.5 mL/L	1 mL/L
Antifoam A/204	0.5 mL/L	0.2 mL/L
Glucose	10 g/L	19.3 g/L
Culture Parameters		
Volume	1 L	0.65 L
Aeration Rate	1 L/min	0.15 L/min
pH	3.4	4
Cell Density	OD 8.5	$5 \cdot 10^{11}$ cells/L

5.7 Conclusion

The mathematical model developed in this Chapter describes the main force driving the YMC as a transcriptional, rather than metabolic, oscillator. When an acetate kick is introduced into the model, there is a phase response time of approximately 10 minutes. Note that this result depends only on the strength of the coupling constant term k_α included in Eqn. (5.1), whose value was calculated in Subsection 5.1.1, using only chemostat dissolved oxygen data including the YMC synchronisation time, before the acetate spiking experiments discussed in Section 5.4. This calculation agrees

with the phase induction time of 11 minutes observed in the experiments [118], therefore this result can be taken as a further validation of the functional form of Eqn. (5.1). I have predicted that there does not exist a threshold in acetate concentration increase required for cells to transition into OX, but the weaker the increase, the fewer cells make the transition, in an approximately linear manner. After this incomplete transition, I predict that under low concentration acetate spiking, the system will evolve in the manner described by the simulation results shown in Fig. 5.5.a,b: by exhibiting incomplete entrances into the OX phase that can be experimentally observed in the form of incomplete dips in dissolved oxygen at the beginning of the fresh OX phase. This way, we have confirmed that the intercellular synchronisation model developed and discussed in this Chapter is a system where the oscillation of metabolites such as acetyl-CoA acts as the pace-keeper of the cellular metabolic processes [116–118]. The calculations on the strength of this oscillator (see Subsections 5.1.1 and 5.6.1) reveal that intercellular synchronisation is much weaker than the internal transcriptional machinery, since the intracellular mechanism is able to change both the phase and the frequency of the oscillations, while cellular synchronisation via metabolites in the media can only change their phase.

Chapter 6

Conclusion

The aim of this doctoral thesis was to broaden the understanding of the YMC's dynamics and its underlying regulatory machinery, using bioinformatics analysis on multiple datasets first to identify the TFs beneath YMC regulation and then building a mathematical model able to explain the multicellular synchronisation mechanisms in a wide variety of experimental conditions in order to predict the system behaviour under a series of perturbations.

In Chapter 2, I described how H3K4me$_3$ cycles weakly, and accumulates for long windows of time after RNA expression has peaked. This means H3K4me$_3$ is unlikely to be instructive for transcription for the vast majority of genes during the YMC. Its appearance immediately after RNA expression, and for at least one YMC phase, suggests that the trimethylation mark is left by the Pol II to signal an already-transcribed gene that is not to be expressed until the mark is erased, i.e. until the next round of transcription or YMC iteration.

On the other hand, H3 tail lysine acetylations mirror RNA cycling. Correlation does not entail causation, but the appearance of early-tail lysine acetylations H3K9ac, H3K14ac, and H3K18ac precisely at the same temporal window as transcripts provides some evidence for 'just in time' acetylation supply of YMC gene products. However, the employed ChIP-seq data lack the time resolution required to fully test this hypothesis, since acetylation could occur after or before transcription, and it would not be detectable if the time lag is smaller than the sequencing time resolution. Finally, I proved how mRNAs produced at different YMC phases have different lifespans, suggesting a regulatory mechanism maintained, at least partially, by controlling mRNA decay and setting a basal gene expression level via mRNA

turnover. Most catabolic, protein degradation processes, occur in fact in saturated media, where degradation is stronger and, for a given gene, degradation occurs at a constant rate.

In Chapter 3, I explored different ways of clustering the yeast genome according to the individual genes progression during the YMC to learn what characteristics (in the form of GO terms) are shared by groups of genes with similar expression dynamics. I showed that the YMC is a highly compartmentalized system on which different yeast metabolic processes are segregated into different temporal moments, potentially to increase their metabolic efficiency. By making the hypothesis that genes belonging to the same cluster are co-regulated, I used YEASTRACT literature to trace the TFs associated with each of the three YMC phases, obtaining similar results to other TFs-retrieving algorithms in the literature, while individuating Sfp1p, Cup9p, and Ace2p as the chief TFs regulating the OX phase, Msn2p, Msn4p, Gis1p, and Adr1p for RB, and Xbp1p for RC.

Clustering multiple-datasets found a strong relationship between transcription and system progression through the YMC. This result provides good grounds for the hypothesis that the YMC is, essentially, a transcriptional oscillator on which intercellular synchronisation plays a relatively minor part. In this sense, the metagenes retrieved from the multiple-datasets clustering algorithms can be used as the building blocks for a mathematical model of the YMC transcriptional oscillator. Finally, by clustering multiple-datasets, I confirmed the result of Chapter 2 that early-tail H3 lysine acetylations H3K9ac and H3K18ac are highly dynamic and peak with the appearance of transcripts, while confirming that the ChIP-seq data provided for H4K5ac and H4K16ac are not reliable.

In Chapter 4, I built a single-cell transcriptional oscillator in which fluctuations in the levels of RNA and protein concentrations were able to retrieve the metagene profiles previously obtained in Chapter 3. Reproducing the YMC as a robust three-phases transcriptional oscillator with the appropriate metagenes dynamics provides good grounds for maintaining that the YMC regulation is transcriptionally rather than metabolically controlled. Such an oscillator also reproduces the phenomenon discussed in Appendix A.7: coded protein concentration peaks slightly after the corresponding mRNA.

In Chapter 5, I merged a plethora of single-cell oscillators by synchronizing them via a mean-field Hamiltonian term, a new method that explores the mechanisms un-

derlying the intercellular interactions that ultimately lead to the YMC macroscopic behavior, in order to explain why yeast populations set and maintain synchronized dynamics. These mathematical models reinforce the idea that the main force driving the YMC is a transcriptional, rather than a metabolic oscillator.

I have built an intercellular synchronisation model where the oscillation of metabolites such as acetyl-CoA acts as the pace-keeper of the cellular metabolic processes. Calculations on the strength of this oscillator reveal that intercellular synchronization is much weaker than the internal transcriptional machinery, since the cellular internal mechanism can tune both the phase and the frequency of the YMC oscillations, while the intercellular component can only change their phase.

This theoretical work has also led to new research questions. In the following Section, I propose experimental work that would contribute towards gaining a deeper understanding of the YMC machinery, corroborate the hypotheses presented in this thesis, and pose new questions.

6.1 Proposed future work

Multicellular sequencing data cannot reveal if, during the YMC, all of the cells in the chemostat are in the same metabolic state, or if a fraction of them are in quiescence while the rest are cycling. A study of the YMC yeast transcriptome at the single-cell level would reveal what fraction of the cells are cycling, and whether all the cells exhibit metabolic cycles of an identical form. The YMC is a low-energy system on which the low glucose environment could very easily induce quiescence in most of the cells [976–978]. Therefore, the single-cell simulation results in Sections 4.9 and 5.6 could be tested employing FISH analysis [979] or single-cell RNA-seq [938] to ascertain the behaviour of not synchronised cells. If the sequencing techniques' noise could be significantly lowered [387], then it would be possible to evaluate the system noise k_β constant introduced in Section 5.3 and compare single-cell RNA expression dynamics with the prediction of a single yeast cell RNA dynamics shown in Fig. 5.8.a.

A second question arising from this thesis concerns the precise nature of the interplay between the YMC and the CDC [128, 129, 134]. An accurate measurement of the time evolution of cell density would allow us to calculate the rate at which cells are budding during the YMC. Such measurement is not be straightforward, since the YMC is coupled to the CDC [128, 134], which means cells change in size during the YMC, and changes in OD cannot be attributed solely to changes in cells concentra-

tion. If these changes were measured, an accurate model of the YMC-CDC coupling could be built by using the chemostat dilution equations [932] and corresponding cellular concentration evolution dynamics [980, 981]. Microscope imaging of YMC samples at high time-resolution would let us to accurately identify CDC progression by looking at the cells' shape, which indicates the cell division stage [982, 983].

The co-culture experiment proposed in Section 5.6 could reveal whether the fundamentally different shape of the YMC oscillations observed by [3] on one hand and by [108, 109] on the other is due to the same biochemical process with very minor variations caused by the different strain and environmental conditions. This would imply that the YMC frequency is highly tunable, as some authors predict [134, 138]. A related experiment would be to mix a batch of cells progressing through the YMC with an identical batch grown under the same conditions, but without a glucose starvation phase, so that the latter does not enter the YMC. According to my model, this is equivalent to a new system where the distribution of cells in phase space is not described by Eqn. (5.4), but rather by a mixture of a Dirac δ function [984, 985] (see Fig. 5.2.d) and a uniform distribution (Fig. 5.2.a) which converges quickly into the predicted synchronised state. The convergence speed to such state is a robust test for the $k_\alpha = 0.1\ \mathrm{h}^{-1}$ value obtained in Subsection 5.1.1.

Cells undergoing the YMC are some times observed to lose their synchrony after a couple of weeks, with the last few oscillations having a significantly shorter period [136, 222, 888]. Whether this is a consequence of some experimental disruption during cycling is hard to ascertain, since there are a number of technical difficulties involved in keeping the chemostat running for long periods of time. It would be useful to compare carbon-rich metabolite concentrations during both the first and the last YMC oscillations to learn if some accumulating metabolite is bringing the YMC to its end, since low media glucose levels are necessary for maintaining metabolic cycles [136, Fig. 5]. The mathematical description of this phenomenon would consist of a term in the form of Eqn. (5.1), but having $k_\alpha < 0$, creating hence the necessary divergence for the system synchrony to be lost.

The simulation of different strengths and lengths of the acetate spiking described in Subsection 5.4.2 has not been confirmed experimentally; trying to induce YMC OX phases with decreasingly lower changes in acetate (or analogous C-source) concentration would test the simulation results displayed in Fig. 5.5. Similarly, it would be interesting to replicate the argon pumping experiment [157] discussed in Subsection 5.4.2, using N_2 rather than Ar, and definitely for a longer time: if it takes

10 minutes for the OX phase to be induced, it does not make sense to perturb the system for just 2 minutes; an over 10 minutes perturbation is required to see whether dissolved oxygen in the media levels are actually affecting the YMC synchrony.

Finally, it would be relevant to pursue the idea that acetate pouring induces a fresh OX phase because cells change their preferred acetyl-CoA synthesis pathway from glycolisis to ethanol and acetate conversion.[1] This means yeast cells would shift from a fasted survival mechanism -the YMC, where glucose is scarce- to fed growth in abundant carbon-sources, hence resulting in increased levels of histone acetylation [989] that would be experimentally measurable [127]. The acetate $\rightarrow$ acetyl-CoA reaction is mediated by an enzyme encoded by $ACS1$ [990], but the glucose $\rightarrow$ acetyl-CoA pathway is more complex and has pyruvate as a necessary intermediate [991]. The pyruvate $\rightarrow$ acetyl-CoA part of the reaction is mediated by enzymes encoded by the $PDA1$ [991], $PDB1$ [992], and $PDX1$ [993] genes. Therefore, another way of testing my hypothesis would be to measure (via RNA-seq) the expression patterns of these four genes both during the YMC, where the acetate $\rightarrow$ acetyl-CoA pathway would be negligible and therefore $ACS1$ should be seldom expressed, and during an acetate-abundant culture, where $PDA1$, $PDB1$, and $PDX1$ would not be expressed, since the pyruvate $\rightarrow$ acetyl-CoA reaction would not be so frequent.

By using mathematical modelling techniques and bioinformatic analysis on a wide variety of sequencing data, in this DPhil thesis I have elucidated some aspects of the convoluted YMC regulatory machinery. My work has also raised questions that can be addressed experimentally to lead to a better understanding of the YMC regulation.

[1] Yeast acetyl-CoA metabolism has two chief pathways: either glycolysis in the cytosol [986, 987] or directly from acetate either in the cytosol, the peroxisome, or the nucleus [986, Fig. 1], [988].

Appendix A

Appendix

The purpose of this Appendix is to expand discussions and calculations that, despite being relevant to understand some of the concepts presented in this thesis, have lengths and complexities suggesting they be separated from the main text to enhance its readability. In particular, they are important to understand the technicalities of data processing underlying the bioinformatics work presented in the main text.

In this Appendix I describe the different datasets used in this thesis (Appendix A.1) and their limitations in terms of quality and noise (Appendix A.2). Large sequencing datasets are rendered manageable using information condensation techniques (Appendix A.3) whose accuracy is verified using metagenes analysis (Appendix A.4). Afterwards, two mathematical proofs are included justifying the claims made in Subsection 3.4.2 that both cluster overlap metrics discussed in that Subsection -see Eqns. (3.9) and (3.12)- verify the properties of a distance measure (Appendix A.5). Appendix A.6 discusses some of the computational details underlying the numerical solutions of the ODEs systems presented in Chapters 4 and 5. Finally, there is a discussion on the possible role of proteins concentration (Appendix A.7) and iron-sulfur proteins (Appendix A.8) in the YMC regulation.

A.1 Data description

The purpose of this Section is to explain the nature and quality of the sequencing data used in this work. It must be noted that these data are highly inhomogeneous, and the most relevant differences in experimental conditions need to be kept in mind when comparing results and drawing conclusions. It is important to understand the

different dataset limitations, since sequencing techniques are complex, noisy, and prone to a wide variety of errors throughout the experiment, which means that the raw published data cannot be used straightforwardly for the bioinformatics analysis.

A.1.1 RNA-seq and Microarray

Boeke 2014 data

Boeke *et. al.* [127] measured genome-wide RNA-seq of WT yeast CEN.PK113-7D strain undergoing the YMC at the 16 timepoints shown in Figure A.1. These data are published in the Gene Expression Omnibus database [994–997] under accession code GSE52339 (samples GSM1263339 to GSM1263354). These data are not strand-specific.

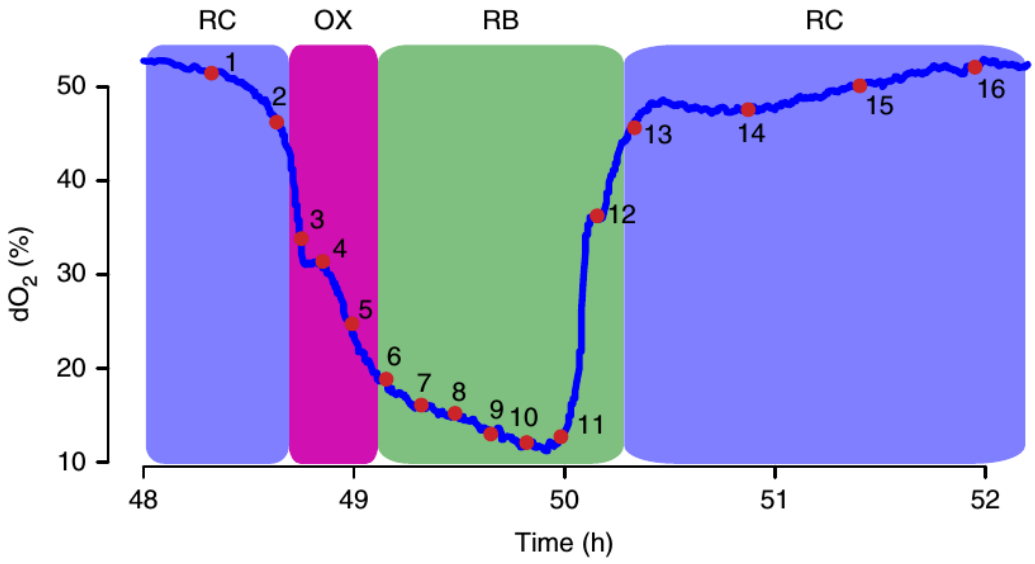


Figure A.1: Sampling timepoints of the Boeke 2014 RNA-seq data featuring 16 time points selected across one period of the YMC; y-axis is the media dissolved O_2 concentration (dO_2), as measured by the chemostat probe, x-axis is time. Background: OX, RB, and RC YMC phases associated with the processes identified in Table. 1.1. Taken from [127].

Tu 2005 data

Tu *et. al.* [3] used Microarrays to sample 36 time points selected across three consecutive periods of the metabolic cycle of WT yeast CEN.PK113-7D strain. These data are not strand-specific. Microarray data can be found at the DOI `10.1126/science.1120499`.

Li 2006 data

Li *et. al.* [108] used Microarrays to sample 48 time points selected across four consecutive periods of the metabolic cycle of WT yeast IFO 0233 strain, as shown in Figure A.2. After the first cycle, a 3 mM dose of phenelzine was added, as discussed in Subsection 2.1.1. These data are not strand-specific. Microarray data can be found at the DOI `10.1073/pnas.0604860103`.

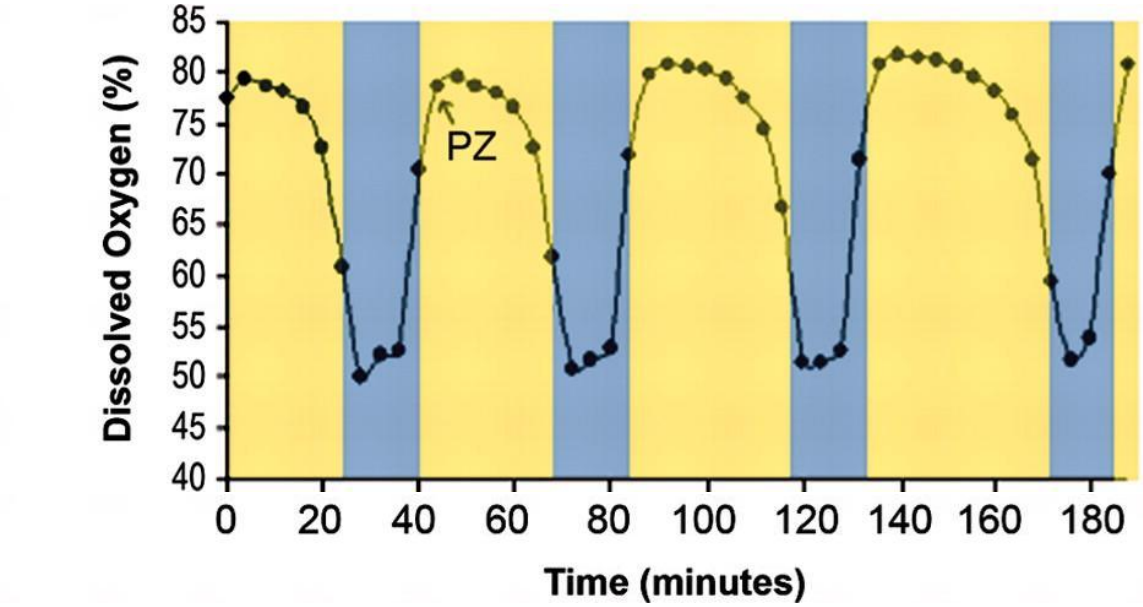


Figure A.2: Sampling timepoints of the Li 2006 [108] sequencing data. 48 time points selected across four consecutive metabolic periods. PZ marks the Phenelzine addition point. Background: high (blue) and low (yellow) oxygen consumption parts of the YMC. Taken from [108].

A.1.2 ChIP-seq

Boeke *et. al.* [127] sequenced eight histone marks over one YMC period at 16 points: histone H3 lysines 9,14,27, and 56 acetylation, histone H4 lysines 5 and 16 acetylation, and histone H3 lysines 4 and 36 trimethylation.

The 16 ChIP-seq sampling timepoints are not coincident with the 16 RNA-seq sampling timepoints in the same paper (see Subsection A.1.1). Both datasets are compared using a smooth curve with a penalized B-spline [376,998], where the coefficient of each basis function is estimated via a penalized optimization problem. These data are not strand-specific. Data are in the Gene Expression Omnibus database under accession code GSE52339 (samples GSM1263355 to GSM1263498).

ChIP-seq data was generated in two batches, H3K9ac being independently measured in both of them to minimize batch-to-batch variations [127, Methods]. These data have a resolution of approximately 100bp. The same lab has sequenced H3K18ac at the same 16 timepoints in a collaboration with the Mellor lab (data not published).

Possible misreads in some HMs

There is a possibility that some of the ChIP-seq antibodies did not work as expected due to problems with antibody quality [466]. Antibody adherence is a recurring problem in the field of sequencing [464], being responsible for many paper retractions [465] and wrongly conducted experiments [999]. The paper did not perform a specific antibodies validation, even if the quality of antibodies can vary among batches of the same antibody [149]. In particular, my bioinformatics analysis casts doubts about the batches of Millipore 07-327 and Millipore 07-329 Merck antibodies used to sequence H4K5ac and H4K16ac. From my data analysis, I am almost certain that H4K5ac and H4K16ac sequencing data in Boeke 2014 [127] are incorrect, and have therefore been excluded from my analysis, as discussed in Subsection 3.2.1 of the main text. On the same line, anomalous results at the $t = 1$ and $t = 16$ datapoints for several HMs are divergent with the literature [1000–1002], suggesting that these samples may not have the $> 10^7$ quantity of cells required for proper sequencing [1003, 1004].

A.1.3 NET-seq

The Mellor lab [373] has obtained NET-seq data at 12 timepoints of WT CEN.PK strain yeast undergoing the YMC, as shown in Figure A.3. Measuring the nascent, actively transcribing RNA polymerase II is relevant to explore if the YMC is transcription-driven in Chapter 3. NET-seq data are strand-specific [241]; the plus (sense transcription) strand was selected to co-cluster the data with the non strand-specific RNA-seq data in Boeke *et. al.* [127].

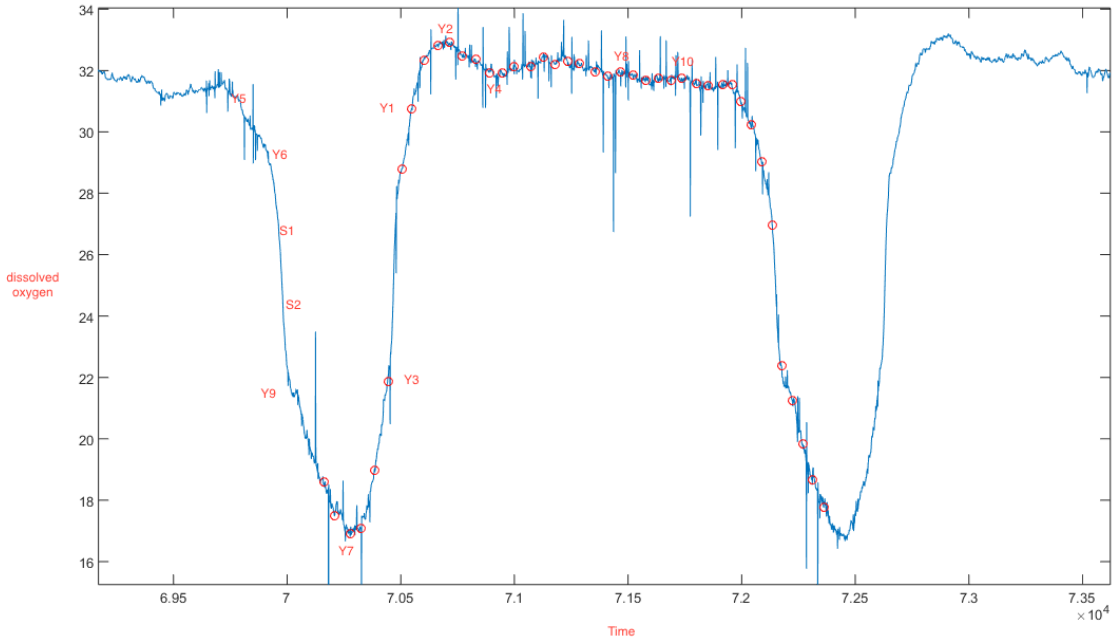


Figure A.3: 12 NET-seq sampling points [373]. x-axis: time; y-axis: dissolved oxygen concentration in the media. Points labelled Y and S were sampled and sequenced. Other marked circles correspond to the 3' end sequencing points used to compare the trace with the Boeke 2014 [127] data. Image credit: Shidong Xi [373].

A.2 Data quality and possible error sources

The first limitation that must be considered when handling sequencing data is that they are extraordinarily noisy [387]. Noise limitations entail the necessity of establishing a 2-fold change for claiming a significant variation in the sequencing output [572, 578]. The strategy used in this thesis to minimize this problem is not to

consider individual genes reads by their own, but rather to cluster similarly behaving genes to minimize the effect of noise in individual measurements.

Nominally, RNA-seq and NET-seq have single base pair resolution [582, 1005], while the resolution of ChIP-seq data depends on the DNA fragments' size (approximately 150 bp on average), but they cannot derive gene and isomorph expression values from raw sequencing reads [1006] while accounting for variability in measurements across biological replicates [1007, 1008]. Read mapping uncertainty and the frequent yeast isoforms mean that expression levels cannot be accurately calculated for some genes [835].

A particular concern with the RNA-seq data in Boeke 2014 [127] dataset is that the Illumina HiSeq 2000 readings intensity seems to saturate rather quickly [1009, 1010]. In this sense, gene by gene analysis of the RNA-seq reads reveals that the maximum Illumina intensity (read 100 in the normalized units provided by the Boeke lab) is reached for long periods of time (data not shown), raising concerns about the precise location of peaks in RNA expression. RNA-seq is more accurate and reproducible than Microarrays [1011], therefore the concerns raised before for RNA-seq apply also to the Tu 2005 [3] and Li 2006 [108] datasets analysis.

Errors in the sequencing reading or mapping: data 'Inf' and 'NaN' inputs were set to a zero value to make different datasets more easily compared, while reflecting the fact that a zero reads count in the Illumina signal can be due to either the actual absence of transcription or to some error in the sequencing protocol.

A.3 Information condensation

This Appendix is oriented towards making all of the relevant data readily available, in a compatible and systematic format, for the rest of the project. It also investigates the error margins and reliability of the different datasets. This task has a crucial importance since base-pair resolution sequencing files on the Cuffidff format are approximately 5Mb in size; the yeast genome is 12.5 Mbp long [1012, 1013], therefore studying just one sequencing datapoint 'by eye', without condensing its information, would be an extraordinarily lengthy task. The aim of this Appendix is hence to form a transition from bp-resolution sequencing data to more manageable datasets. Its goal is to condense information minimizing content loss in order to obtain smaller, homogeneous, datasets on which to perform the relevant bioinformatics analysis.

I use different techniques to explore the overall sequencing data and learn how the different YMC indicators behave at the most fundamental level, in order to form a preliminary model of their spatial and temporal relationships with one another. The first simplification I make is to consider chromosome regions in the neighbourhood of a gene only. For my analysis, I only consider the yeast verified genes, disregarding uncharacterised and dubious genes, which is a common practice in the literature [1014, 1015]. The rationale underlying this approach is my interest in establishing regulatory networks and compare the relevant genes cellular functions as provided by their associated GOIDs, but if the genes have not been characterized, this cannot be done.

Since I am interested in the evolution of sequencing quantities for each of the relevant genes, I look for a reliable way of condensing the rich, base pair resolved information, into one number that can provide us with a comprehensive understanding of the amount of RNA or a particular HM present at a given gene.

Nomenclature: Gene-related windows in brackets indicate positions relative to the TSS: for example, [-100 200] is the genomic region spanning 100 bp upstream the TSS and 200 bp downstream.

A.3.1 From base pairs to genes: appropriate size of genomic windows

This Subsection is devoted to obtaining an appropriate gene normalizing window from which to calculate the relevant sequenced magnitude at a given gene; in short, to obtain a single data point for each gene and sequencing dataset while balancing computational simplicity and resolution.

Single data points will become the first 'building blocks' used for the bioinformatic analysis in Chapters 2 and clustering analysis in Chapter 3, from which the mathematical models of Chapter 4 and 5 are derived. The relevant bioinformatic questions to be answered are:

Which genetic region, $\mathcal{R}$, is relevant for this study? **and** *How can the information encoded in $\mathcal{R}$ be condensed into a form that can be more easily modelled and understood?*

Such analysis, as we can see in Figures A.4 and A.5 reflecting the metagene analysis of Appendix A.4, reveals that for both RNA-seq and ChIP-seq data, the overall metagene shape does not change significantly through the YMC at different genomic windows, small changes being observed only as an amplitude adjustment of the accumulation levels. Therefore, it is appropriate to take the gene body + promoter average accumulation level as a value representative of the overall genetic dynamics. Merging both results, the best way of ordering the data is through a function $(g, t, h) \to f(g, t, h)$ which gives the average value, in the [-250 TTS] region of gene g, of the accumulation level of sequenced magnitude h at time t.

Conclusion

The [TSS-250 TTS] region average of RNA or HMs expression/accumulation level is a representative metric for the measurement of overall HM accumulation level that accurately reflects temporal evolution within the gene. This metric does not privilege longer genes, and rather than on intergenic HMs distribution, it is primarily concerned with capturing time evolution, which is the magnitude this bioinformatics analysis and mathematical model are mostly concerned with.

Therefore, the bp-resolution information provided by the datasets discussed in Appendix A.1 is condensed using the [TSS-250 TTS] region average. Mathematically expressed, this corresponds to the function

$$\begin{aligned} f &: \mathcal{G} \to \mathbb{R} \\ &\mathcal{G}_i \mapsto \texttt{RM}([-250 \;\; \text{TTS}]_i), \end{aligned} \tag{A.1}$$

where $\mathcal{G}$ is the set of the 5770 yeast characterized genes $\mathcal{G}_i$ and RM is the reads mean on the [TSS-250 TTS] interval.

A.3.2 Comparison with Boeke 2014 windows

Boeke 2014 has opted for *ad hoc* windows for each of the HMs, on which read counts are averaged over the following regions:

- H3K9ac, H3K14ac, H3K56ac, H4K5ac and H3K4me$_3$: [-100 400] window
- H3K36me$_3$, H4K16ac and H3: [TSS TTS] window

In Appendix A.4 it will be reiterated how the genomic window choice has a negligible impact on the results.

A.4 Metagene construction

This Appendix aims to build metagene profiles of RNA and HM time evolution during the YMC to understand how they evolve at the different parts of an average, representative, gene body, by using the data in Appendix A.1, and finally comparing HM dynamics at different loci of the gene body and surrounding genomic regions.

The metagene associated with RNA-seq or ChIP-seq data is constructed using the following algorithm:

1. For each gene, the [-250 TTS] region is binned:

   1.1 Counting in the downstream direction, the 250 bp in the promoter region [-250 TSS] are divided into 25 bins, and take the average read in each of these 10 bp bins.

   1.2 The [TSS TTS] region is divided into 100 bins also downstream. Bin lengths are rounded to the closest integer. These bins have different lengths depending on the gene length.

   The outcome is, for each gene, a list of 125 values, 25 for the promoter [TSS-250 TSS] region and 100 for the gene body [TSS TTS].

2. For each bin, we consider the average value of all the genes belonging to one of the three YMC clusters: OX, RB, and RC [3, 127]. This results in the metagenes shown in Figures A.4 and A.5.

I observe that H3K36me$_3$ accumulates roughly in the middle of the gene body [Fig. A.4.b], while H3K14ac and H3K18ac are more frequently expressed right after the TSS and in the gene promoter [Fig. A.5.c]. For H3K9ac, accumulation is also more frequent around the TSS region, but it also accumulates on the gene body [Fig. A.4.c]. Peaks are time-dependent, consistent with the oscillatory dynamics observed in gene expression [3, 127].

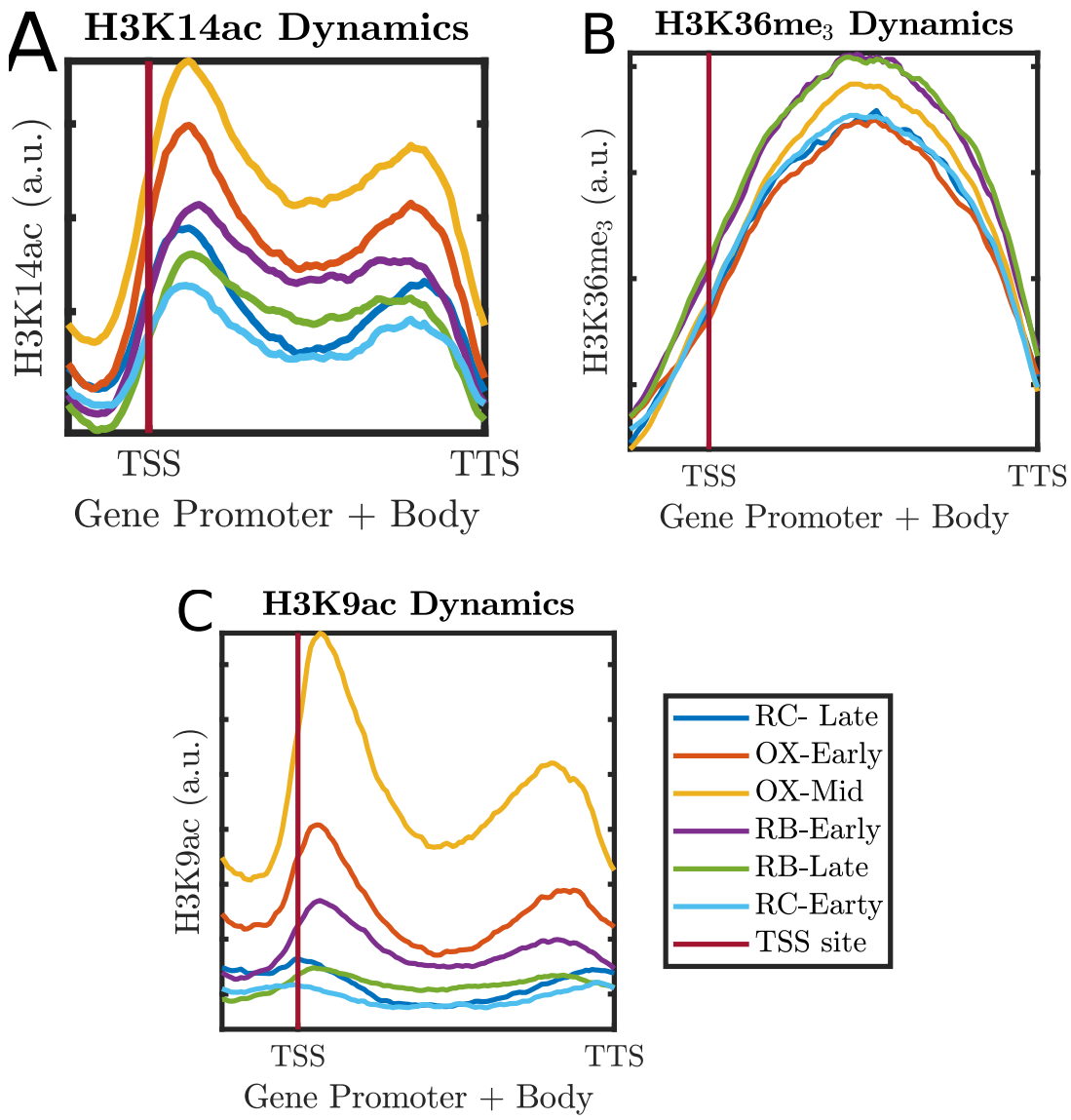


Figure A.4: Lines represent levels of (a) H3K14ac (b) H3K36me$_3$ (c) H3K9ac at a given time across the OX metagene (average values for all the genes belonging to the Boeke 2014 OX cluster [127]) constructed with the algorithm discussed in the main text. x-axis: genomic region; y-axis: HM accumulation level in arbitrary units.

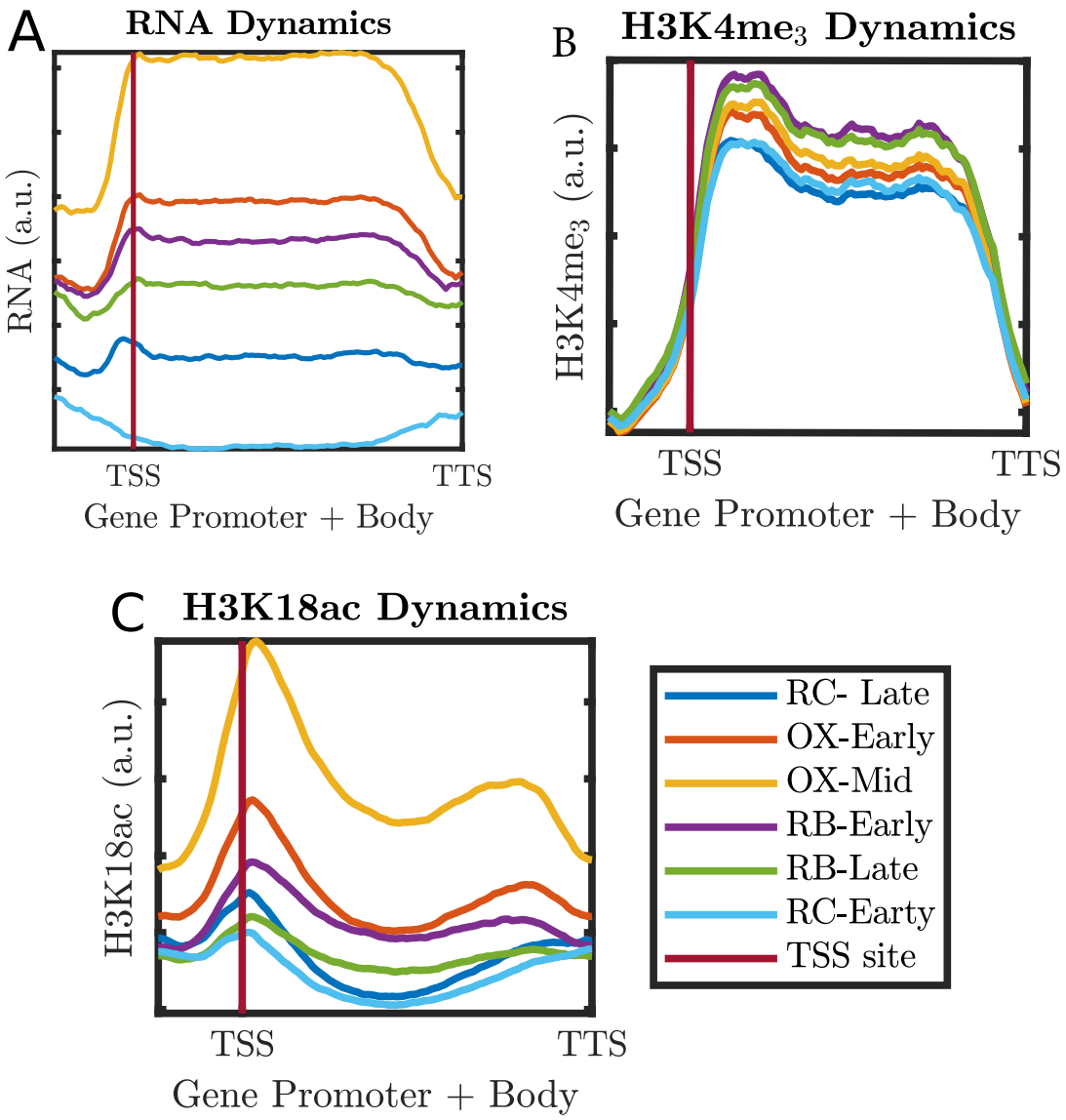


Figure A.5: Lines represent levels of (a) RNA (b) H3K4me$_3$, (c) H3K18ac at a given time across the OX metagene (average values for all the genes belonging to the Boeke 2014 OX cluster [127]) constructed with the algorithm discussed in the main text. x-axis: genomic region; y-axis: RNA expression or HM accumulation level in arbitrary units. Crucially, the curves corresponding to different timepoints do have very similar shapes and in general are non-intersecting. This feature is extraordinarily relevant to justify the methodology employed in Appendix A.3, since it reveals sequencing data time-evolution results are fairly independent on the genomic window employed in the bioinformatic analysis, leading to the conclusion that the average across a wide variety of genomic regions is a good proxy for HMs dynamics. Fairly equidistant, non-intersecting lines mean that window choice is fairly irrelevant for the gene-by-gene analysis, since the same time evolution is observed approximately for a given metagene under different windows. This result gives us some relief in the widely-discussed literature topic of which is the 'right' genomic window for each HM [591, 1016–1019].

A.5 Proof of theorems in Subsection 3.4.2

This Appendix presents the mathematical proof of the two theorems presented in Subsection 3.4.2, which for clarity purposes is stated here, since it is too lengthy to be included in the main text in a detailed manner

Theorem 1. *Function d, as defined by Eqn.* (3.9)*, verifies the properties of a distance measure: non-negativity, symmetry, and the triangle inequality.*

Proof. 1. Non-negativity. $A \cap B \in B \rightarrow |A \cap B| \leq |A| \rightarrow |A| - |A \cap B| \geq 0$. Similarly, $|B| - |A \cap B| \geq 0 \rightarrow d(A,B) \equiv |A| + |B| - 2 \cdot |A \cap B| \geq 0$

2. Symmetry. $A \cap B = B \cap A \rightarrow d(A,B) \equiv |A| + |B| - 2 \cdot |A \cap B| = |A| + |B| - 2 \cdot |B \cap A| = d(B,A)$

3. Triangle inequality. Note that

   $|B| = |B \backslash (A \cup B)| + |A \cap B| + |B \cap C| - |A \cap B \cap C| \rightarrow |A \cap B \cap C| \geq |A \cap B| + |B \cap C| - |B|$,

   where $B \setminus (A \cup B)$ is the set of elements in B not belonging to either A or C.

   Then, we have that

   $|A \cap C| \geq |A \cap B \cap C| \geq |A \cap B| + |B \cap C| - |B| \rightarrow 2 \cdot |A \cap C| - 2 \cdot |A \cap B| - 2 \cdot |B \cap C| + 2|B| \geq 0.$

   And applying the definition of d, $d(A,B) + d(B,C) - d(A,C) = (|A| + |B| - 2 \cdot |A \cap B|) + (|B| + |C| - 2 \cdot |B \cap C|) - (|A| + |C| - 2 \cdot |A \cap C|) = 2|B| - 2 \cdot |A \cap B| - 2 \cdot |B \cap C| + 2 \cdot |A \cap C| \geq 0$.

□

Theorem 2. *Function D, as defined by Eqn.* (3.12)*, verifies the properties of non-negativity and symmetry.*

Proof. 1. Non-negativity. $D(C_1, C_2) = \frac{1}{n+m} \left(\sum_{i=1}^{n} d(S_i, S'(S_i)) + \sum_{j=1}^{m} d(S'_j, S(S'_j)) \right)$. Because $n, m \in \mathbb{N}$, $1/(n+m) > 0$. Using Theorem 1, $d(S_i, S'(S_i)) \geq 0 \rightarrow \sum_{i=1}^{n} d(S_i, S'(S_i)) \geq 0$, and the same follows for the second summand in the RHS. Therefore, $D(C_1, C_2) \geq 0$.

2. Symmetry. Follows trivially from the fact that the sum operation verifies the commutative property, and so does function d ex Theorem 1.

□

A.6 Computational details

A.6.1 Code in Subsection 4.3.5

From the system of Equations (4.12), we can write the Matlab SimBiology equations describing the different circuits [765]. For example, a very simplified form of a $RNA_1 \dashv RNA_2 \dashv RNA_1$ loop as the one sketched in Figure 4.2 can be implemented by the code below. Note that the values of the different constants must be obtained first through the simulated annealing algorithm presented in Subsection 4.6.1; in this case, arbitrary values for the parameters have been selected.

```
M = sbiomodel('cell'); % Simbiology Model

p = addparameter(M,'p',3); % Numerical values of the different ...
constants
KM = addparameter(M,'KM',6);
KDegOX = addparameter(M,'KDeg1',7);
KDegRB = addparameter(M,'KDeg2',8);
KsOX = addparameter(M,'Ks1',10);
KsRB = addparameter(M,'Ks2',11);
KdpOX = addparameter(M,'Kdp1',13);
KdpRB = addparameter(M,'Kdp2',14);
KRsOX = addparameter(M,'KRs1',16);
KRsRB = addparameter(M,'KRs2',17);

compObj = addcompartment(M,'Cell');

RNAOX = addspecies(compObj,'RNA1'); % Phase 1 RNA Concentration
RNARB = addspecies(compObj,'RNA2');
PrOX = addspecies(compObj, 'Prot1'); % Phase 1 Protein Concentration
PrRB = addspecies(compObj,'Prot2');

compObj.Species(1).InitialAmount = 0; % RNA 1 Initial Amount
compObj.Species(2).InitialAmount = 0;

compObj.Species(3).InitialAmount = 10; % Protein 1 Initial Amount
compObj.Species(4).InitialAmount = 10;

RNA1eq = addrule(M,'RNA1 = (KRs1/(1+Prot2^p) - ...
    KDeg1*RNA1)','rate'); % RNA1 Rate Equation
RNA2eq = addrule(M,'RNA2 = (KRs2/(1+Prot1^p) - KDeg2*RNA2)','rate');

Prot1eq = addrule(M,'Prot1= (Ks1*RNA1 - Kdp1*Prot1/(KM + ...
    Prot1))','rate'); % Protein1 Rate Equation
```

```
Prot2eq = addrule(M,'Prot2= (Ks2* RNA2 - Kdp2*Prot2/(KM ...
    +Prot2))','rate');
configset = getconfigset(M);
configset.CompileOptions.UnitConversion = false;
Sol = get(configset, 'SolverOptions');
Sol.MaxStep = 20;
set(configset, 'SolverType', 'ode23t') % Use ode23 algorithm to ...
    solve the ODE system
Stime = 10000;
set(configset, 'StopTime', Stime)

[success, variant out] = sbiosteadystate(M) % Assess if the ...
    system reaches a steady state
SD = sbiosimulate(M); % Numerical simulation of the system ...
    described by the parameters above
Plots = SD.DataNames(1:4);
Out = selectbyname(SD,Plots);
SBP = sbioplot(Out);

[Times,Vals,Names]= sbiosimulate(M);
```

A.6.2 Details of the simulated annealing algorithm in Subsection 4.6.1

In order to find both local and global minima of the function ϕ, a score of the difference between the simulation and the experimental values as defined by Eqn. (4.19), which is to be maximized by a simulated annealing algorithm, a first approach was made using the MatLab `simulannealbnd` function [1020], a function to retrieve the minimum of a function using a simulated annealing algorithm. However, because of its construction, `simulannealbnd` is unable to optimize a function of more than 10 real variables [1021]. An efficient method in this cases is to construct a Gaußian Martingale [1022–1024] such that the parameter values attempted at each step (or jump) are

$$\boldsymbol{k} = \boldsymbol{k}_0 + \mathcal{N}(\vec{0}, \boldsymbol{k}_{best}), \tag{A.2}$$

where $\boldsymbol{k}_0$ is the current best-performing simulation minimizing the $\phi(\boldsymbol{k})$ function as defined by Eqn. (4.19) and $\mathcal{N}$ is a multivariate normal distribution of mean zero and standard deviation k_i across each of the $\boldsymbol{k}_0$ dimensions.

When the random scattering algorithm finds a new $\boldsymbol{k}_0'$ value such that $\phi(\boldsymbol{k}_0') < \phi(\boldsymbol{k}_0)$, the algorithm moves from that point to a nearby one-dimensional local-

maxima point, following a 'jump algorithm' [764, 1025, 1026]. The algorithm stops when it finds a $\phi(\boldsymbol{k}) < 0.001$ value, since at that point the match is good to approximately a 99.9% precision. After running this simulation, I found a parameter set with the sought ratios, which is presented at Fig. 4.6 in the main text.

Every time MatLab restarts itself, it generates the same random numbers list, therefore at the beginning of each simulation we ran the MatLab `rng('shuffle');` command to alter the random numbers generator, since otherwise simulated annealing and Monte Carlo simulations perform an effective $n = 1$ simulation.

A.7 Proteomics

Collaborations with Shabaz Mohammed (SM) at Oxford Biochemistry and Benedikt Kessler (BK) at the Nuffield Department of Medicine measured fluctuations in the levels of proteins (SM) and phosphoproteins (BK) through the YMC by using Time-Of-Flight Mass Spectrometry (TOFMS) [1027–1029]. Understanding protein dynamics is a huge step into decoding the YMC regulatory dynamics. Unfortunately, the information content that can be drawn from these datasets is not entirely meaningful for the following reasons:

1. Number of available timepoints. Because of the experimental difficulties of TOFMS [1030, 1031], only a few timepoints were sequenced. This makes establishing peaks and causal relations hard to ascertain. Only one of the timepoints belongs to the OX phase, implying a possible miss of the protein peaks on that phase.

2. Data normalization. Fission yeast spiking [1032, 1033] was used to normalize protein levels cross-samples, since overall protein levels are expected to fluctuate during the YMC [112], but there were contamination issues with its sequencing.

3. Error margin. Variations in typical protein concentrations across the YMC were found to be approximately $\pm 20\%$, and specially for the least abundant ones, this is outnumbered by the spectrometer error range.

4. Proteins concentration range. There is a 1500-fold difference between the concentrations of the most and the least expressed proteins in yeast. This difference poses a risk for protein normalization based on overall proteinic content, since the normalization term will be effectively dominated by a handful of genes, and would have an unbearable error propagation.

Results

For these reasons, I have confined the proteomics analysis to the 100 most expressed proteins, looking for genes with wide oscillations in their associated proteins, since they are good candidates to be YMC regulators. The most relevant oscillation in protein levels corresponds to Tef1p. $TEF1$ is a translational elongation factor EF-1α [1034]. A GTP-bound active form [1035], it binds to and delivers aminoacylated tRNA to the A-site of ribosomes for elongation of nascent polypeptides [1036]. As we see in Figure A.6.a, $TEF1$ is very clearly an OX gene, while Tef1p is mostly abundant during the RB YMC phase, reaching a minimum during the OX phase, as seen in Figure A.6.b. The combination of $TEF1$ and Tef1p is coherent with the transcriptional oscillator presented in Chapter 4, on which genes highly expressed in one YMC phase lead to the associate protein being produced during the next phase, since it would be expected that the controlling circuit proteins be out of phase with the expression of the genes they regulate [1037, 1038].

Other genes with abundant associated protein and observed to cycle during the YMC are $HYP2$ (an early RC gene with a protein peak during OX) and $EFT1$ (an OX gene whose protein is not significantly cycling[1]). The fourth gene encountered is $YEF3$, Yef3p protein abundance being 10-fold smaller than Tef1p. It is remarkable that all of these abundant proteins correspond to genes that are related to the mRNA translation machinery [1039, 1040]. Some key regulators of transitions between phases of the yeast metabolic cycle appear to be Hap1 (RC), Hap4 (RB), Gcn4 (RC), Msn4 (Non-Cyc), Swi6 (RB), and Adr1 (RC) [472], all of them yeast Transcription Factors (TFs) [1041]. Unfortunately, no proteomics analysis can produce an exhaustive measurement of the proteome [675, 676]. Crucially, TFs, which are key proteins for the genetic regulatory machinery, have a very low concentration (few molecules per cell) [681,682], which means they are not detectable by proteomics analysis.

The final analysis in this Appendix is concerned with the time lag between RNA expression and protein concentration peaks represented in Figure A.7, aiming to elucidate biological functions associated with the temporal lags between RNA and protein by performing GO analyses in five different regions of Fig. A.7: the diagonal (no time lag), short ($t < 30$ min) time lag for RNA or protein, and long ($t > 30$ min) time lag. Unfortunately, none of these regions revealed significant (high p-value) GO (GOrilla) terms, except for the organonitrogen compound metabolic process and cy-

[1]The oscillation amplitude is smaller than the TOFMS error range.

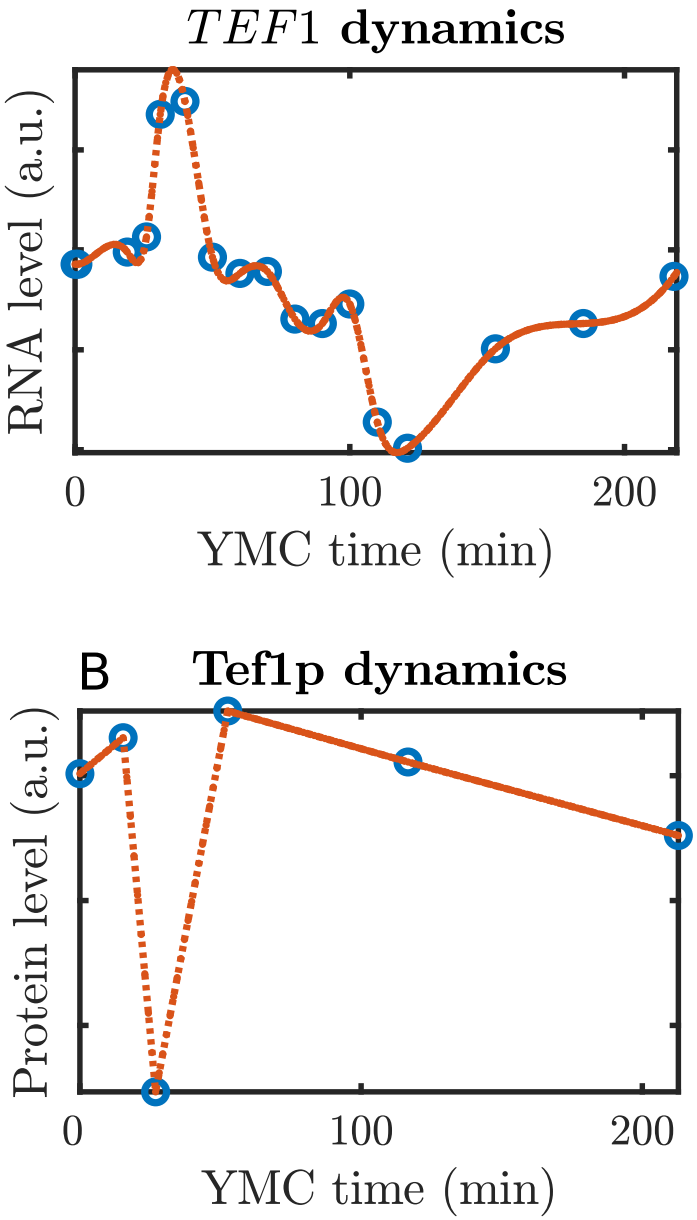


Figure A.6: *TEF*1 & Tef1p dynamics during the YMC. (Top) *TEF*1 RNA-seq for the [TSS-250 TTS] region (Bottom) Associate protein Tef1p (P02994) abundance. The OX phase drop in Tef1p level relies on only one measurement. Blue circles: sample measurements; red dotted lines: interpolation using cubic spline.

toplasmic translation (GO:1901564 & GO:0002181) terms obtained with $p = 0.03$ for the diagonal line genes.

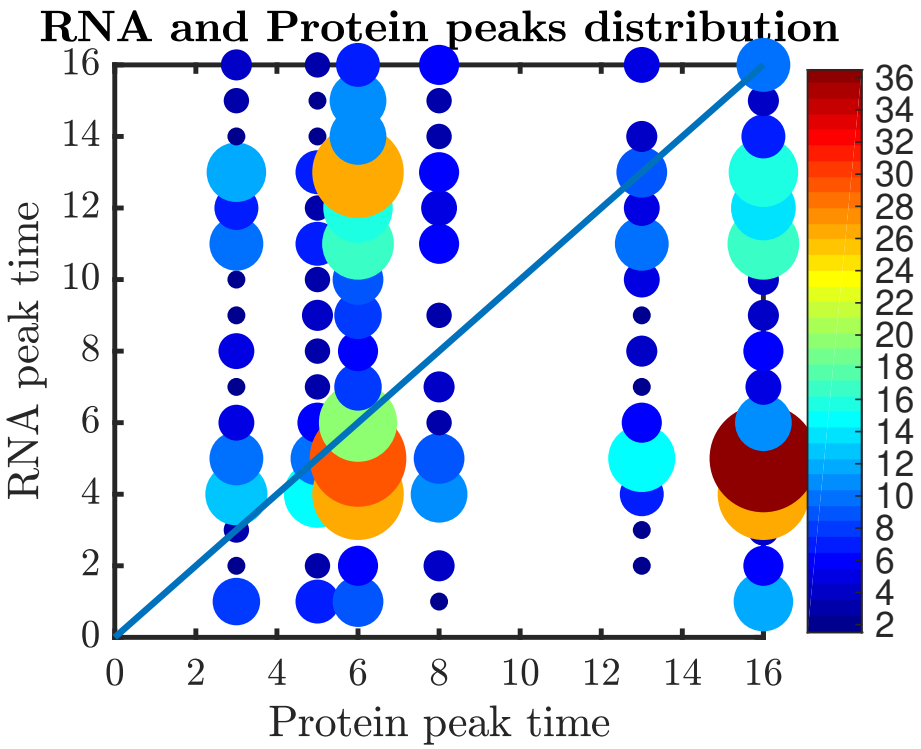


Figure A.7: RNA and protein peaks distribution in time in the YMC. A dot's area indicates the number of genes having their RNA and protein peaks at the corresponding temporal coordinates, which is also indicated by the dots color as per the colorbar. x-axis: point of the YMC with highest protein concentration measured in the proteomics experiment discussed in the main text (note that only six points were sampled, therefore the only possible values correspond to the y-axis equivalent of points 3, 5, 6, 8, 13, and 16); y-axis: point of the YMC with highest RNA concentration. Peaks are defined as the sample time point with maximum RNA expression or protein concentration. The diagonal line comprises the genes with RNA and protein co-occurrence.

A.8 Iron-Sulfur proteins, Fe^{2+} ion, and $xrn1\Delta$

This Appendix investigates the relationship between the iron-sulfur cluster proteins, focusing on the role of the Fe^{2+} ion, and the exonuclease $XRN1$. In the context of the YMC, it is observed that $xrn1\Delta$ mutant defects in the RNA degradation machinery [1042] imply a larger RNA content within the cells [1043], which are also considerably larger in size [1044] and exhibit a weaker cell growth [1045,1046]. These changes are believed to be caused by both TFs and chromatin remodellers [1047].

Iron metabolism in general and Fe^{2+} ions in particular are fundamental for mRNAs degradation [1048,1049], since the stability of transferrin receptor mRNA is regulated by iron availability in humans [1050]. Iron deficiency can affect both mRNA degradation in yeast [1051] and cell growth in plants [1052], suggesting this way a direct relationship between $xrn1\Delta$ and the mRNA degradation machinery.

A.8.1 Iron proteins

The Mellor group has found [1053] that, when $FeSO_4$ concentration in the media is largely increased[2], the YMC period is increased from 2h to 3h. This result suggests that the Fe^{2+} ion is playing a crucial part in the yeast metabolic regulation. Therefore, I assess if the set of genes associated with iron ions (either Fe^{2+} or Fe^{3+}) exhibits some form of specific behavior during the YMC.

A.8.2 Why Fe^{2+} and not SO_4^{2-}?

In an aqueous medium (as in the chemostat), we have [1054]:

$$FeSO_4 \leftrightarrow Fe^{2+} + SO_4^{2-}$$

The YMC period increase described in A.8.1 is caused by a change in $FeSO_4$ concentration, so *a priori* it could be either Fe^{2+} or SO_4^{2-} causing the increase.

The SO_4^{2-} ion group interactions of $FeSO_4$ in an aqueous medium are analogous to the SO_4^{2-} group interactions in $(NH_4)_2SO_4$ [1055]. In the YMC media, $[(NH_4)_2SO_4] \gg [FeSO_4 \cdot 7H_2O]$[3]; if SO_4^{2-} was to be the relevant ion, then the 18-fold increase in $[FeSO_4]$ would virtually have no effect in the YMC period, since it would

[2] By an 18-fold, the $FeSO_4 \cdot 7H_2O$ concentration is increased from 0.02g/L to 0.36g/L.

[3] $[(NH_4)_2SO_4] = 5$g/L, while $[FeSO_4 \cdot 7H_2O]$ changes from 0.02g/L to 0.36g/L.

increase overall $[SO_4^{2-}]$ by less than 5%.

I am interested in knowing if a similar YMC period increase can be achieved via a large increase in the concentration of other compounds in the YMC media, like $ZnSO_4$ or $CuSO_4$[4], exhibiting similar functional ions [1056–1059]. I do not predict a YMC period increase in this case, since $FeSO_4$ has been shown to exhibit remarkably different interactions to $ZnSO_4$ or $CuSO_4$ [1060], and transcription of the yeast iron regulon does not respond directly to iron but rather to iron-sulfur cluster biosynthesis [1061].

Iron-sulfur cluster proteins are encoded by essential genes and perform a wide variety of metabolic functions in yeast [1062, Table 1], actively regulating gene expression via some TFs like SoxR, IRP, and IscR. [1063, Table 1]. It is remarkable that the FNR (fumarate and nitrate reduction) regulon $[4Fe\text{-}4S]^{2+}$ cluster is a global regulator that performs oxygen sensing under oxygen limiting conditions in *E. coli* [204].

Another Fe-S cluster relevant for YMC maintenance could be $NFU1$. Yeast strains that lack Nfu1p exhibit loss of Fe-S protein activities [1064, 1065], revealing a translational regulator dependent on the presence of a $[4Fe–4S]^{2+}$ cluster [1066] via alternative splicing [1067].

These results are linked to the oxygen sensing conclusions in Subsection 1.5.1 via the NADPH and its oscillations during the YMC [1068–1070], since NADPH has been shown to be coordinated with both the $[4Fe–4S]^{2+}$ proteins [204] and oxygen sensing [210].

A.8.3 Iron genes analysis

According to the Saccharomyces Genome Database (SGD) [1071–1076], the genes in Table A.1 are associated with either the Fe^{2+} or Fe^{3+} ions, and their YMC dynamics expression is depicted in Figure A.8.

[4] A full description of the media composition is given in Boeke *et. al.* [127, Suppl. Mat.].

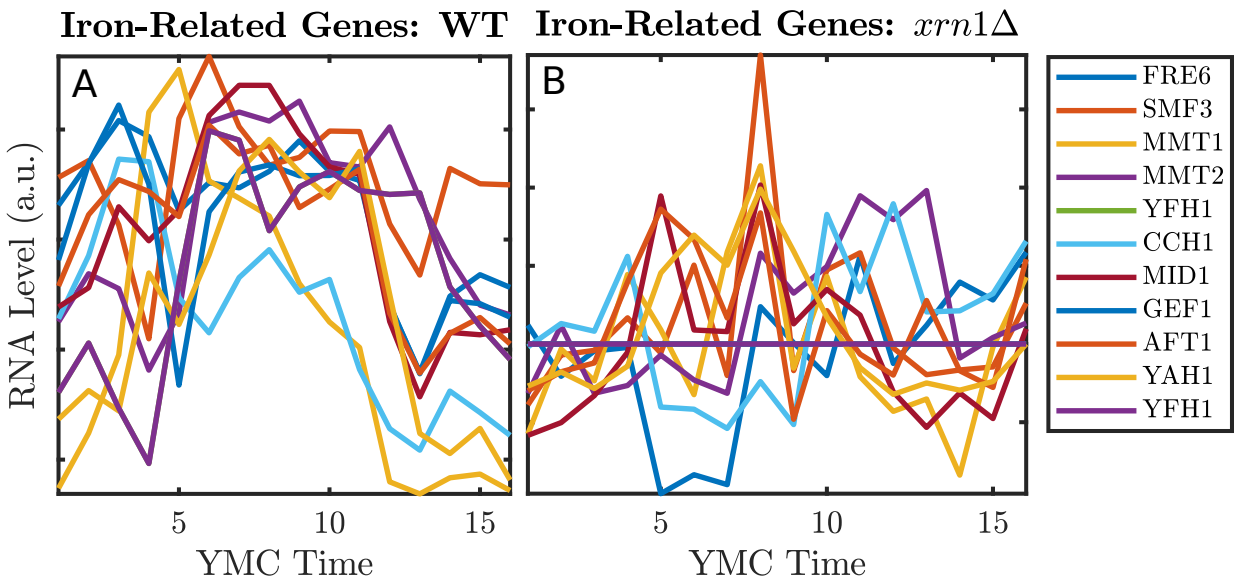


Figure A.8: (a) WT YMC dynamics of genes whose phenotype is associated with the iron ions Fe^{2+} and Fe^{3+} as per Table A.1. A large number of them are RB genes. (b) *xrn*1Δ YMC expression levels for the same group of genes. x-axis: time in the YMC in arbitrary units; y-axis: gene expression in arbitrary units.

Table A.1: Genes whose phenotype is associated with iron ions Fe^{2+} or Fe^{3+}

Fe^{2+}

Gene	Reference
FRE6	Singh A, et al. (2007) [1077]
SMF3	Singh A, et al. (2007) [1077]
MMT1	Li L and Kaplan J (1997) [1078]
MMT2	Li L and Kaplan J (1997) [1078]
YFH1	Foury F, et al. (2007) [1079]
CCH1	Peiter E, et al. (2005) [1080]
MID1	Peiter E, et al. (2005) [1080]
GEF1	Greene JR, et al. (1993) [1081]

Fe^{3+}

Gene	Reference
AFT1	Batista L, et al. (2013) [1082]
YAH1	Miao R, et al. (2008) [1083]
YFH1	Karthikeyan G, et al. (2002) [1084]

Bibliography

[1] J. E. Baldwin and H. Krebs, "The evolution of metabolic cycles," *Nature*, vol. 291, pp. 381–382, jun 1981.

[2] D. Lloyd and D. B. Murray, "Ultradian metronome: Timekeeper for orchestration of cellular coherence," *Trends in Biochemical Sciences*, vol. 30, no. 7, pp. 373–377, 2005.

[3] B. P. Tu, A. Kudlicki, M. Rowicka, and S. L. McKnight, "Logic of the Yeast Metabolic Cycle: Temporal Compartmentalization of Cellular Processes," *Science*, vol. 310, pp. 1152–1158, 2005.

[4] M. Sharma and P. Kumar, "Chemical oscillations," *Resonance*, vol. 11, pp. 43–50, aug 2006.

[5] A. M. Zhabotinsky, "A history of chemical oscillations and waves," *Chaos*, vol. 1, pp. 379–386, dec 1991.

[6] A. Goldbeter and M. Berridge, *Biochemical oscillations and cellular rhythms: The molecular bases of periodic and chaotic behaviour*. Cambridge: Cambridge University Press, 1996.

[7] B. P. Belousov and A. Zhabotinsk, "A periodic reaction and its mechanism," *Biofizika*, vol. 9, pp. 306–311, 1964.

[8] P. Valery, G. Vilmos, M. Honathan, and S. Jenneth, "Controlling chaos in the Belousov-Zhabotinsky reaction," *Nature*, vol. 361, p. 240, jan 1993.

[9] G. Nicolis and I. Prigogine, "Fluctuations in nonequilibrium systems," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 68, pp. 2102–7, sep 1971.

[10] R. M. May, "Biological Populations with Nonoverlapping Generations: Stable Points, Stable Cycles, and Chaos," *Science*, vol. 186, pp. 645–647, nov 1977.

[11] A. T. Winfree, *The Geometry of Biological Time*, vol. 12. Springer, 2001.

[12] M. Rosenlicht, "Toroidal Algebraic Groups," *Proceedings of the American Mathematical Society*, vol. 12, pp. 984–988, dec 1961.

[13] W. Vance and J. Ross, "Entrainment, phase resetting, and quenching of chemical oscillations," *The Journal of Chemical Physics*, vol. 103, pp. 2472–2481, aug 1995.

[14] J. Wolf and R. Heinrich, "Dynamics of two-component biochemical systems in interacting cells; Synchronization and desynchronization of oscillations and multiple steady states," *BioSystems*, vol. 43, pp. 1–24, may 1997.

[15] M. J. Hauser, U. Kummer, A. Z. Larsen, and L. F. Olsen, "Oscillatory dynamics protect enzymes and possibly cells against toxic substances," *Faraday Discussions*, vol. 120, pp. 215–227, jan 2001.

[16] J. Aldridge and E. K. Pye, "Cell density dependence of oscillatory metabolism," *Nature*, vol. 259, pp. 670–671, feb 1976.

[17] M. H. Vitaterna, J. S. Takahashi, and F. W. Turek, "Overview of circadian rhythms," *Alcohol research & health: the journal of the National Institute on Alcohol Abuse and Alcoholism*, vol. 25, no. 2, pp. 85–93, 2001.

[18] R. G. Foster and L. Kreitzmann, *Rhythms of life: the biological clocks that control the daily lives of every living thin*. Profile, 2004.

[19] D. Lloyd and D. B. Murray, "The temporal architecture of eukaryotic growth," *FEBS Letters*, vol. 580, pp. 2830–2835, may 2006.

[20] J. Leland N Edmunds, *Cell Cycle Clocks*. Routledge/Taylor & Francis Group, 1984.

[21] J. Rutter, M. Reick, and S. L. McKnight, "Metabolism and the Control of Circadian Rhythms," *Annual Review of Biochemistry*, vol. 71, no. 1, pp. 307–331, 2002.

[22] J. Zantke, T. Ishikawa-fujiwara, E. Arboleda, C. Lohs, K. Schipany, and K. Tessmar-Raible, "Circadian and Circalunar Clock Interactions in a Marine Annelid," *Cell Reports*, vol. 5, no. 1, pp. 99–113, 2013.

[23] D. Wilcockson and L. Zhang, "Circatidal clocks," *Current Biology*, vol. 18, pp. R753–R755, sep 2008.

[24] E. Naylor, "Crab clockwork: The case for interactive circatidal and circadian oscillators controlling rhythmic locomotor activity of Carcinus maenas," *Chronobiology International*, vol. 13, pp. 153–161, jan 1996.

[25] M. Rosbash, "The Implications of Multiple Circadian," *PLoS Biology*, vol. 7, no. 3, pp. 421–425, 2009.

[26] C. G. Warman and E. Naylor, "Evidence for multiple, cue-specific circatidal clocks in the shore crab Carcinus maenas," *Journal of Experimental Marine Biology and Ecology*, vol. 189, pp. 93–101, jun 1995.

[27] Z. Deng, S. Arsenault, C. Caranica, J. Griffith, T. Zhu, A. Al-Omari, H.-B. Schüttler, J. Arnold, and L. Mao, "Synchronizing stochastic circadian oscillators in single cells of Neurospora crassa," *Scientific Reports*, vol. 6, no. 1, p. 35828, 2016.

[28] V. G. Bruce, "Environmental entrainment of circadian rhythms," in *Cold Spring Harbor Symposia on Quantitative Biology*, vol. 25, pp. 29–48, Cold Spring Harbor Laboratory Press, jan 1960.

[29] K. A. Thomas, R. L. Burr, and S. Spieker, "Light and maternal influence in the entrainment of activity circadian rhythm in infants 4–12 weeks of age," *Sleep and Biological Rhythms*, vol. 14, pp. 249–255, jul 2016.

[30] J. K. Paulose, E. B. Rucker, and V. M. Cassone, "Toward the Beginning of Time: Circadian Rhythms in Metabolism Precede Rhythms in Clock Gene Expression in Mouse Embryonic Stem Cells," *PLoS ONE*, vol. 7, p. e49555, nov 2012.

[31] F. Kippert and P. Hunt, "Ultradian clocks in eukaryotic microbes: from behavioural observation to functional genomics," *BioEssays*, vol. 22, no. 1, pp. 16–22, 2000.

[32] V. Dvornyk, O. Vinogradova, and E. Nevo, "Origin and evolution of circadian clock genes in prokaryotes," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 100, pp. 2495–2500, apr 2003.

[33] E. Tauber, K. S. Last, P. J. Olive, and C. P. Kyriacou, "Clock Gene Evolution and Functional Divergence," *Journal of Biological Rhythms*, vol. 19, pp. 445–458, oct 2004.

[34] P. Janich, A. B. Arpat, V. Castelo-Szekely, M. Lopes, and D. Gatfield, "Ribosome profiling reveals the rhythmic liver translatome and circadian clock regulation by upstream open reading frames," *Genome Research*, vol. 25, pp. 1848–1859, dec 2015.

[35] J. S. Menet, J. Rodriguez, K. C. Abruzzi, and M. Rosbash, "Nascent-Seq reveals novel features of mouse circadian transcriptional regulation," *eLife*, vol. 2012, no. 1, pp. 1–25, 2012.

[36] P. Hardin, J. Hall, and M. Rosbash, "Feedback of the Drosophila period gene product on circadian cycling of its messenger RNA levels," *Nature*, vol. 343, no. 6258, pp. 536–40, 1990.

[37] U. Schibler and F. Naef, "Cellular oscillators: rhythmic gene expression and metabolism," *Current opinion in cell biology*, vol. 17, no. 2, pp. 223–229, 2005.

[38] H. Wijnen and M. W. Young, "Interplay of circadian clocks and metabolic rhythms," *Annu. Rev. Genet.*, vol. 40, pp. 409–448, 2006.

[39] P. Smolen, D. A. Baxter, and J. H. Byrne, "Modeling circadian oscillations with interlocking positive and negative feedback loops," *The Journal of neuroscience: the official journal of the Society for Neuroscience*, vol. 21, pp. 6644–6656, sep 2001.

[40] S. Becker-Weimann, J. Wolf, H. Herzel, and A. Kramer, "Modeling feedback loops of the mammalian circadian oscillator," *Biophysical Journal*, vol. 87, pp. 3023–3034, nov 2004.

[41] T. Pavlidis, "Populations of interacting oscillators and circadian rhythms," *Journal of theoretical biology*, vol. 22, no. 3, pp. 418–436, 1969.

[42] R. Heinrich, S. M. Rapoport, and T. A. Rapoport, "Metabolic regulation and mathematical models," *Progress in Biophysics and Molecular Biology*, vol. 32, no. C, pp. 1–82, 1978.

[43] R. Refinetti, G. C. Lissen, and F. Halberg, "Procedures for numerical analysis of circadian rhythms," *Biol Rhythm Res*, vol. 38, no. 4, pp. 275–325, 2007.

[44] H. Kacser and J. A. Burns, "The control of flux," *Symposia of the Society for Experimental Biology*, vol. 27, pp. 65–104, 1973.

[45] H. V. Westerhoff, "Thermodynamics and control of proton motive free-energy transduction," *Biomed Biochim Acta*, vol. 44, no. 6, pp. 929–941, 1985.

[46] S. Ray and A. B. Reddy, "Cross-talk between circadian clocks, sleep-wake cycles, and metabolic networks: Dispelling the darkness," *BioEssays*, vol. 38, no. 4, pp. 394–405, 2016.

[47] P. E. Hardin, J. C. Hallt, and M. Rosbash, "Circadian oscillations in period gene mRNA levels are transcriptionally regulated," *Proc Natl Acad Sci U S A*, vol. 89, no. December, pp. 11711–11715, 1992.

[48] H. Zeng, Z. Qian, M. P. Myers, and M. Rosbash, "A light-entrainment mechanism for the drosophila circadian clock," *Nature*, vol. 380, pp. 129–135, mar 1996.

[49] G. Asher, D. Gatfield, M. Stratmann, H. Reinke, C. Dibner, F. Kreppel, R. Mostoslavsky, F. W. Alt, and U. Schibler, "SIRT1 Regulates Circadian Clock Gene Expression through PER2 Deacetylation," *Cell*, vol. 134, pp. 317–328, jul 2008.

[50] Z. Eelderink-Chen, G. Mazzotta, M. Sturre, J. Bosman, T. Roenneberg, and M. Merrow, "A circadian clock in Saccharomyces cerevisiae," *Proceedings of the National Academy of Sciences*, vol. 107, pp. 2043–2047, feb 2010.

[51] Z. Eelderink-Chen, M. Olmedo, J. Bosman, and M. Merrow, "Using circadian entrainment to find cryptic clocks," *Methods in Enzymology*, vol. 551, pp. 73–93, jan 2015.

[52] M. Bier, B. Teusink, B. N. Kholodenko, and H. V. Westerhoff, "Control analysis of glycolytic oscillations," *Biophysical Chemistry*, vol. 62, pp. 15–24, nov 1996.

[53] I. M. de la Fuente, L. Martinez, and J. Veguillas, "Dynamic behavior in glycolytic oscillations with phase shifts," *BioSystems*, vol. 35, pp. 1–13, jan 1995.

[54] Y. Termonia and J. Ross, "Oscillations and control features in glycolysis: analysis of resonance effects," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 78, pp. 3563–3566, jun 1981.

[55] B. Teusink, B. M. Bakker, and H. V. Westerhoff, "Control of frequency and amplitudes is shared by all enzymes in three models for yeast glycolytic oscillations," *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, vol. 1275, pp. 204–212, jul 1996.

[56] J. B. Robertson, C. R. Davis, and C. H. Johnson, "Visible light alters yeast metabolic rhythms by inhibiting respiration," *Proceedings of the National Academy of Sciences*, vol. 110, pp. 21130–21135, dec 2013.

[57] P. Novick and D. Botstein, "Phenotypic analysis of temperature-sensitive yeast actin mutants," *Cell*, vol. 40, pp. 405–416, feb 1985.

[58] M. J. Miller, N. H. Xuong, and E. P. Geiduschek, "A response of protein synthesis to temperature shift in the yeast Saccharomyces cerevisiae.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 76, pp. 5222–5225, oct 1979.

[59] M. Zakhartsev and M. Reuss, "Cell size and morphological properties of yeast Saccharomyces cerevisiae in relation to growth temperature.," *FEMS yeast research*, vol. 18, sep 2018.

[60] F. N. Arroyo López, S. Orlić, A. Querol, and E. Barrio, "Effects of temperature, pH and sugar concentration on the growth parameters of Saccharomyces cerevisiae, S. kudriavzevii and their interspecific hybrid," *International Journal of Food Microbiology*, vol. 131, pp. 120–127, may 2009.

[61] K. Oishi, N. Ohkura, K. Kadota, M. Kasamatsu, K. Shibusawa, J. Matsuda, K. Machida, S. Horie, and N. Ishida, "Clock mutation affects circadian regulation of circulating blood cells," *Journal of Circadian Rhythms*, vol. 4, p. 13, oct 2006.

[62] J. S. O-Neill and A. B. Reddy, "Circadian clocks in human red blood cells," *Nature*, vol. 469, pp. 498–504, jan 2011.

[63] S. P. Klinken, "Cells in focus Red blood cells," *The International Journal of Biochemistry & Cell Biology*, vol. 34, pp. 1513–1518, 2002.

[64] S. Kabanova, P. Kleinbongard, J. Volkmer, B. Andrée, M. Kelm, and T. W. Jax, "Gene expression analysis of human red blood cells," *International Journal of Medical Sciences*, vol. 6, no. 4, pp. 156–159, 2009.

[65] B. C. Goodwin and M. H. Cohen, "A phase-shift model for the spatial and temporal organization of developing systems," *Journal of Theoretical Biology*, vol. 25, pp. 49–107, oct 1969.

[66] R. K. Poole, D. Lloyd, and R. B. Kemp, "Respiratory Oscillations and Heat Evolution in Synchronously Dividing Cultures of the Fission Yeast Schizosaccharomyces pombe 972h-," *Microbiology*, vol. 77, pp. 209–220, jul 1973.

[67] C. Edwards, M. Statham, and D. Lloyd, "The Preparation of Large-scale Synchronous Cultures of the Trypanosomatid, Crithidia fasciculata, by Cell-size Selection: Changes in Respiration and Adenylate Charge through the Cell-cycle," *Journal of General Microbiology*, vol. 88, pp. 141–152, may 1975.

[68] A. Chagla and A. Griffiths, "Synchronous Cultures of Acanthamoeba castellanii and Their Use in the Study of Encystation," *Journal of General Microbiology*, vol. 108, no. 1, pp. 39–43, 1978.

[69] S. W. Edwards and D. Lloyd, "Oscillations of Respiration and Adenine Nucleotides in Synchronous Cultures of Acanthamoeba castellanii: Mitochondrial Respiratory Control in vivo," *Microbiology*, vol. 108, no. 2, pp. 197–204, 1978.

[70] C. L. Bashford, B. Chance, D. Lloyd, and R. K. Poole, "Oscillations of redox states in synchronously dividing cultures of Acanthamoeba castellanii and Schizosaccharomyces pombe," *Biophysical Journal*, vol. 29, pp. 1–11, jan 1980.

[71] C. Woffendin and A. J. Griffiths, "Changes in respiratory activity and total protein during synchronous growth of Dictyostelium discoideum," *FEMS Microbiol. Lett.*, vol. 29, pp. 263–267, sep 1985.

[72] W. J. Ingledew and R. K. Poole, "The respiratory chains of Escherichia coli," *Microbiol Rev*, vol. 48, no. 3, pp. 222–271, 1984.

[73] H. Degn and D. E. Harrisson, "Theory of oscillations of respiration rate in continuous culture of Klebsiella aerogenes," *Journal of Theoretical Biology*, vol. 22, pp. 238–248, feb 1969.

[74] D. Lloyd, S. W. Edwards, and J. C. Fry, "Temperature-compensated oscillations in respiration and cellular protein content in synchronous cultures of Acanthamoeba castellanii," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 79, no. 12, pp. 3785–8, 1982.

[75] J. Kader and D. Lloyd, "Respiratory oscillations and heat evolution in synchronous cultures of Candida utilis," *Journal of General Microbiology*, vol. 114, pp. 455–461, oct 1979.

[76] R. J. Cho, M. J. Campbell, E. A. Winzeler, L. Steinmetz, A. Conway, L. Wodicka, T. G. Wolfsberg, A. E. Gabrielian, D. Landsman, D. J. Lockhart, and R. W. Davis, "A genome-wide transcriptional analysis of the mitotic cell cycle," *Molecular Cell*, vol. 2, pp. 65–73, jul 1998.

[77] P. Spellman, G. Sherlock, M. Zhang, V. Iyer, K. Anders, M. Eisen, P. Brown, D. Botstein, and B. Futcher, "Comprehensive identification of cell cycle-regulated genes of the yeast Saccharomyces cerevisiae by microarray hybridization," *Mol. Biol. Cell*, vol. 9, pp. 3273–3297, dec 1998.

[78] S. L. Harmer, J. B. Hogenesch, M. Straume, H. S. Chang, B. Han, T. Zhu, X. Wang, J. A. Kreps, and S. A. Kay, "Orchestrated transcription of key pathways in Arabidopsis by the circadian clock," *Science*, vol. 290, pp. 2110–2113, dec 2000.

[79] M. J. McDonald and M. Rosbash, "Microarray analysis and organization of circadian gene expression in Drosophila," *Cell*, vol. 107, pp. 567–578, nov 2001.

[80] Y. Lin, M. Han, B. Shimada, L. Wang, T. M. Gibler, A. Amarakone, T. A. Awad, G. D. Stormo, R. N. Van Gelder, and P. H. Taghert, "Influence of the period-dependent circadian clock on diurnal, circadian, and aperiodic gene expression in Drosophila melanogaster," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, pp. 9562–7, jul 2002.

[81] K.-F. Storch, O. Lipan, I. Leykin, N. Viswanathan, F. C. Davis, W. H. Wong, and C. J. Weitz, "Extensive and divergent circadian gene expression in liver and heart," *Nature*, vol. 417, pp. 78–83, may 2002.

[82] S. Panda, M. P. Antoch, B. H. Miller, A. I. Su, A. B. Schook, M. Straume, P. G. Schultz, S. A. Kay, J. S. Takahashi, and J. B. Hogenesch, "Coordinated transcription of key pathways in the mouse by the circadian clock," *Cell*, vol. 109, no. 3, pp. 307–320, 2002.

[83] R. A. Akhtar, A. B. Reddy, E. S. Maywood, J. D. Clayton, V. M. King, A. G. Smith, T. W. Gant, M. H. Hastings, and C. P. Kyriacou, "Circadian cycling of the mouse liver transcriptome, as revealed by cDNA microarray, is driven

by the suprachiasmatic nucleus," *Current Biology*, vol. 12, pp. 540–550, apr 2002.

[84] G. E. Duffield, J. D. Best, B. H. Meurers, A. Bittner, J. J. Loros, and J. C. Dunlap, "Circadian programs of transcriptional activation, signaling, and protein turnover revealed by microarray analysis of mammalian cells," *Current Biology*, vol. 12, pp. 551–557, apr 2002.

[85] J. Liu, A. Prindle, J. Humphries, M. Gabalda-Sagarra, M. Asally, D. Y. D. Lee, S. Ly, J. Garcia-Ojalvo, and G. M. Süel, "Metabolic co-dependence gives rise to collective oscillations within biofilms," *Nature*, vol. 523, no. 7562, pp. 550–554, 2015.

[86] J. J. Tyson, K. C. Chen, and B. Novak, "Sniffers, buzzers, toggles and blinkers: Dynamics of regulatory and signaling pathways in the cell," *Current Opinion in Cell Biology*, vol. 15, pp. 221–231, apr 2003.

[87] U. S. Bhalla and R. Iyengar, "Emergent properties of networks of biological signaling pathways," *Science*, vol. 283, pp. 381–387, jan 1999.

[88] B. N. Kholodenko, G. C. Brown, and J. B. Hoek, "Diffusion control of protein phosphorylation in signal transduction pathways.," *Biochemical Journal*, vol. 350 Pt 3, pp. 901–907, sep 2000.

[89] G. C. Brown and B. N. Kholodenko, "Spatial gradients of cellular phosphoproteins," *FEBS Letters*, vol. 457, pp. 452–454, sep 1999.

[90] B. N. Kholodenko, "MAP kinase cascade signaling and endocytic trafficking: A marriage of convenience?," apr 2002.

[91] P. Niethammer, P. Bastiaens, and E. Karsenti, "Stathmin-Tubutin Interaction Gradients in Motile and Mitotic Cells," *Science*, vol. 303, pp. 1862–1866, mar 2004.

[92] B. Goldstein, J. R. Faeder, and W. S. Hlavacek, "Mathematical and computational models of immune-receptor signalling," *Nature Reviews Immunology*, vol. 4, pp. 445–456, jun 2004.

[93] D. Bray, "Reductionism for biochemists: How to survive the protein jungle," *Trends in Biochemical Sciences*, vol. 22, pp. 325–326, sep 1997.

[94] A. Ghosh and B. Chance, "Oscillations of glycolytic intermediates in yeast cells," *Biochemical and Biophysical Research Communications*, vol. 16, no. 2, pp. 174–181, 1964.

[95] P. Richard, "The rhythm of yeast," *FEMS Microbiology Reviews*, vol. 27, pp. 547–557, oct 2003.

[96] B. Chance, R. Estabrook, and A. Ghosh, "Damped sinusoidal oscillations of cytoplasmic reduced pyridine nucleotide in yeast cells," *Proc Natl Acad Sci U S A*, vol. 51, pp. 1244–1251, 1964.

[97] B. Chance, B. Schoener, and S. Elsaesser, "Control of the waveform of oscillations of the reduced pyridine nucleotide level in a cell-free extract.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 52, pp. 337–41, aug 1964.

[98] F. A. Hommes, "Oscillatory Reductions of Pyridine Nucleotides Yeast during Anaerobic Glycolysis in Brewers' Yeast," *Archives of Biochemistry and Biophysics*, vol. 108, pp. 36–46, 1964.

[99] A. Betz and B. Chance, "Phase Relationship of Glycolytic Intermediates in Yeast Cells With Oscillatory Metabolic Control," *Archives of biochemistry and biophysics*, vol. 109, pp. 585–594, 1965.

[100] H. K. Meyenburg, "Energetics of the Budding Cycle of Saccharomyces cerevisiae during Ghicose Limited Aerobic Growth," *Archiv für Mikrobiologie*, vol. 303, no. 66, pp. 289–303, 1969.

[101] M. Bier, B. M. Bakker, and H. V. Westerhoff, "How Yeast Cells Synchronize their Glycolytic Oscillations: A Perturbation Analytic Treatment," *Biophysical Journal*, vol. 78, no. 3, pp. 1087–1093, 2000.

[102] F. A. Chandra, G. Buzi, and J. C. Doyle, "Glycolytic oscillations and limits on robust efficiency," *Science*, vol. 333, no. 6039, pp. 187–192, 2011.

[103] A. Kloster and L. F. Olsen, "Oscillations in glycolysis in Saccharomyces cerevisiae: The role of autocatalysis and intracellular ATPase activity," *Biophysical Chemistry*, vol. 165-166, pp. 39–47, 2012.

[104] D. Porro, E. Martegani, B. M. Ranzi, and L. Alberghina, "Oscillations in continuous cultures of budding yeast: A segregated parameter analysis," *Biotechnology and Bioengineering*, vol. 32, no. 4, pp. 411–417, 1988.

[105] A. D. Satroutdinov, H. Kuriyama, and H. Kobayashi, "Oscillatory metabolism of Saccharomyces cerevisiae in continuous culture," *FEMS Microbiology Letters*, vol. 98, pp. 261–267, nov 1992.

[106] M. Keulers, T. Suzuki, A. D. Satroutdinov, and H. Kuriyama, "Autonomous metabolic oscillation in continuous culture of Saccharomyces cerevisiae grown on ethanol," *FEMS Microbiology Letters*, vol. 142, pp. 253–258, sep 1996.

[107] S. Danø, P. G. Sørensen, and F. Hynne, "Sustained oscillations in living cells," *Nature*, vol. 402, pp. 320–322, nov 1999.

[108] C. M. Li and R. R. Klevecz, "A rapid genome-scale response of the transcriptional oscillator to perturbation reveals a period-doubling path to phenotypic change," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103, no. 44, pp. 16254–9, 2006.

[109] R. R. Klevecz, J. Bolen, G. Forrest, and D. B. Murray, "A genomewide oscillation in transcription gates DNA replication and cell cycle," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 5, pp. 1200–1205, 2004.

[110] A. Novick and L. Szilard, "Experiments with the Chemostat on Spontaneous Mutations of Bacteria," *Source: Proceedings of the National Academy of Sciences of the United States of America*, vol. 36, no. 12, pp. 708–719, 1950.

[111] Orbit Biotech, "Chemostat (a continous culture)." Online resource. http://orbitbiotech.com/chemostat-a-continous-culture-culture-microbial-growth-chemostat/. Date accessed: 01-04-2020.

[112] B. P. Tu, R. E. Mohler, J. C. Liu, K. M. Dombek, E. T. Young, R. E. Synovec, and S. L. McKnight, "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*, vol. 104, no. 43, pp. 16886–16891, 2007.

[113] R. E. Mohler, B. P. Tu, K. M. Dombek, J. C. Hoggard, E. T. Young, and R. E. Synovec, "Identification and evaluation of cycling yeast metabolites in two-dimensional comprehensive gas chromatography-time-of-flight-mass spectrometry data," *Journal of Chromatography A*, vol. 1186, no. 1-2, pp. 401–411, 2008.

[114] B. P. Tu and S. L. McKnight, "Metabolic cycles as an underlying basis of biological oscillations," *Nature Reviews Molecular Cell Biology*, vol. 7, no. 9, pp. 696–701, 2006.

[115] B. P. Tu and S. L. Mcknight, "The yeast metabolic cycle: Insights into the life of a eukaryotic cell," *Cold Spring Harbor Symposia on Quantitative Biology*, vol. 72, pp. 339–343, 2007.

[116] L. Cai and B. P. Tu, "Acetyl-CoA drives the transcriptional growth program in yeast," *Cell Cycle*, vol. 10, no. 18, pp. 3045–3046, 2011.

[117] L. Cai and B. P. Tu, "On acetyl-CoA as a gauge of cellular metabolic state," *Cold Spring Harbor Symposia on Quantitative Biology*, vol. 76, pp. 195–202, 2011.

[118] L. Cai, B. M. Sutter, B. Li, and B. P. Tu, "Acetyl-CoA Induces Cell Growth and Proliferation by Promoting the Acetylation of Histones at Growth Genes," *Molecular Cell*, vol. 42, no. 4, pp. 426–437, 2011.

[119] H. Y. Sohn, D. B. Murray, and H. Kuriyama, "Ultradian oscillation of Saccharomyces cerevisiae during aerobic continuous culture: Hydrogen sulphide mediates population synchrony," *Yeast*, vol. 16, no. 13, pp. 1185–1190, 2000.

[120] D. B. Murray, R. R. Klevecz, and D. Lloyd, "Generation and maintenance of synchrony in Saccharomyces cerevisiae continuous culture," *Experimental Cell Research*, vol. 287, no. 1, pp. 10–15, 2003.

[121] Z. Palkova, "Ammonia Pulses and Metabolic Oscillations Guide Yeast Colony Development," *Molecular Biology of the Cell*, vol. 13, pp. 3901–3914, nov 2002.

[122] G. Z. Wang, S. L. Hickey, L. Shi, H. C. Huang, P. Nakashe, N. Koike, B. P. Tu, J. S. Takahashi, and G. Konopka, "Cycling Transcriptional Networks Optimize Energy Utilization on a Genome Scale," *Cell Reports*, vol. 13, no. 9, pp. 1868–1880, 2015.

[123] Mettler Toledo, "Dissolved oxygen measurement for cell culture and fermentation." White paper on eliminating noisy oxygen measurements in fermentation and cell culture. Online resource. https://www.mt.com/us/en/home/library/white-papers/process-analytics/improved-oxygen-measurement.html. Date accessed: 26-03-2020.

[124] J. Feltham and S. Xi, "Glucose effect on the YMC period," 2018. Personal communication concerning unpublished data.

[125] L. E. Salgado, D. B. D. Murray, D. Lloyd, E. Salgado, D. B. D. Murray, and D. Lloyd, "Some Antidepressant Agents (Li+, Monoamine Oxidase Type A Inhibitors) Perturb the Ultradian Clock in Saccharomyces cerevisiae," *Biological rhythm research*, vol. 33, no. 3, pp. 351–361, 2002.

[126] H. Reinke and D. Gatfield, "Genome-wide oscillation of transcription in yeast," *Trends in Biochemical Sciences*, vol. 31, no. 4, pp. 189–191, 2006.

[127] Z. Kuang, L. Cai, X. Zhang, H. Ji, B. P. Tu, and J. D. Boeke, "High-temporal-resolution view of transcription and chromatin states across distinct metabolic states in budding yeast," *Nature Structural and Molecular Biology*, vol. 21, no. 10, pp. 854–863, 2014.

[128] L. Cai and B. P. Tu, "Driving the cell cycle through metabolism," *Annual review of cell and developmental biology*, vol. 28, pp. 59–87, 2012.

[129] N. Slavov, D. Botstein, and T. D. Fox, "Coupling among growth rate response, metabolic cycle, and cell division cycle in yeast," *Molecular biology of the cell*, vol. 22, pp. 1997–2009, 2011.

[130] B. Futcher, "Metabolic cycle, cell cycle, and the finishing kick to start," *Genome Biology*, vol. 7, p. 107, apr 2006.

[131] D. Müller, S. Exler, L. Aguilera-Vázquez, E. Guerrero-Martín, and M. Reuss, "Cyclic AMP mediates the cell cycle dynamics of energy metabolism in Saccharomyces cerevisiae," *Yeast*, vol. 20, pp. 351–367, mar 2003.

[132] Z. Xu and K. Tsurugi, "A potential mechanism of energy-metabolism oscillation in an aerobic chemostat culture of the yeast Saccharomyces cerevisiae," *FEBS Journal*, vol. 273, pp. 1696–1709, apr 2006.

[133] M. Rowicka, A. Kudlicki, B. P. Tu, and Z. Otwinowski, "High-resolution timing of cell cycle-regulated gene expression," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 104, no. 43, pp. 16892–16897, 2007.

[134] A. Papagiannakis, B. Niebel, E. C. Wit, and M. Heinemann, "Autonomous metabolic oscillations robustly gate the early and late cell cycle," *Molecular Cell*, vol. 65, no. 2, pp. 285–295, 2017.

[135] A. B. Emerman, Z.-R. Zhang, O. Chakrabarti, and R. S. Hegde, "Trehalose Is a Key Determinant of the Quiescent Metabolic State That Fuels Cell Cycle Progression upon Return to Growth," *Molecular biology of the cell*, vol. 21, pp. 4325–4337, jun 2010.

[136] S. Laxman, B. M. Sutter, and B. P. Tu, "Behavior of a metabolic cycling population at the single cell level as visualized by fluorescent gene expression reporters," *PLoS ONE*, vol. 5, no. 9, pp. 1–11, 2010.

[137] B. Daignan-Fornier and I. Sagot, "Proliferation/quiescence: the controversial" aller-retour"," *Cell division*, vol. 6, no. 1, p. 10, 2011.

[138] A. J. Burnetti, M. Aydin, and N. E. Buchler, "Cell cycle Start is coupled to entry into the yeast metabolic cycle across diverse strains and growth rates," *Molecular Biology of the Cell*, vol. 27, no. 1, pp. 64–74, 2016.

[139] N. Slavov, E. M. Airoldi, A. van Oudenaarden, and D. Botstein, "A conserved cell growth cycle can account for the environmental stress responses of divergent eukaryotes," *Molecular Biology of the Cell*, vol. 23, no. 10, pp. 1986–1997, 2012.

[140] N. Slavov, J. Macinskas, A. Caudy, and D. Botstein, "Metabolic cycling without cell division cycling in respiring yeast," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 108, no. 47, pp. 19090–19095, 2011.

[141] R. R. Klevecz, "Quantized generation time in mammalian cells as an expression of the cellular clock," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 73, no. 11, pp. 4012–6, 1976.

[142] S. W. Chisholm and J. C. Costello, "Influence of Environmental Factors and Population Composition on the Timing of Cell Division in Thalassiosira Fluviatilis (Bacillariophyceae) Grown on Light/Dark Cycles," *Journal of Phycology*, vol. 16, no. 3, pp. 375–383, 1980.

[143] K. Homma and J. W. Hastings, "Cell cycle synchronization of Gonyaulax polyedra by filtration: quantized generation times," *Journal of Biological Rhythms*, vol. 3, no. 1, pp. 49–58, 1988.

[144] M. R. Parthun, J. Widom, and D. E. Gottschling, "The major cytoplasmic histone acetyltransferase in yeast: Links to chromatin replication and histone metabolism," *Cell*, vol. 87, no. 1, pp. 85–94, 1996.

[145] T. Suganuma and J. L. Workman, "Signals and combinatorial functions of histone modifications.," *Annual review of Biochemistry*, vol. 80, pp. 473–499, 2011.

[146] C. Lu and C. B. Thompson, "Metabolic regulation of epigenetics," *Cell Metabolism*, vol. 16, no. 1, pp. 9–17, 2012.

[147] B. D. Strahl and C. D. Allis, "The language of covalent histone modifications," *Nature*, vol. 403, no. 6765, pp. 41–45, 2000.

[148] A. J. Bannister and T. Kouzarides, "Regulation of chromatin by histone modifications," *Cell Research*, vol. 21, pp. 381–395, mar 2011.

[149] P. J. Park, "ChIP-Seq: advantages and challenges of a maturing technology," *Nature reviews. Genetics*, vol. 10, no. 10, pp. 669–680, 2009.

[150] Z. Kuang, Z. Ji, J. D. Boeke, and H. Ji, "Dynamic motif occupancy (DynaMO) analysis identifies transcription factors and their binding sites driving dynamic biological processes," *Nucleic Acids Research*, vol. 46, no. 1, pp. 1–16, 2018.

[151] Z. Kuang, S. Pinglay, H. Ji, and J. D. Boeke, "Msn2/4 regulate expression of glycolytic enzymes and control transition from quiescence to growth," *eLife*, vol. 6, p. e29938, 2017.

[152] Z. Kuang, H. Ji, and J. D. Boeke, "Stress response factors drive regrowth of quiescent cells," *Current Genetics*, vol. 64, pp. 807–810, aug 2018.

[153] M. Wyart, D. Botstein, and N. S. Wingreen, "Evaluating Gene Expression Dynamics Using Pairwise RNA FISH Data," *PLoS Computational Biology*, vol. 6, no. 11, pp. 1–14, 2010.

[154] S. J. Silverman, A. A. Petti, N. Slavov, L. Parsons, R. Briehof, S. Y. Thiberge, D. Zenklusen, S. J. Gandhi, D. R. Larson, R. H. Singer, and D. Botstein, "Metabolic cycling in single yeast cells from unsynchronized steady-state populations limited on glucose or phosphate," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 107, no. 15, pp. 6946–6951, 2010.

[155] A. Papagiannakis, J. J. De Jonge, Z. Zhang, and M. Heinemann, "Quantitative characterization of the auxin-inducible degron: A guide for dynamic protein depletion in single yeast cells," *Scientific Reports*, vol. 7, no. 1, pp. 1–13, 2017.

[156] Z. Chen, E. A. Odstrcil, B. P. Tu, and S. L. McKnight, "Restriction of DNA Replication to the Cycle Protects Genome Integrity," *Science*, vol. 316, no. June, pp. 1916–1919, 2007.

[157] B. P. Tu and S. L. McKnight, "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*, vol. 106, no. 34, pp. 14293–6, 2009.

[158] J. Feltham, "Gene expression in the S. cerevisiae xrn1Δ mutant during the yeast metabolic cycle," 2018. Personal communication concerning unpublished data.

[159] A. Sveiczer, B. Novak, J. M. Mitchison, B. Novák, and J. M. Mitchison, "The size control of fission yeast revisited," *Journal of Cell Science*, vol. 109, no. 12, pp. 2947–2957, 1996.

[160] A. Lovrics, A. Csikász-nagy, I. G. Zsély, J. Zádor, T. Turányi, and B. Novák, "Time scale and dimension analysis of a budding yeast cell cycle model," *BMC Bioinformatics*, vol. 11, no. 7:494, pp. 1–11, 2006.

[161] B. Novak and J. J. Tyson, "Modeling the control of DNA replication in fission yeast," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 94, no. August, pp. 9147–9152, 1997.

[162] A. Sveiczer, A. Csikasz-Nagy, B. Gyorffy, J. J. Tyson, and B. Novák, "Modeling the fission yeast cell cycle: quantized cycle times in wee1- cdc25Delta mutant cells," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 97, no. 14, pp. 7865–70, 2000.

[163] J. J. Tyson, K. Chen, and B. Novak, "Network dynamics and cell physiology," *Nature reviews. Molecular cell biology*, vol. 2, no. 12, pp. 908–916, 2001.

[164] B. Novak, J. J. Tyson, B. Gyorffy, and A. Csikasz-Nagy, "Irreversible cell-cycle transitions are due to systems-level feedback," *Nature Cell Biology*, vol. 9, pp. 724–728, jul 2007.

[165] K. C. Chen, L. Calzone, A. Csikasz-nagy, F. R. Cross, B. Novák, and J. J. Tyson, "Integrative Analysis of Cell Cycle Control in Budding Yeast," *Molecular Biology of the Cell*, vol. 15, no. August, pp. 3841–3862, 2004.

[166] W. M. van Gulik and J. J. Heijnen, "A metabolic network stoichiometry analysis of microbial growth and product formation," *Biotechnology and Bioengineering*, vol. 48, pp. 681–698, dec 1995.

[167] P. A. Vanrolleghem, P. De Jong-Gubbels, W. M. Van Gulik, J. T. Pronk, J. P. Van Dijken, and S. Heijnen, "Validation of a metabolic network for Saccharomyces cerevisiae using mixed substrate studies," *Biotechnology Progress*, vol. 12, pp. 434–448, aug 1996.

[168] T. L. Nissen, U. Schulze, J. Nielsen, and J. Villadsen, "Flux distributions in anaerobic, glucose-limited continuous cultures of Saccharomyces cerevisiae," jan 1997.

[169] J. Förster, I. Famili, P. Fu, B. Palsson, and J. Nielsen, "Genome-scale reconstruction of the Saccharomyces cerevisiae metabolic network," *Genome Research*, vol. 13, pp. 244–253, feb 2003.

[170] M. J. Herrgard, N. Swainston, P. Dobson, W. B. Dunn, K. Y. Arga, M. Arvas, N. Bluthgen, S. Borger, R. Costenoble, M. Heinemann, M. Hucka, N. Le Novere, P. Li, W. Liebermeister, M. L. Mo, A. P. Oliveira, D. Petranovic, S. Pettifer, E. Simeonidis, K. Smallbone, I. Spasic, D. Weichart, R. Brent, D. S. Broomhead, H. V. Westerhoff, B. Kirdar, M. Penttila, E. Klipp, B. O. Palsson, U. Sauer, S. G. Oliver, P. Mendes, J. Nielsen, and D. B. Kell, "A consensus yeast metabolic network reconstruction obtained from a community approach to systems biology," *Nature Biotechnology*, vol. 26, pp. 1155–1160, oct 2008.

[171] P. D. Dobson, K. Smallbone, D. Jameson, E. Simeonidis, K. Lanthaler, P. Pir, C. Lu, N. Swainston, W. B. Dunn, P. Fisher, D. Hull, M. Brown, O. Oshota, N. J. Stanford, D. B. Kell, R. D. King, S. G. Oliver, R. D. Stevens, and P. Mendes, "Further developments towards a genome-scale metabolic model of yeast," *BMC Systems Biology*, vol. 4, p. 145, oct 2010.

[172] D. B. Murray and R. Machne, "The yin and yang of yeast transcription: elements of a global feedback system between metabolism and chromatin," *PLoS ONE*, vol. 7, no. 6, pp. 1–21, 2012.

[173] C. Amariei, R. Machné, V. Stolc, T. Soga, M. Tomita, and D. B. Murray, "Time resolved DNA occupancy dynamics during the respiratory oscillation uncover a global reset point in the yeast growth program," *Microbial Cell*, vol. 1, no. 9, pp. 279–288, 2014.

[174] D. B. Murray, M. Beckmann, and H. Kitano, "Regulation of yeast oscillatory dynamics," *Proc Natl Acad Sci U S A*, vol. 104, no. 7, pp. 2241–2246, 2007.

[175] E. J. Molinelli, A. Korkut, W. Wang, M. L. Miller, N. P. Gauthier, X. Jing, P. Kaushik, Q. He, G. Mills, D. B. Solit, C. A. Pratilas, M. Weigt, A. Braunstein, A. Pagnani, R. Zecchina, and C. Sander, "Perturbation Biology: Inferring Signaling Networks in Cellular Systems," *PLoS Computational Biology*, vol. 9, p. e1003290, dec 2013.

[176] R. Christiano, N. Nagaraj, F. Fröhlich, and T. C. Walther, "Global Proteome Turnover Analyses of the Yeasts S.cerevisiae and S.pombe," *Cell Reports*, vol. 9, pp. 1959–1966, dec 2014.

[177] R. S. Edgar, E. W. Green, Y. Zhao, G. van Ooijen, M. Olmedo, X. Qin, Y. Xu, M. Pan, U. K. Valekunja, K. A. Feeney, E. S. Maywood, M. H. Hastings, N. S. Baliga, M. Merrow, A. J. Millar, C. H. Johnson, C. P. Kyriacou, J. S. O'Neill, and A. B. Reddy, "Peroxiredoxins are conserved markers of circadian rhythms," *Nature*, vol. 485, no. 7399, pp. 459–464, 2012.

[178] P. E. Hardin and W. Yu, "Circadian Transcription: Passing the HAT to CLOCK," *Cell*, vol. 125, no. 3, pp. 424–426, 2006.

[179] M. Doi, J. Hirayama, and P. Sassone-Corsi, "Circadian Regulator CLOCK Is a Histone Acetyltransferase," *Cell*, vol. 125, no. 3, pp. 497–508, 2006.

[180] Y. Nakahata, B. Grimaldi, S. Sahar, J. Hirayama, and P. Sassone-Corsi, "Signaling to the circadian clock: plasticity by chromatin remodeling," *Current Opinion in Cell Biology*, vol. 19, no. 2, pp. 230–237, 2007.

[181] J. E. Gangwisch, "Epidemiological evidence for the links between sleep, circadian rhythms and metabolism," *Obesity Reviews*, vol. 10, pp. 37–45, nov 2009.

[182] A. Ribas-Latre and K. Eckel-Mahan, "Interdependence of nutrient metabolism and the circadian clock system: Importance for metabolic health," *Molecular Metabolism*, vol. 5, pp. 133–152, mar 2016.

[183] R. A. Cairns, I. S. Harris, and T. W. Mak, "Regulation of cancer cell metabolism," *Nature Reviews Cancer*, vol. 11, pp. 85–95, feb 2011.

[184] U. E. Martínez-Outschoorn, M. Peiris-Pagés, R. G. Pestell, F. Sotgia, and M. P. Lisanti, "Cancer metabolism: A therapeutic perspective," *Nature Reviews Clinical Oncology*, vol. 14, pp. 11–31, jan 2017.

[185] G. Kroemer and J. Pouyssegur, "Tumor Cell Metabolism: Cancer's Achilles' Heel," *Cancer Cell*, vol. 13, pp. 472–482, jun 2008.

[186] R. J. De Berardinis and N. S. Chandel, "Fundamentals of cancer metabolism," *Science Advances*, vol. 2, no. 5, 2016.

[187] C. T. Walsh, B. P. Tu, and Y. Tang, "Eight Kinetically Stable but Thermodynamically Activated Molecules that Power Cell Metabolism," *Chemical Reviews*, 2017.

[188] W. Trevelyan and J. Harrison, "Studies on yeast metabolism. 7. Yeast carbohydrate fractions. Separation from nucleic acid, analysis, and behaviour during anaerobic fermentation," *Biochemical Journal*, vol. 63, pp. 23–33, may 1956.

[189] P. M. Bruinenberg, P. H. de Bot, J. P. van Dijken, and W. A. Scheffers, "The role of redox balances in the anaerobic fermentation of xylose by yeasts," *European Journal of Applied Microbiology and Biotechnology*, vol. 18, no. 5, pp. 287–292, 1983.

[190] S. R. Hackett, V. R. T. Zanotelli, W. Xu, J. Goya, J. O. Park, D. H. Perlman, P. A. Gibney, D. Botstein, J. D. Storey, and J. D. Rabinowitz, "Systems-level analysis of mechanisms regulating yeast metabolic flux," *Science*, vol. 354, no. 6311, pp. aaf2786–aaf2786, 2016.

[191] Z. Xu and K. Tsurugi, "Destabilization of energy-metabolism oscillation in the absence of trehalose synthesis in the chemostat culture of yeast," *Archives of Biochemistry and Biophysics*, vol. 464, no. 2, pp. 350–358, 2007.

[192] D. J. Acheson, *Elementary Fluid Dynamics*. New York: Oxford University Press, 1991.

[193] G. H. Wadhams and J. P. Armitage, "Making sense of it all: bacterial chemotaxis," *Nat.Rev.Mol.Cell Biol.*, vol. 5, pp. 1024–1037, dec 2004.

[194] J. Green and M. S. Paget, "Bacterial redox sensors," *Nature Reviews Microbiology*, vol. 2, pp. 954–966, dec 2004.

[195] B. L. Taylor, "How do bacteria find the optimal concentration of oxygen?," *Trends in Biochemical Sciences*, vol. 8, pp. 438–441, dec 1983.

[196] B. L. Taylor and I. B. Zhulin, "PAS domains: internal sensors of oxygen, redox potential, and light.," *Microbiology and molecular biology reviews*, vol. 63, pp. 479–506, jun 1999.

[197] M. A. Goldberg, S. P. Dunning, and H. F. Bunn, "Regulation of the erythropoietin gene: evidence that the oxygen sensor is a heme protein.," *Science*, vol. 242, pp. 1412–1415, dec 1988.

[198] A. C. Redfield, "The biological control of chemical factors in the environent," *American Scientist*, vol. 46, p. 230, 1958.

[199] D. B. Roszak and R. R. Colwell, "Survival strategies of bacteria in the natural environment.," *Microbiological reviews*, vol. 51, pp. 365–379, sep 1987.

[200] J. P. Armitage, *Bacterial tactic responses*, vol. 41. Academic Press, jan 1999.

[201] K. K. Jefferson, "What drives bacteria to produce a biofilm?," *FEMS Microbiol. Lett.*, vol. 236, pp. 163–173, jul 2004.

[202] S. I. Bibikov, R. O. Y. Biran, and K. E. Rudd, "A signal transducer for aerotaxis in Escherichia A Signal Transducer for Aerotaxis in Escherichia coli," *Journal of Bacteriology*, vol. 179, no. 12, pp. 4075–4079, 1997.

[203] E. K. Monson, M. Weinstein, G. S. Ditta, and D. R. Helinski, "The FixL protein of Rhizobium meliloti can be separated into a heme-binding oxygen-sensing domain and a functional C-terminal kinase domain.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 89, no. 10, pp. 4280–4, 1992.

[204] P. J. Kiley and H. Beinert, "Oxygen sensing by the global regulator, FNR: the role of the iron- sulfur cluster," *FEMS Microbiol Rev*, vol. 22, no. 5, pp. 341–52., 1998.

[205] E. L. Mettert and P. J. Kiley, "Reassessing the Structure and Function Relationship of the O_2 Sensing Transcription Factor FNR," *Antioxidants & Redox Signaling*, vol. 00, no. 00, p. ars.2017.7365, 2017.

[206] W. Gong, B. Hao, S. S. Mansy, G. Gonzalez, M. a. Gilles-Gonzalez, and M. K. Chan, "Structure of a biological oxygen sensor: a new mechanism for heme-driven signal transduction.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 95, no. 26, pp. 15177–15182, 1998.

[207] J. Green, J. C. Crack, A. J. Thomson, and N. E. LeBrun, "Bacterial sensors of oxygen," *Current Opinion in Microbiology*, vol. 12, no. 2, pp. 145–151, 2009.

[208] W. G. Kaelin and P. J. Ratcliffe, "Oxygen Sensing by Metazoans: The Central Role of the HIF Hydroxylase Pathway," *Molecular Cell*, vol. 30, pp. 393–402, may 2008.

[209] M. Ivan, K. Kondo, H. Yang, W. Kim, J. Valiando, M. Ohh, A. Salic, J. M. Asara, W. S. Lane, and W. G. Kaelin, "HIFalpha Targeted for VHL-Mediated Destruction by Proline Hydroxylation: Implications for O2 Sensing," *Science*, vol. 292, no. 5516, pp. 464–468, 2001.

[210] D. Wang, C. Youngson, V. Wong, H. Yeger, M. C. Dinauer, E. Vega-Saenz Miera, B. Rudy, and E. Cutz, "NADPH-oxidase and a hydrogen peroxide-sensitive K+ channel may function as an oxygen sensor complex in airway chemoreceptors and small cell lung carcinoma cell lines.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 93, no. 23, pp. 13182–7, 1996.

[211] M. A. McDonough, V. Li, E. Flashman, R. Chowdhury, C. Mohr, B. M. R. Lienard, J. Zondlo, N. J. Oldham, I. J. Clifton, J. Lewis, L. A. McNeill, R. J. M. Kurzeja, K. S. Hewitson, E. Yang, S. Jordan, R. S. Syed, and C. J. Schofield, "Cellular oxygen sensing: Crystal structure of hypoxia-inducible factor prolyl hydroxylase (PHD2)," *Proceedings of the National Academy of Sciences*, vol. 103, no. 26, pp. 9814–9819, 2006.

[212] A. A. Urrutia and J. Aragonés, "HIF Oxygen Sensing Pathways in Lung Biology," *Biomedicines*, vol. 6, jun 2018.

[213] K. E. Kwast, P. V. Burke, and R. O. Poyton, "Oxygen sensing and the transcriptional regulation of oxygen-responsive genes in yeast," *The Journal of experimental biology*, vol. 201, no. Pt 8, pp. 1177–1195, 1998.

[214] K. E. Kwast, P. V. Burke, B. T. Staahl, and R. O. Poyton, "Oxygen sensing in yeast: Evidence for the involvement of the respiratory chain in regulating

the transcription of a subset of hypoxic genes," *Proceedings of the National Academy of Sciences*, vol. 96, no. 10, pp. 5446–5451, 1999.

[215] R. O. Poyton, R. Dirmeier, K. O'Brien, P. David, and A. Dodd, "Experimental strategies for analyzing oxygen sensing in yeast," in *Oxygen Sensing*, vol. 381 of *Methods in Enzymology*, pp. 644 – 662, Academic Press, 2004.

[216] A. L. Hughes, B. L. Todd, and P. J. Espenshade, "SREBP Pathway Responds to Sterols and Functions as an Oxygen Sensor in Fission Yeast," *Cell*, vol. 120, no. 6, pp. 831 – 842, 2005.

[217] B. S. J. Davies and J. Rine, "A role for sterol levels in oxygen sensing in saccharomyces cerevisiae," *Genetics*, vol. 174, no. 1, pp. 191–201, 2006.

[218] C. J. Schofield and P. J. Ratcliffe, "Oxygen sensing by HIF hydroxylases," *Nature Reviews Molecular Cell Biology*, vol. 5, pp. 343–354, may 2004.

[219] H. F. Bunn and R. O. Poyton, "Oxygen sensing and molecular adaptation to hypoxia," *Physiological reviews*, vol. 76, pp. 839–885, jul 1996.

[220] J. T. van Dongen and F. Licausi, "Oxygen Sensing and Signaling," *Annual Review of Plant Biology*, vol. 66, pp. 345–367, apr 2015.

[221] N. R. Prabhakar and G. L. Semenza, "Oxygen Sensing and Homeostasis," *Physiology*, vol. 30, pp. 340–348, sep 2015.

[222] J. Feltham, "Chemostat dissolved oxygen data during the YMC," 2017. Personal communication concerning unpublished data.

[223] C. G. Evans, D. Herbert, and D. W. Tempest, "The Continuous Cultivation of Micro-organisms: 2. Construction of a Chemostat," *Methods in Microbiology*, vol. 2, pp. 277–327, jan 1970.

[224] B. Leimkuhler and S. Reich, *Simulating hamiltonian dynamics*, vol. 14. Cambridge University Press, 2004.

[225] R. Peikert, H. Hauser, H. Carr, and R. Fuchs, *Topological Methods in Data Analysis and Visualization II: Theory, Algorithms, and Applications*. Springer Science & Business Media, 2012.

[226] W. K. Lewis and W. G. Whitman, "Absorption symposium," *Industrial and Engineering Chemnisty*, vol. 16, no. 12, pp. 1215–1220, 1924.

[227] C. A. Ward, "The rate of gas absorption at a liquid interface," *The Journal of Chemical Physics*, vol. 67, no. 1, pp. 229–235, 1977.

[228] A. Gross, C. M. Li, F. Remacle, and R. D. Levine, "Free energy rhythms in Saccharomyces cerevisiae: A dynamic perspective with implications for ribosomal biogenesis," *Biochemistry*, vol. 52, no. 9, pp. 1641–1648, 2013.

[229] J. Tokunaga, "Solubilities of Oxygen, Nitrogen, and Carbon Dioxide in Aqueous Alcohol Solutions," *Journal of Chemical and Engineering Data*, vol. 20, no. 1, pp. 41–46, 1975.

[230] F. Vendruscolo, M. J. Rossi, W. Schmidell, and J. L. Ninow, "Determination of Oxygen Solubility in Liquid Media," *ISRN Chemical Engineering*, vol. 2012, pp. 1–5, jun 2012.

[231] R. F. Weiss, "The solubility of nitrogen, oxygen and argon in water and seawater," *Deep-Sea Research and Oceanographic Abstracts*, vol. 17, no. 4, pp. 721–735, 1970.

[232] D. J. Erickson Iii and D. J. Erickson, "A Stability Dependent Theory For Air-Sea Gas Exchange," *Journal of Geophysical Research*, vol. 98, no. 15, pp. 8471–8488, 1993.

[233] J. Korsgaard, "House-dust mites and absolute indoor humidity," *Allergy*, vol. 38, no. 2, pp. 85–92, 1983.

[234] L. Zhang and J. Niu, "Indoor humidity behaviors associated with decoupled cooling in hot and humid climates," *Building and Environment*, vol. 38, no. 1, pp. 99–107, 2003.

[235] Z. Wang, M. Gerstein, and M. Snyder, "RNA-Seq: a revolutionary tool for transcriptomics," *Nature reviews. Genetics*, vol. 10, no. 1, pp. 57–63, 2009.

[236] S. Marguerat and J. Bähler, "RNA-seq: From technology to biology," *Cellular and Molecular Life Sciences*, vol. 67, pp. 569–579, feb 2010.

[237] M. D. Robinson and A. Oshlack, "A scaling normalization method for differential expression analysis of RNA-seq data," *Genome Biology*, vol. 11, p. R25, mar 2010.

[238] Y. Chu and D. R. Corey, "RNA sequencing: platform selection, experimental design, and data interpretation.," *Nucleic acid therapeutics*, vol. 22, pp. 271–4, aug 2012.

[239] M. Garber, M. G. Grabherr, M. Guttman, and C. Trapnell, "Computational methods for transcriptome annotation and quantification using RNA-seq," *Nature Methods*, vol. 8, pp. 469–477, jun 2011.

[240] L. S. Churchman and J. S. Weissman, "Nascent transcript sequencing visualizes transcription at nucleotide resolution," *Nature*, vol. 469, no. 7330, pp. 368–373, 2011.

[241] L. S. Churchman and J. S. Weissman, "Native elongating transcript sequencing (NET-seq)," *Current Protocols in Molecular Biology*, vol. 1, no. SUPPL.98, pp. 1–17, 2012.

[242] A. Barski, S. Cuddapah, K. Cui, T. Y. Roh, D. E. Schones, Z. Wang, G. Wei, I. Chepelev, and K. Zhao, "High-Resolution Profiling of Histone Methylations in the Human Genome," *Cell*, vol. 129, no. 4, pp. 823–837, 2007.

[243] G. Robertson, M. Hirst, M. Bainbridge, M. Bilenky, Y. Zhao, T. Zeng, G. Euskirchen, B. Bernier, R. Varhol, A. Delaney, N. Thiessen, O. L. Griffith, A. He, M. Marra, M. Snyder, and S. Jones, "Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing," *Nature Methods*, vol. 4, pp. 651–657, aug 2007.

[244] D. S. Johnson, A. Mortazavi, R. M. Myers, and B. Wold, "Genome-wide mapping of in vivo protein-DNA interactions," *Science*, vol. 316, pp. 1497–1502, feb 2007.

[245] T. S. Mikkelsen, M. Ku, D. B. Jaffe, B. Issac, E. Lieberman, G. Giannoukos, P. Alvarez, W. Brockman, T. K. Kim, R. P. Koche, W. Lee, E. Mendenhall, A. O'Donovan, A. Presser, C. Russ, X. Xie, A. Meissner, M. Wernig, R. Jaenisch, C. Nusbaum, E. S. Lander, and B. E. Bernstein, "Genome-wide maps of chromatin state in pluripotent and lineage-committed cells," *Nature*, vol. 448, pp. 553–560, aug 2007.

[246] P. V. Kharchenko, M. Y. Tolstorukov, and P. J. Park, "Design and analysis of ChIP-seq experiments for DNA-binding proteins," *Nature Biotechnology*, vol. 26, pp. 1351–1359, dec 2008.

[247] T. Jenuwein and C. D. Allis, "Translating the histone code," *Science*, vol. 293, pp. 1074–1080, aug 2001.

[248] A. Angel, J. Song, C. Dean, and M. Howard, "A Polycomb-based switch underlying quantitative epigenetic memory," *Nature*, vol. 476, no. 7358, pp. 105–108, 2011.

[249] J. Song, A. Angel, M. Howard, and C. Dean, "Vernalization - a cold-induced epigenetic switch," *Journal of cell science*, vol. 125, no. Pt 16, pp. 3723–3731, 2012.

[250] M. H. Kuo and C. D. Allis, "Roles of histone acetyltransferases and deacetylases in gene regulation," *BioEssays*, vol. 20, pp. 615–626, dec 1998.

[251] S. L. Berger, "Histone modifications in transcriptional regulation," *Current Opinion in Genetics and Development*, vol. 12, pp. 142–148, apr 2002.

[252] A. B. Reddy and G. Rey, "Metabolic and Nontranscriptional Circadian Clocks: Eukaryotes," *Annual Review of Biochemistry*, vol. 83, no. 1, pp. 165–189, 2014.

[253] M. Foo, D. G. Bates, and O. E. Akman, "A simplified modelling framework facilitates more complex representations of plant circadian clocks," *PLOS Computational Biology*, vol. 16, pp. 1–34, 03 2020.

[254] I. T. Tokuda, O. E. Akman, and J. C. Locke, "Reducing the complexity of mathematical models for the plant circadian clock by distributed delays," *Journal of Theoretical Biology*, vol. 463, pp. 155 – 166, 2019.

[255] A. Angel, J. Song, H. Yang, J. I. Questa, C. Dean, and M. Howard, "Vernalizing cold is registered digitally at FLC.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 112, no. 13, pp. 4146–4151, 2015.

[256] M. Depledge, "Disruption of endogenous rhythms in carcinus maenas (l.) following exposure to mercury pollution," *Comparative Biochemistry and Physiology Part A: Physiology*, vol. 78, no. 2, pp. 375–379, 1984.

[257] T. Klein, H. Martens, D.-J. Dijk, R. E. Kronauer, E. W. Seely, and C. A. Czeisler, "Circadian sleep regulation in the absence of light perception: Chronic non-24-hour circadian rhythm sleep disorder in a blind man with a regular 24-hour sleep—wake schedule," *Sleep*, vol. 16, no. 4, pp. 333–343, 1993.

[258] B. Sharma and S. Feinsilver, "Circadian rhythm sleep disorders: An update," *Sleep and Biological Rhythms*, vol. 7, no. 2, pp. 113–124, 2009.

[259] L. Shi and B. P. Tu, "Protein acetylation as a means to regulate protein function in tune with metabolic state," *Biochemical Society Transactions*, vol. 42, no. 4, pp. 1037–1042, 2014.

[260] J. J. Tyson, C. I. Hong, C. D. Thron, and B. Novak, "A simple model of circadian rhythms based on dimerization and proteolysis of PER and TIM.," *Biophysical journal*, vol. 77, no. 5, pp. 2411–7, 1999.

[261] D. B. Murray, S. Roller, H. Kuriyama, and D. Lloyd, "Clock control of ultradian respiratory oscillation found during yeast continuous culture," *Journal of Bacteriology*, vol. 183, no. 24, pp. 7253–7259, 2001.

[262] M. W. Young and S. A. Kay, "Time zones: a comparative genetics of circadian clocks," *Nature Reviews*, vol. 2, no. September, pp. 702–715, 2001.

[263] H. Wijnen and M. W. Young, "Interplay of circadian clocks and metabolic rhythms.," *Annual review of genetics*, vol. 40, pp. 409–448, 2006.

[264] M. Hastings, J. S. O'Neill, and E. S. Maywood, "Circadian clocks: regulators of endocrine and metabolic rhythms," *Journal of Endocrinology*, vol. 195, no. 2, pp. 187–198, 2007.

[265] T. L. T. Deborah Bell-Pedersen, Vincent M. Cassone, David J. Earnest, Susan S. Golden, Paul E. Hardin and M. J. Zoran, "Circadian rhythms from multiple oscillators: lessons from diverse organisms," *Nature Reviews Drug Discovery*, vol. 4, no. 2, pp. 121–130, 2005.

[266] J. M. Hurley, J. J. Loros, and J. C. Dunlap, "Circadian Oscillators: Around the Transcription–Translation Feedback Loop and on to Output," *Trends in Biochemical Sciences*, vol. 41, no. 10, pp. 1–13, 2016.

[267] Z. Chen and S. L. McKnight, "A conserved DNA damage response pathway responsible for coupling the cell division cycle to the circadian and metabolic cycles," *Cell Cycle*, vol. 6, no. 23, pp. 2906–2912, 2007.

[268] H. C. Causton, K. A. Feeney, C. A. Ziegler, and J. S. O'Neill, "Metabolic cycles in yeast share features conserved among circadian rhythms," *Current Biology*, vol. 25, no. 8, pp. 1056–1062, 2015.

[269] J. Mellor, "The molecular basis of metabolic cycles and their relationship to circadian rhythms," *Nature Structural & Molecular Biology*, vol. 23, no. 12, pp. 1035–1044, 2016.

[270] A. Nieborak and R. Schneider, "Metabolic intermediates - Cellular messengers talking to chromatin modifiers," *Molecular Metabolism*, pp. 1–14, 2018.

[271] V. Moses and K. Lonberg-Holm, "The study of metabolic compartmentalization," *Journal of theoretical biology*, vol. 10, no. 2, pp. 336–355, 1966.

[272] A. H. Chen and P. A. Silver, "Designing biological compartmentalization," *Trends in cell biology*, vol. 22, no. 12, pp. 662–670, 2012.

[273] Y. Liu, N. F. Tsinoremas, C. H. Johnson, N. V. Lebedeva, S. S. Golden, M. Ishiura, and T. Kondo, "Circadian orchestration of gene expression in cyanobacteria.," *Genes & development*, vol. 9, no. 12, pp. 1469–1478, 1995.

[274] A. Claridge-Chang, H. Wijnen, F. Naef, C. Boothroyd, N. Rajewsky, and M. W. Young, "Circadian regulation of gene expression systems in the drosophila head," *Neuron*, vol. 32, no. 4, pp. 657–671, 2001.

[275] M. Kawahata, K. Masaki, T. Fujii, and H. Iefuji, "Yeast genes involved in response to lactic acid and acetic acid: acidic conditions caused by the organic acids in saccharomyces cerevisiae cultures induce expression of intracellular metal metabolism genes regulated by aft1p," *FEMS yeast research*, vol. 6, no. 6, pp. 924–936, 2006.

[276] B. Sonnleitner, "Dynamics of yeast metabolism and regulation," *Bioprocess engineering*, vol. 6, no. 5, pp. 187–193, 1991.

[277] G. H. Fleet, "Yeast spoilage of foods and beverages," in *The yeasts*, pp. 53–63, Elsevier, 2011.

[278] S. Dashko, N. Zhou, C. Compagno, and J. Piškur, "Why, when, and how did yeast evolve alcoholic fermentation?," *FEMS yeast research*, vol. 14, no. 6, pp. 826–832, 2014.

[279] W. G. Kaelin and S. L. McKnight, "Influence of metabolism on epigenetics and disease," *Cell*, vol. 153, no. 1, pp. 56–69, 2013.

[280] M. C. Palumbo, L. Farina, and P. Paci, "Kinetics effects and modeling of mRNA turnover," *WIREs RNA*, vol. 6, no. 3, pp. 327–336, 2015.

[281] C. K. Lee, Y. Shibata, B. Rao, B. D. Strahl, and J. D. Lieb, "Evidence for nucleosome depletion at active regulatory regions genome-wide," *Nature Genetics*, vol. 36, pp. 900–905, aug 2004.

[282] P. Chi, C. D. Allis, and G. G. Wang, "Covalent histone modifications-miswritten, misinterpreted and mis-erased in human cancers," *Nature Reviews Cancer*, vol. 10, pp. 457–469, jul 2010.

[283] V. A. Spencer and J. R. Davie, "Role of covalent modifications of histones in regulating gene expression," *Gene*, vol. 240, pp. 1–12, nov 1999.

[284] J. A. Latham and S. Y. Dent, "Cross-regulation of histone modifications," *Nature Structural and Molecular Biology*, vol. 14, pp. 1017–1024, nov 2007.

[285] S. D. Taverna, H. Li, A. J. Ruthenburg, C. D. Allis, and D. J. Patel, "How chromatin-binding modules interpret histone modifications: Lessons from professional pocket pickers," *Nature Structural and Molecular Biology*, vol. 14, pp. 1025–1040, nov 2007.

[286] R. K. Richmond, D. F. Sargent, T. J. Richmond, K. Luger, and A. W. Ma, "Crystal structure of the nucleosome core particle at 2 .8 A resolution," *Nature*, vol. 389, pp. 251–260, 1997.

[287] D. E. Sterner and S. L. Berger, "Acetylation of Histones and Transcription-Related Factors," *Microbiology and Molecular Biology Reviews*, vol. 64, pp. 435–459, jun 2000.

[288] T. Kouzarides, "Chromatin Modifications and Their Function," *Cell*, vol. 128, pp. 693–705, feb 2007.

[289] C. Martin and Y. Zhang, "The diverse functions of histone lysine methylation," *Nat. Rev. Mol. Cell Biol.*, vol. 6, pp. 838–849, nov 2005.

[290] D. Y. Lee, J. J. Hayes, D. Pruss, and A. P. Wolffe, "A positive role for histone acetylation in transcription factor access to nucleosomal DNA," *Cell*, vol. 72, pp. 73–84, jan 1993.

[291] M. Grunstein, "Histone acetylation in chromatin structure and transcription," *Nature*, vol. 389, pp. 349–352, sep 1997.

[292] L. S. Beese, V. Derbyshire, and T. A. Steitz, "Structure of DNA polymerase I Klenow fragment bound to duplex DNA," *Science*, vol. 260, pp. 352–355, apr 1993.

[293] T. Nojima, T. Gomes, A. R. F. Grosso, H. Kimura, M. J. Dye, S. Dhir, M. Carmo-Fonseca, and N. J. Proudfoot, "Mammalian NET-seq reveals genome-wide nascent transcription coupled to RNA processing," *Cell*, vol. 161, no. 3, pp. 526–540, 2015.

[294] J. Houseley and D. Tollervey, "The Many Pathways of RNA Degradation," *Cell*, vol. 136, no. 4, pp. 763–776, 2009.

[295] C.-D. Lee and B. P. Tu, "Glucose-regulated phosphorylation of the puf protein puf3 regulates the translational fate of its bound mrnas and association with rna granules," *Cell reports*, vol. 11, no. 10, pp. 1638–1650, 2015.

[296] C. E. Brown and A. B. Sachs, "Poly(A) tail length control in Saccharomyces cerevisiae occurs by message-specific deadenylation.," *Molecular and cellular biology*, vol. 18, pp. 6548–59, nov 1998.

[297] M. Tucker, R. R. Staples, M. A. Valencia-sanchez, D. Muhlrad, and R. Parker, "Ccr4p is the catalytic subunit of a Ccr4p / Pop2p / Notp mRNA deadenylase complex in Saccharomyces cerevisiae," *EMBO Journal*, vol. 21, pp. 1427–1436, mar 2002.

[298] D. C. Schwartz and R. Parker, "mRNA Decapping in Yeast Requires Dissociation of the Cap Binding Protein, Eukaryotic Translation Initiation Factor 4E," *Molecular and Cellular Biology*, vol. 20, pp. 7933–7942, nov 2000.

[299] S. Huch, J. Gommlich, M. Muppavarapu, C. Beckham, and T. Nissan, "Membrane-association of mRNA decapping factors is independent of stress in budding yeast," *Scientific Reports*, vol. 6, p. 25477, jul 2016.

[300] C. Charenton, C. Gaudon-Plesse, Z. Fourati, V. Taverniti, R. Back, O. Kolesnikova, B. Séraphin, and M. Graille, "A unique surface on Pat1 C-terminal domain directly interacts with Dcp2 decapping enzyme and Xrn1 5°–3° mRNA exonuclease in yeast," *Proceedings of the National Academy of Sciences*, vol. 114, pp. E9493–E9501, oct 2017.

[301] N. Arndt, D. Ross-Kaschitza, A. Kojukhov, A. A. Komar, and M. Altmann, "Properties of the ternary complex formed by yeast eIF4E, p20 and mRNA," *Scientific Reports*, vol. 8, p. 6707, apr 2018.

[302] J. S. Anderson and R. P. Parker, "The 3' to 5' degradation of yeast mRNAs is a general mechanism for mRNA turnover that requires the SKI2 DEVH box

protein and 39 to 59 exonucleases of the exosome complex," *EMBO Journal*, vol. 17, no. 5, pp. 1497–1506, 1998.

[303] C. Allmang, J. Kufel, G. Chanfreau, P. Mitchell, E. Petfalski, and D. Tollervey, "Functions of the exosome in rRNA, snoRNA and snRNA synthesis Christine," *EMBO Journal*, vol. 18, pp. 5399–5410, feb 1999.

[304] M. Lubas, M. S. Christensen, M. S. Kristiansen, M. Domanski, L. G. Falkenby, S. Lykke-Andersen, J. S. Andersen, A. Dziembowski, and T. H. Jensen, "Interaction Profiling Identifies the Human Nuclear Exosome Targeting Complex," *Molecular Cell*, vol. 43, pp. 624–637, aug 2011.

[305] F. Cacace, P. Paci, V. Cusimano, A. Germani, and L. Farina, "Stochastic Modeling of Expression Kinetics Identifies Messenger Half-Lives and Reveals Sequential Waves of Co-ordinated Transcription and Decay," *PLoS Computational Biology*, vol. 8, no. 11, 2012.

[306] Y. Wang, C. L. Liu, J. D. Storey, R. J. Tibshirani, D. Herschlag, and P. O. Brown, "Precision and functional specificity in mRNA decay," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, no. 9, pp. 5860–5865, 2002.

[307] S. E. Munchel, R. K. Shultzaberger, N. Takizawa, and K. Weis, "Dynamic profiling of mRNA turnover reveals gene-specific and system-wide regulation of mRNA decay," *Molecular biology of the cell*, vol. 22, no. 15, pp. 2787–95, 2011.

[308] J. Grigull, S. Mnaimneh, J. Pootoolal, D. Mark, M. D. Robinson, and T. R. Hughes, "Genome-Wide Analysis of mRNA Stability Using Transcription Inhibitors and Microarrays Reveals Posttranscriptional Control of Ribosome Biogenesis Factors," *Molecular and cellular biology*, vol. 24, no. 12, pp. 5534–5547, 2004.

[309] A. von Gabain, J. G. Belasco, J. L. Schottel, A. C. Chang, and S. N. Cohen, "Decay of mRNA in Escherichia coli: investigation of the fate of specific segments of transcripts.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 80, pp. 653–7, feb 1983.

[310] J. R. Cole and M. Nomura, "Changes in the half-life of ribosomal protein messenger RNA caused by translational repression," *Journal of Molecular Biology*, vol. 188, pp. 383–392, apr 1986.

[311] E. Yang, E. van Nimwegen, M. Zavolan, N. Rajewsky, M. Schroeder, M. Magnasco, and J. E. Darnell, "Decay rates of human mRNAs: Correlation with functional characteristics and sequence attributes," *Genome Research*, vol. 13, pp. 1863–1872, aug 2003.

[312] B. M. Turner, "Histone acetylation and an epigenetic code," *BioEssays*, vol. 22, no. 9, pp. 836–845, 2000.

[313] J. Nakayama, J. C. Rice, B. D. Strahl, C. D. Allis, and S. I. S. Grewal, "Role of histone H3 lysine 9 methylation in epigenetic control of heterochromatin assembly," *Science*, vol. 292, pp. 110–113, apr 2001.

[314] C. L. Peterson and M.-A. Laniel, "Histones and histone modifications," *Current Biology*, vol. 14, pp. R546–R551, jul 2004.

[315] A. Lennartsson and K. Ekwall, "Histone modification patterns and epigenetic codes," *Biochimica et Biophysica Acta - General Subjects*, vol. 1790, no. 9, pp. 863–868, 2009.

[316] S. Henikoff and A. Shilatifard, "Histone modification: Cause or cog?," *Trends in Genetics*, vol. 27, pp. 389–396, oct 2011.

[317] S. L. Berger, "The complex language of chromatin regulation during transcription.," *Nature*, vol. 447, no. 7143, pp. 407–412, 2007.

[318] S. B. Ficarro, M. L. McCleland, P. T. Stukenberg, D. J. Burke, M. M. Ross, J. Shabanowitz, D. F. Hunt, and F. M. White, "Phosphoproteome analysis by mass spectrometry and its application to Saccharomyces cerevisiae," *Nature Biotechnology*, vol. 20, pp. 301–305, mar 2002.

[319] I. H. Lee and T. Finkel, "Metabolic regulation of the cell cycle," *Current Opinion in Cell Biology*, vol. 25, pp. 724–729, dec 2013.

[320] T. Kawakami, Y. Mishima, H. Hojo, and I. Suetake, "Synthesis of ubiquitylated histone H3 using a thiirane linker for chemical ligation," *Journal of Peptide Science*, vol. 23, pp. 532–538, jul 2017.

[321] B. R. Sabari, Z. Tang, H. Huang, V. Yong-Gonzalez, H. Molina, H. E. Kong, L. Dai, M. Shimada, J. R. Cross, Y. Zhao, *et al.*, "Intracellular crotonyl-coa stimulates transcription through p300-catalyzed histone crotonylation," *Molecular cell*, vol. 58, no. 2, pp. 203–215, 2015.

[322] Y. Li, D. Zhao, Z. Chen, and H. Li, "Yeats domain: Linking histone crotonylation to gene regulation," *Transcription*, vol. 8, no. 1, pp. 9–14, 2017.

[323] S. D. Briggs, M. Bryk, B. D. Strahl, W. L. Cheung, J. K. Davie, S. Y. Dent, F. Winston, and C. David Allis, "Histone H3 lysine 4 methylation is mediated by Set1 and required for cell growth and rDNA silencing in Saccharomyces cerevisiae," *Genes and Development*, vol. 15, pp. 3286–3295, dec 2001.

[324] G. Liang, J. C. Y. Lin, V. Wei, C. Yoo, J. C. Cheng, C. T. Nguyen, D. J. Weisenberger, G. Egger, D. Takai, F. A. Gonzales, and P. A. Jones, "Distinct localization of histone H3 acetylation and H3-K4 methylation to the transcription start sites in the human genome," *Proceedings of the National Academy of Sciences*, vol. 101, pp. 7357–7362, may 2004.

[325] P. A. Grant, A. Eberharter, S. John, R. G. Cook, B. M. Turner, and J. L. Workman, "Expanded lysine acetylation specificity of Gcn5 in native complexes," *Journal of Biological Chemistry*, vol. 274, pp. 5895–5900, feb 1999.

[326] D. K. Pokholok, C. T. Harbison, S. Levine, M. Cole, N. M. Hannett, I. L. Tong, G. W. Bell, K. Walker, P. A. Rolfe, E. Herbolsheimer, J. Zeitlinger, F. Lewitter, D. K. Gifford, and R. A. Young, "Genome-wide map of nucleosome acetylation and methylation in yeast," *Cell*, vol. 122, no. 4, pp. 517–527, 2005.

[327] Z. Nagy and L. Tora, "Distinct GCN5/PCAF-containing complexes function as co-activators and are involved in transcription factor and global histone acetylation," aug 2007.

[328] T. Y. Roh, W. C. Ngau, K. Cui, D. Landsman, and K. Zhao, "High-resolution genome-wide mapping of histone modifications," *Nature Biotechnology*, vol. 22, pp. 1013–1016, aug 2004.

[329] S. K. Kurdistani, S. Tavazoie, and M. Grunstein, "Mapping global histone acetylation patterns to gene expression," *Cell*, vol. 117, pp. 721–733, jun 2004.

[330] L. L. Chih, T. Kaplan, M. Kim, S. Buratowski, S. L. Schreiber, N. Friedman, and O. J. Rando, "Single-nucleosome mapping of histone modifications in S. cerevisiae," *PLoS Biology*, vol. 3, p. e328, aug 2005.

[331] T. Agalioti, G. Chen, and D. Thanos, "Deciphering the transcriptional histone acetylation code for a human gene," *Cell*, vol. 111, pp. 381–392, nov 2002.

[332] Z. Wang, C. Zang, J. A. Rosenfeld, D. E. Schones, A. Barski, S. Cuddapah, K. Cui, T.-Y. Y. Roh, W. Peng, M. Q. Zhang, and K. Zhao, "Combinatorial patterns of histone acetylations and methylations in the human genome," *Nature Genetics*, vol. 40, pp. 897–903, jul 2008.

[333] N. J. Krogan, M. Kim, A. Tong, A. Golshani, G. Cagney, V. Canadien, D. P. Richards, B. K. Beattie, A. Emili, C. Boone, A. Shilatifard, S. Buratowski, and J. Greenblatt, "Methylation of Histone H3 by Set2 in Saccharomyces cerevisiae Is Linked to Transcriptional Elongation by RNA Polymerase II," *Molecular and Cellular Biology*, vol. 23, pp. 4207–4218, jun 2003.

[334] P. Kolasinska-zwierz, T. Down, I. Latorre, T. Liu, X. S. Liu, and J. Ahringer, "Differential chromatin marking of introns and expressed exons by H3K36me3," *Nature Genetics*, vol. 41, no. 3, pp. 376–381, 2009.

[335] G. Le Martelot, D. Canella, L. Symul, E. Migliavacca, F. Gilardi, R. Liechti, O. Martin, K. Harshman, M. Delorenzi, B. Desvergne, W. Herr, B. Deplancke, U. Schibler, J. Rougemont, N. Guex, N. Hernandez, and F. Naef, "Genome-Wide RNA Polymerase II Profiles and RNA Accumulation Reveal Kinetics of Transcription and Associated Epigenetic Changes During Diurnal Cycles," *PLoS Biology*, vol. 10, p. e1001442, nov 2012.

[336] F. Xu, K. Zhang, and M. Grunstein, "Acetylation in histone H3 globular domain regulates gene expression in yeast," *Cell*, vol. 121, pp. 375–385, may 2005.

[337] T. Kaplan, C. L. Liu, J. A. Erkmann, J. Holik, M. Grunstein, P. D. Kaufman, N. Friedman, and O. J. Rando, "Cell cycle- and chaperone-mediated regulation of H3K56ac incorporation in yeast," *PLoS Genetics*, vol. 4, p. e1000270, nov 2008.

[338] M. Grunstein, "Histones as Regulators of Genes," *Scientific American*, vol. 267, no. 4, pp. 68–75, 1992.

[339] B. Blackwell, E. Marley, J. Price, and D. Taylor, "Hypertensive interactions between monoamine oxidase inhibitors and foodstuffs.," *The British journal of psychiatry : the journal of mental science*, vol. 113, pp. 349–365, apr 1967.

[340] E. H. Hare, J. Dominian, and L. Sharpe, "Phenelzine and Dexamphetamine in Depressive Illness," *British Medical Journal*, vol. 1, pp. 9–12, jan 1962.

[341] G. B. Baker, R. T. Coutts, K. F. McKenna, and R. L. Sherry-McKenna, "Insights into the mechanisms of action of the MAO inhibitors phenelzine and tranylcypromine: A review," *Journal of Psychiatry and Neuroscience*, vol. 17, pp. 206–214, nov 1992.

[342] M. G. Lee, C. Wynder, D. M. Schmidt, and D. G. Mccafferty, "Histone H3 Lysine 4 Brief Communication Demethylation Is a Target of Nonselective Antidepressive Medications," *Chemistry & Biology*, vol. 13, pp. 563–567, jun 2006.

[343] D. A. Price Evans, K. Davison, and R. T. C. Pratt, "The influence of acetylator phenotype on the effects of treating depression with phenelzine," *Clinical Pharmacology & Therapeutics*, vol. 6, pp. 430–435, jul 1965.

[344] P. Prusevich, J. H. Kalin, S. A. Ming, M. Basso, J. Givens, X. Li, J. Hu, M. S. Taylor, A. M. Cieniewicz, P. Y. Hsiao, R. Huang, H. Roberson, N. Adejola, L. B. Avery, R. A. Casero, S. D. Taverna, J. Qian, A. J. Tackett, R. R. Ratan, O. G. McDonald, A. P. Feinberg, and P. A. Cole, "A selective phenelzine analogue inhibitor of histone demethylase LSD1," *ACS Chemical Biology*, vol. 9, pp. 1284–1293, jun 2014.

[345] P. O. Brook, M. M. Perry, I. M. Adcock, and A. L. Durham, "Epigenome-modifying tools in asthma," *Epigenomics*, vol. 7, pp. 1017–1032, sep 2015.

[346] A. M. Fortress and K. M. Frick, "Pharmacologically manipulating learning and memory," in *The Maze Book: Theories, Practice, and Protocols for Testing Rodent Cognition*, pp. 165–210, Humana Press, New York, NY, 2015.

[347] J. H. Waterborg, "Dynamics of histone acetylation in vivo. A function for acetylation turnover?," *Biochemistry and Cell Biology*, vol. 80, no. 3, pp. 363–378, 2002.

[348] H. Xu, L. Handoko, X. Wei, C. Ye, J. Sheng, C. L. Wei, F. Lin, and W. K. Sung, "A signal-noise model for significance analysis of ChIP-seq with negative control," *Bioinformatics*, vol. 26, no. 9, pp. 1199–1204, 2010.

[349] T. S. Furey, "ChIP-seq and beyond: New and improved methodologies to detect and characterize protein-DNA interactions," *Nature Reviews Genetics*, vol. 13, no. 12, pp. 840–852, 2012.

[350] S. Pepke, B. Wold, and A. Mortazavi, "Computation for chip-seq and rna-seq studies," *Nature Methods*, vol. 6, p. S22, nov 2009.

[351] M. Imme, "A noise peak elimination filter," *CVGIP: Graphical Models and Image Processing*, vol. 53, no. 2, pp. 204–211, 1991.

[352] D. Dick and T. M. Shay, "Ultrahigh-noise rejection optical filter," *Optics letters*, vol. 16, no. 11, pp. 867–869, 1991.

[353] A. A. Alekseyenko, S. Peng, E. Larschan, A. A. Gorchakov, O. K. Lee, P. Kharchenko, S. D. McGrath, C. I. Wang, E. R. Mardis, P. J. Park, and M. I. Kuroda, "A Sequence Motif within Chromatin Entry Sites Directs MSL Establishment on the Drosophila X Chromosome," *Cell*, vol. 134, pp. 599–609, aug 2008.

[354] K. Pearson, "Biometrika Trust Notes on the History of Correlation," *Biometrika*, vol. 13, pp. 25–45, oct 2009.

[355] Y. Yu, C. Song, Q. Zhang, P. A. DiMaggio, B. A. Garcia, A. York, M. F. Carey, and M. Grunstein, "Histone h3 lysine 56 methylation regulates dna replication through its interaction with pcna," *Molecular cell*, vol. 46, no. 1, pp. 7–17, 2012.

[356] H. Wurtele, G. S. Kaiser, J. Bacal, E. St-Hilaire, E.-H. Lee, S. Tsao, J. Dorn, P. Maddox, M. Lisby, P. Pasero, and A. Verreault, "Histone H3 Lysine 56 Acetylation and the Response to DNA Replication Fork Damage," *Molecular and Cellular Biology*, vol. 32, pp. 154–172, jan 2012.

[357] J. V. Tjeertes, K. M. Miller, and S. P. Jackson, "Screen for DNA-damage-responsive histone modifications identifies H3K9Ac and H3K56Ac in human cells," *EMBO Journal*, vol. 28, pp. 1878–1889, jul 2009.

[358] Q. Li, H. Zhou, H. Wurtele, B. Davies, B. Horazdovsky, A. Verreault, and Z. Zhang, "Acetylation of Histone H3 Lysine 56 Regulates Replication-Coupled Nucleosome Assembly," *Cell*, vol. 134, pp. 244–255, jul 2008.

[359] J. Feltham, "Microscope imaging of chemostat cells during the YMC," 2018. Personal communication concerning unpublished data.

[360] B. Li, M. Gogol, M. Carey, D. Lee, C. Seidel, and J. L. Workman, "Combined action of PHD and chromo domains directs the Rpd3S HDAC to transcribed chromatin," *Science*, vol. 316, pp. 1050–1054, may 2007.

[361] K. P. Nightingale, L. P. O'Neill, and B. M. Turner, "Histone modifications: Signalling receptors and potential elements of a heritable epigenetic code," *Current Opinion in Genetics and Development*, vol. 16, pp. 125–136, apr 2006.

[362] T. Zhang, S. Cooper, and N. Brockdorff, "The interplay of histone modifications - writers that read," *EMBO reports*, vol. 16, pp. 1467–1481, nov 2015.

[363] K. Struhl, "Histone acetylation and transcriptional regulatory mechanisms.," *Genes and Development development*, vol. 12, pp. 599–606, mar 1998.

[364] D. Herrick, R. Parker, and A. Jacobson, "Identification and comparison of stable and unstable mRNAs in Saccharomyces cerevisiae," *Molecular and Cellular Biology*, vol. 10, pp. 2269–2284, may 1990.

[365] R. Parker, D. Herrick, S. W. Peltz, and A. Jacobson, "Measurement of mRNA decay rates in Saccharomyces cerevisiae," *Methods in Enzymology*, vol. 194, pp. 415–423, jan 1991.

[366] A. Jacobson, "Interrelationships of the Pathways of mRNA Decay and Translation in Eukaryotic Cells," *Annual Review of Biochemistry*, vol. 65, pp. 693–739, jun 1996.

[367] J. Coller and R. Parker, "Eukaryotic mRNA Decapping," *Annual Review of Biochemistry*, vol. 73, pp. 861–890, jun 2004.

[368] R. Parker and U. Sheth, "P Bodies and the Control of mRNA Translation and Degradation," *Molecular Cell*, vol. 25, pp. 635–646, mar 2007.

[369] L. E. Maquat and M. Kiledjian, "RNA turnover in eukaryotes: nucleases, pathways and analysis of mRNA decay. Preface.," *Methods in enzymology*, vol. 448, pp. xxi–xxii, 2008.

[370] F. C. Holstege, E. G. Jennings, J. J. Wyrick, T. I. Lee, C. J. Hengartner, M. R. Green, T. R. Golub, E. S. Lander, R. A. Young, T. I. Lee, F. C. Holstege, R. A. Young, E. G. Jennings, J. J. Wyrick, C. J. Hengartner, M. R. Green, and T. R. Golub, "Dissecting the regulatory circuitry of a eukaryotic genome," *Cell*, vol. 95, no. 5, pp. 717–728, 1998.

[371] J. García-Martínez, A. Aranda, and J. E. Pérez-Ortín, "Genomic run-on evaluates transcription rates for all yeast genes and identifies gene regulatory mechanisms," *Molecular Cell*, vol. 15, pp. 303–313, jul 2004.

[372] M. Rabani, J. Z. Levin, L. Fan, X. Adiconis, R. Raychowdhury, M. Garber, A. Gnirke, C. Nusbaum, N. Hacohen, N. Friedman, I. Amit, and A. Regev, "Metabolic labeling of RNA uncovers principles of RNA production and degradation dynamics in mammalian cells," *Nature Biotechnology*, vol. 29, pp. 436–442, may 2011.

[373] S. Xi, "NET-seq measurements on the Yeast Metabolic Cycle," 2018. Personal communication concerning unpublished data.

[374] M. Wery, C. Gautier, M. Descrimes, M. Yoda, H. Vennin-Rendos, V. Migeot, D. Gautheret, D. Hermand, and A. Morillon, "Native elongating transcript sequencing reveals global anti-correlation between sense and antisense nascent transcription in fission yeast," *RNA*, vol. 24, no. 2, pp. 196–208, 2018.

[375] C. de Boor, "A Practical Guide to Splines," *Mathematics of Computation*, vol. 34, no. 149, p. 325, 1980.

[376] P. H. C. Eilers and B. D. Marx, "Flexible smoothing with B-splines and penalties," *Statistical Science*, vol. 11, no. 2, pp. 89–121, 1996.

[377] E. Eden, D. Lipson, S. Yogev, and Z. Yakhini, "Discovering Motifs in Ranked Lists of DNA Sequences," *PLoS Computational Biology*, vol. 3, no. 3, p. e39, 2007.

[378] E. Eden, R. Navon, I. Steinfeld, D. Lipson, and Z. Yakhini, "GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists," *BMC Bioinformatics*, vol. 10, no. 1, p. 48, 2009.

[379] M. Ashburner, C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis, K. Dolinski, S. S. Dwight, J. T. Eppig, M. A. Harris, D. P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J. C. Matese, J. E. Richardson, G. M. Rubin, G. Sherlock, M. Ringwald, G. M. Rubin, and G. Sherlock, "Gene Ontology: tool for the unification of biology," *Nature Genetics*, vol. 25, pp. 25–29, may 2000.

[380] E. I. Boyle, S. Weng, J. Gollub, H. Jin, D. Botstein, J. M. Cherry, and G. Sherlock, "GO::TermFinder - Open source software for accessing Gene Ontology information and finding significantly enriched Gene Ontology terms associated with a list of genes," *Bioinformatics*, vol. 20, no. 18, pp. 3710–3715, 2004.

[381] S. Carbon, A. Ireland, C. J. Mungall, S. Shu, B. Marshall, S. Lewis, A. Hub, and W. Presence, "AmiGO: online access to ontology and annotation data," *Bioinformatics*, vol. 25, no. 2, pp. 288–289, 2009.

[382] F. Al-Shahrour, R. Díaz-Uriarte, and J. Dopazo, "FatiGO: A web tool for finding significant associations of Gene Ontology terms with groups of genes," *Bioinformatics*, vol. 20, no. 4, pp. 578–580, 2004.

[383] S. Maere, K. Heymans, and M. Kuiper, "BiNGO: A Cytoscape plugin to assess overrepresentation of Gene Ontology categories in Biological Networks," *Bioinformatics*, vol. 21, pp. 3448–3449, aug 2005.

[384] M. Mu, J. L. Souciet, J. D. Montigny, and D. Go, "CYGD: the Comprehensive Yeast Genome Database," *Nucleic acids research*, vol. 33, no. Database, pp. 364–368, 2017.

[385] M. Aslett and V. Wood, "Gene Ontology annotation status of the fission yeast genome: preliminary coverage approaches 100 %," *Yeast*, vol. 23, pp. 913–919, 2006.

[386] S. Y. Rhee, V. Wood, K. Dolinski, and S. Draghici, "Use and misuse of the gene ontology annotations," *Nature Reviews Genetics*, vol. 9, no. July, pp. 509–515, 2008.

[387] C. Bleidorn, "Sequencing Techniques," in *Phylogenomics* (Springer, ed.), ch. 3, pp. 43–60, Cham: Springer International Publishing, 2017.

[388] J. Taylor, *Introduction to error analysis, the study of uncertainties in physical measurements*. University Science Books, 1997.

[389] M. C. Teixeira, P. T. Monteiro, N. D. Mendes, S. D'orey, S. Tenreiro, N. P. Mira, H. Pais, A. P. Francisco, A. M. Carvalho, A. B. Lourenço, I. Sá-Correia, A. I. Oliveira, and A. T. Freitas, "The YEASTRACT database: a tool for the analysis of transcription regulatory associations in Saccharomyces cerevisiae," *Nucleic Acids Research*, vol. 34, no. 90001, pp. D446–D451, 2006.

[390] M. C. Teixeira, P. T. Monteiro, M. Palma, C. Costa, C. P. Godinho, P. Pais, M. Cavalheiro, M. Antunes, A. Lemos, T. Pedreira, and I. Sá-Correia, "YEASTRACT: an upgraded database for the analysis of transcription regulatory networks in Saccharomyces cerevisiae," *Nucleic Acids Research*, vol. 46, no. January, pp. 348–353, 2018.

[391] M. C. Teixeira, P. T. Monteiro, J. F. Guerreiro, J. P. Gonçalves, N. P. Mira, S. C. Dos Santos, T. R. Cabrito, M. Palma, C. Costa, A. P. Francisco, S. C. Madeira, A. L. Oliveira, A. T. Freitas, and I. Sá-Correia, "The YEASTRACT database: An upgraded information system for the analysis of gene and genomic transcription regulation in Saccharomyces cerevisiae," *Nucleic Acids Research*, vol. 42, no. D1, pp. 161–166, 2014.

[392] D. Abdulrehman, P. T. Monteiro, M. C. Teixeira, N. P. Mira, A. B. Lourenço, S. C. Dos Santos, T. R. Cabrito, A. P. Francisco, S. C. Madeira, R. S. Aires, A. L. Oliveira, I. Sá-Correia, and A. T. Freitas, "YEASTRACT: Providing a programmatic access to curated transcriptional regulatory associations in Saccharomyces cerevisiae through a web services interface," *Nucleic Acids Research*, vol. 39, no. SUPPL. 1, pp. 136–140, 2011.

[393] P. T. Monteiro, N. D. Mendes, M. C. Teixeira, S. D'orey, S. Tenreiro, N. P. Mira, H. Pais, A. P. Francisco, A. M. Carvalho, A. B. Lourenço, I. Sá-Correia, A. I. Oliveira, and A. T. Freitas, "YEASTRACT-DISCOVERER: New tools to improve the analysis of transcriptional regulatory associations in Saccharomyces cerevisiae," *Nucleic Acids Research*, vol. 36, no. Suppl. 1, pp. 132–136, 2008.

[394] E. Wingender, X. Chen, R. Hehl, H. Karas, I. Liebich, V. Matys, T. Meinhardt, M. Prüß, I. Reuter, and F. Schacherer, "Transfac: an integrated system for gene expression regulation," *Nucleic acids research*, vol. 28, no. 1, pp. 316–319, 2000.

[395] E. Wingender, X. Chen, E. Fricke, R. Geffers, R. Hehl, I. Liebich, M. Krull, V. Matys, H. Michael, R. Ohnhäuser, *et al.*, "The transfac system on gene expression regulation," *Nucleic acids research*, vol. 29, no. 1, pp. 281–283, 2001.

[396] V. Matys, E. Fricke, R. Geffers, E. Gößling, M. Haubrock, R. Hehl, K. Hornischer, D. Karas, A. E. Kel, O. V. Kel-Margoulis, *et al.*, "Transfac®: transcriptional regulation, from patterns to profiles," *Nucleic acids research*, vol. 31, no. 1, pp. 374–378, 2003.

[397] N. A. Kolchanov, O. A. Podkolodnaya, E. A. Ananko, E. V. Ignatieva, I. L. Stepanenko, O. V. Kel-Margoulis, A. E. Kel, T. I. Merkulova, T. Goryachkovskaya, T. Busygina, *et al.*, "Transcription regulatory regions database (trrd): its status in 2000," *Nucleic Acids Research*, vol. 28, no. 1, pp. 298–301, 2000.

[398] N. A. Kolchanov, E. V. Ignatieva, E. A. Ananko, O. A. Podkolodnaya, I. L. Stepanenko, T. I. Merkulova, M. A. Pozdnyakov, N. L. Podkolodny, A. N. Naumochkin, and A. G. Romashchenko, "Transcription regulatory regions database (trrd): its status in 2002," *Nucleic acids research*, vol. 30, no. 1, pp. 312–317, 2002.

[399] O. Kel-Margoulis, A. Kel, M. Frisch, A. Romaschenko, N. Kolchanov, and E. Wingender, "Compel a database on composite regulatory elements," in *Proceedings of the First International Conference on Bioinformatics of Genome Regulation and Structure*, pp. 54–57, 1998.

[400] O. V. Kel-Margoulis, A. G. Romashchenko, N. A. Kolchanov, E. Wingender, and A. E. Kel, "Compel: a database on composite regulatory elements providing combinatorial transcriptional regulation," *Nucleic acids research*, vol. 28, no. 1, pp. 311–315, 2000.

[401] D.-S. Huang and G. W. Irwin, *Computational Intelligence and Bioinformatics: International Conference on Intelligent Computing, ICIC 2006, Kunming, China, August 16-19, 2006, Proceedings*, vol. 4115. Springer, 2006.

[402] S. Das, D. Caragea, S. Welch, and W. H. Hsu, *Handbook of research on computational methodologies in gene regulatory networks*. IGI Global, 2010.

[403] L. Chen, L. Cai, G. Skogerbø, Y. Zhao, and R. Chen, "Assessing TF regulatory relationships of divergently transcribed genes," *Genomics*, vol. 92, no. 5, pp. 316–321, 2008.

[404] T. Schwartz, E. Craig, and D. Kennell, "Inactivation and degradation of messenger ribonucleic acid from the lactose operon of escherichia coli," *Journal of molecular biology*, vol. 54, no. 2, pp. 299–311, 1970.

[405] R. D. Mosteller, J. K. Rose, and C. Yanofsky, "Transcription initiation and degradation of trp mrna," in *Cold Spring Harbor Symposia on Quantitative Biology*, vol. 35, pp. 461–466, Cold Spring Harbor Laboratory Press, 1970.

[406] E. Schneider, M. Blundell, and D. Kennell, "Translation and mRNA decay," *MGG Molecular & General Genetics*, vol. 160, no. 2, pp. 121–129, 1978.

[407] S. P. Gygi, Y. Rochon, B. R. Franza, and R. Aebersold, "Correlation between Protein and mRNA Abundance in Yeast," *Molecular and Cellular Biology*, vol. 19, pp. 1720–1730, mar 1999.

[408] T. Trcek, D. R. Larson, A. Moldón, C. C. Query, and R. H. Singer, "Single-molecule mRNA decay measurements reveal promoter-regulated mRNA stability in yeast," *Cell*, vol. 147, no. 7, pp. 1484–1497, 2011.

[409] D. Apirion, "Degradation of rna in escherichia coli," *Molecular and General Genetics MGG*, vol. 122, no. 4, pp. 313–322, 1973.

[410] S. Pedersen, S. Reeh, and J. D. Friesen, "Functional mRNA half lives in E. coli," *MGG Molecular & General Genetics*, vol. 166, no. 3, pp. 329–336, 1978.

[411] N. Heintz, H. L. Sive, and R. G. Roeder, "Regulation of human histone gene expression: kinetics of accumulation and changes in the rate of synthesis and in the half-lives of individual histone mRNAs during the HeLa cell cycle.," *Molecular and cellular biology*, vol. 3, no. 4, pp. 539–50, 1983.

[412] J. Ross, "mRNA Stability in Mammalian Cells," *Microbiological reviews*, vol. 59, no. 3, pp. 423–450, 1995.

[413] G. Hambraeus, C. Von Wachenfeldt, and L. Hederstedt, "Genome-wide survey of mRNA half-lives in Bacillus subtilis identifies extremely stable mRNAs," *Molecular Genetics and Genomics*, vol. 269, no. 5, pp. 706–714, 2003.

[414] G. M. Church, G. M. Church, C. Rosenow, and C. Rosenow, "Global RNA Half-Life Analysis in Escherichia coli Reveals Positional Patterns of Transcript Degradation," *Genome Research*, vol. 13, pp. 216–223, 2003.

[415] R. Narsai, K. A. Howell, A. H. Millar, N. O'Toole, I. Small, and J. Whelan, "Genome-Wide Analysis of mRNA Decay Rates and Their Determinants in Arabidopsis thaliana," *the Plant Cell Online*, vol. 19, no. 11, pp. 3418–3436, 2007.

[416] D. J. Katzmann, M. Babst, and S. D. Emr, "Ubiquitin-dependent sorting into the multivesicular body pathway requires the function of a conserved endosomal protein sorting complex, escrt-i," *Cell*, vol. 106, no. 2, pp. 145–155, 2001.

[417] R. Kama, G. Gabriely, V. Kanneganti, and J. E. Gerst, "Cdc48 and ubiquilins confer selective anterograde protein sorting and entry into the multivesicular body in yeast," *Molecular biology of the cell*, vol. 29, no. 8, pp. 948–963, 2018.

[418] D. Moir, S. E. Stewart, B. C. Osmond, and D. Botstein, "Cold-sensitive cell-division-cycle mutants of yeast: isolation, properties, and pseudoreversion studies," *Genetics*, vol. 100, no. 4, pp. 547–563, 1982.

[419] M.-T. Hsieh and R.-H. Chen, "Cdc48 and cofactors npl4-ufd1 are important for g1 progression during heat stress by maintaining cell wall integrity in saccharomyces cerevisiae," *PloS one*, vol. 6, no. 4, 2011.

[420] A. Stolz, W. Hilt, A. Buchberger, and D. H. Wolf, "Cdc48: a power machine in protein degradation," *Trends in biochemical sciences*, vol. 36, no. 10, pp. 515–523, 2011.

[421] F. Neri, S. Rapelli, A. Krepelova, D. Incarnato, C. Parlato, G. Basile, M. Maldotti, F. Anselmi, and S. Oliviero, "Intragenic dna methylation prevents spurious transcription initiation," *Nature*, vol. 543, no. 7643, pp. 72–77, 2017.

[422] A. Morillon, N. Karabetsou, A. Nair, and J. Mellor, "Dynamic lysine methylation on histone H3 defines the regulatory phase of gene transcription," *Molecular Cell*, vol. 18, no. 6, pp. 723–734, 2005.

[423] K. Glover-Cutter, S. Larochelle, B. Erickson, C. Zhang, K. Shokat, R. P. Fisher, and D. L. Bentley, "TFIIH-Associated Cdk7 Kinase Functions in Phosphorylation of C-Terminal Domain Ser7 Residues, Promoter-Proximal Pausing, and Termination by RNA Polymerase II," *Molecular and Cellular Biology*, vol. 29, pp. 5455–5464, oct 2009.

[424] M. Kim, H. Suh, E. J. Cho, and S. Buratowski, "Phosphorylation of the yeast Rpb1 C-terminal domain at serines 2,5, and 7," *Journal of Biological Chemistry*, vol. 284, pp. 26421–26426, sep 2009.

[425] L. M. Soares and S. Buratowski, "Yeast Swd2 is essential because of antagonism between Set1 histone methyltransferase complex and APT (associated with Pta1) termination factor," *Journal of Biological Chemistry*, vol. 287, pp. 15219–15231, may 2012.

[426] R. Magnus, N. Friis, and M. C. Schultz, "The timing is right," *Nature Structural & Molecular Biology*, vol. 21, no. 10, pp. 846–847, 2014.

[427] C. Vilela, C. V. Ramirez, B. Linz, C. Rodrigues-Pousada, and J. E. McCarthy, "Post-termination ribosome interactions with the 5' utr modulate yeast mrna stability," *The EMBO journal*, vol. 18, no. 11, pp. 3139–3152, 1999.

[428] L. Melin, H. Friden, E. Dehlin, L. Rutberg, and A. Von Gabain, "The importance of the 5'-region in regulating the stability of sdh mrna in bacillus subtilis," *Molecular microbiology*, vol. 4, no. 11, pp. 1881–1889, 1990.

[429] L. Cai, C. K. Dalal, and M. B. Elowitz, "Frequency-modulated nuclear localization bursts coordinate gene regulation," *Nature*, vol. 455, no. 7212, pp. 485–490, 2008.

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

[431] L. David, W. Huber, M. Granovskaia, J. Toedling, C. J. Palm, L. Bofkin, T. Jones, R. W. Davis, and L. M. Steinmetz, "A high-resolution map of transcription in the yeast genome," *Proceedings of the National Academy of Sciences*, vol. 103, pp. 5320–5325, apr 2006.

[432] B. Dujon, D. Sherman, G. Fischer, P. Durrens, S. Casaregela, I. Lafentaine, J. De Montigny, C. Marck, C. Neuvéglise, E. Talla, N. Goffard, L. Frangeul, M. Algie, V. Anthouard, A. Babour, V. Barbe, S. Barnay, S. Blanchin, J. M. Beckerich, E. Beyne, C. Bleykasten, A. Boisramé, J. Boyer, L. Cattolico, F. Confanioleri, A. De Daruvar, L. Despons, E. Fabre, C. Fairhead, H. Ferry-Dumazet, A. Groppi, F. Hantraye, C. Hennequin, N. Jauniaux, P. Joyot, R. Kachouri, A. Kerrest, R. Koszul, M. Lemaire, I. Lesur, L. Ma, H. Muller, J. M. Nicaud, M. Nikolski, S. Oztas, O. Ozier-Kalogeropoulos, S. Pellenz, S. Potter, G. F. Richard, M. L. Straub, A. Suleau, D. Swennen, F. Tekai, M. Wésolowski-Louvel, E. Westhof, B. Wirth, M. Zeniou-Meyer, I. Zivanovic, M. Bolotin-Fukuhara, A. Thierry, C. Bouchter, B. Caudron, C. Scarpelli,

C. Gaillardin, J. Weissebach, P. Wincker, and J. L. Souciet, "Genome evolution in yeasts," *Nature*, vol. 430, pp. 35–44, jul 2004.

[433] W. Lee, D. Tillo, N. Bray, R. H. Morse, R. W. Davis, T. R. Hughes, and C. Nislow, "A high-resolution atlas of nucleosome occupancy in yeast," *Nature Genetics*, vol. 39, pp. 1235–1244, oct 2007.

[434] B. T. Wilhelm and J.-R. Landry, "Rna-seq quantitative measurement of expression through massively parallel rna-sequencing," *Methods*, vol. 48, no. 3, pp. 249–257, 2009.

[435] S. Zhao, W.-P. Fung-Leung, A. Bittner, K. Ngo, and X. Liu, "Comparison of RNA-Seq and microarray in transcriptome profiling of activated T cells," *PloS one*, vol. 9, no. 1, 2014.

[436] C. Wang, B. Gong, P. R. Bushel, J. Thierry-Mieg, D. Thierry-Mieg, J. Xu, H. Fang, H. Hong, J. Shen, Z. Su, *et al.*, "The concordance between RNA-seq and microarray data depends on chemical treatment and transcript abundance," *Nature biotechnology*, vol. 32, no. 9, p. 926, 2014.

[437] Y. Liu, M. Morley, J. Brandimarto, S. Hannenhalli, Y. Hu, E. A. Ashley, W. W. Tang, C. S. Moravec, K. B. Margulies, T. P. Cappola, *et al.*, "RNA-Seq identifies novel myocardial gene expression signatures of heart failure," *Genomics*, vol. 105, no. 2, pp. 83–89, 2015.

[438] J. Li, R. Hou, X. Niu, R. Liu, Q. Wang, C. Wang, X. Li, Z. Hao, G. Yin, and K. Zhang, "Comparison of microarray and rna-seq analysis of mrna expression in dermal mesenchymal stem cells," *Biotechnology letters*, vol. 38, no. 1, pp. 33–41, 2016.

[439] I. M. Hall, K.-i. Noma, and S. I. Grewal, "Rna interference machinery regulates chromosome dynamics during mitosis and meiosis in fission yeast," *Proceedings of the National Academy of Sciences*, vol. 100, no. 1, pp. 193–198, 2003.

[440] H. M. Al-Hashimi and N. G. Walter, "Rna dynamics: it is about time," *Current opinion in structural biology*, vol. 18, no. 3, pp. 321–329, 2008.

[441] D. L. Beach, E. Salmon, and K. Bloom, "Localization and anchoring of mrna in budding yeast," *Current Biology*, vol. 9, no. 11, pp. 569–S1, 1999.

[442] E. A. Dethoff, J. Chugh, A. M. Mustoe, and H. M. Al-Hashimi, "Functional complexity and regulation through rna dynamics," *Nature*, vol. 482, no. 7385, pp. 322–330, 2012.

[443] E. M. Southern, "DNA microarrays. History and overview," in *Methods in molecular biology*, vol. 170, pp. 1–15, New Jersey: Humana Press, 2001.

[444] S. Jayanthi and C. R. Robin, "A survey on different classification methods for microarray data analysis," *Advances in Environmental Biology*, vol. 11, pp. 13–16, may 2017.

[445] K. Luger, A. W. Mäder, R. K. Richmond, D. F. Sargent, and T. J. Richmond, "Crystal structure of the nucleosome core particle at 2.8 å resolution," *Nature*, vol. 389, no. 6648, pp. 251–260, 1997.

[446] J. F. Nijkamp, M. van den Broek, E. Datema, S. de Kok, L. Bosman, M. A. Luttik, P. Daran-Lapujade, W. Vongsangnak, J. Nielsen, W. H. Heijne, P. Klaassen, C. J. Paddon, D. Platt, P. Kotter, R. C. van Ham, M. J. Reinders, J. T. Pronk, D. D. Ridder, and J.-M. Daran, "De novo sequencing, assembly and analysis of the genome of the laboratory strain Saccharomyces cerevisiae CEN.PK113-7D, a model for modern industrial biotechnology," *Microbial Cell Factories*, vol. 11, no. 1, p. 36, 2012.

[447] J. van Dijken, J. Bauer, L. Brambilla, P. Duboc, J. Francois, C. Gancedo, M. Giuseppin, J. Heijnen, M. Hoare, H. Lange, E. Madden, P. Niederberger, J. Nielsen, J. Parrou, T. Petit, D. Porro, M. Reuss, N. van Riel, M. Rizzi, H. Steensma, C. Verrips, J. Vindeløv, and J. Pronk, "An interlaboratory comparison of physiological and genetic properties of four Saccharomyces cerevisiae strains," *Enzyme and Microbial Technology*, vol. 26, no. 9, pp. 706–714, 2000.

[448] E. F. Lock and D. B. Dunson, "Bayesian consensus clustering," *Bioinformatics*, vol. 29, pp. 2610–2616, 08 2013.

[449] B. Andreopoulos, A. An, X. Wang, and M. Schroeder, "A roadmap of clustering algorithms: finding a match for a biomedical application," *Briefings in Bioinformatics*, vol. 10, pp. 297–314, 02 2009.

[450] J. Oyelade, I. Isewon, F. Oladipupo, O. Aromolaran, E. Uwoghiren, F. Ameh, M. Achas, and E. Adebiyi, "Clustering algorithms: their application to gene

expression data," *Bioinformatics and Biology insights*, vol. 10, pp. BBI–S38316, 2016.

[451] X. Z. Fern and W. Lin, "Cluster ensemble selection," *Statistical Analysis and Data Mining: The ASA Data Science Journal*, vol. 1, no. 3, pp. 128–141, 2008.

[452] G. Tsoumakas, I. Partalas, and I. Vlahavas, "A taxonomy and short review of ensemble selection," in *Workshop on Supervised and Unsupervised Ensemble Methods and Their Applications*, pp. 1–6, 2008.

[453] J. Ghosh and A. Acharya, "Cluster ensembles," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 1, no. 4, pp. 305–315, 2011.

[454] S. Vega-Pons and J. Ruiz-Shulcloper, "A survey of clustering ensemble algorithms," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 25, no. 03, pp. 337–372, 2011.

[455] S. Sarumathi, N. Shanthi, and M. Sharmila, "A review: Comparative analysis of different categorical data clustering ensemble methods," *International Journal of Computer and Information Engineering*, vol. 7, no. 12, pp. 1622–1632, 2014.

[456] Y. Chen, V. Kamat, E. R. Dougherty, M. L. Bittner, P. S. Meltzer, and J. M. Trent, "Ratio statistics of gene expression levels and applications to microarray data analysis," *Bioinformatics*, vol. 18, pp. 1207–1215, aug 2002.

[457] W. Pan, "A comparative review of statistical methods for discovering differentially expressed genes in replicated microarray experiments," *Bioinformatics*, vol. 18, pp. 546–554, apr 2002.

[458] M. D. Robinson, D. J. McCarthy, and G. K. Smyth, "edgeR: A Bioconductor package for differential expression analysis of digital gene expression data," *Bioinformatics*, vol. 26, pp. 139–140, jan 2009.

[459] N. Li, T. McMurry, A. Berg, Z. Wang, S. A. Berceli, and R. Wu, "Functional clustering of periodic transcriptional profiles through arma (p, q)," *PloS one*, vol. 5, no. 4, 2010.

[460] M. Shogren-Knaak, H. Ishii, J.-M. Sun, M. J. Pazin, J. R. Davie, and C. L. Peterson, "Histone h4-k16 acetylation controls chromatin structure and protein interactions," *Science*, vol. 311, no. 5762, pp. 844–847, 2006.

[461] O. Bell, C. Wirbelauer, M. Hild, A. N. Scharf, M. Schwaiger, D. M. MacAlpine, F. Zilbermann, F. Van Leeuwen, S. P. Bell, A. Imhof, *et al.*, "Localized h3k36 methylation states define histone h4k16 acetylation during transcriptional elongation in drosophila," *The EMBO journal*, vol. 26, no. 24, pp. 4974–4984, 2007.

[462] A. Zippo, R. Serafini, M. Rocchigiani, S. Pennacchini, A. Krepelova, and S. Oliviero, "Histone crosstalk between h3s10ph and h4k16ac generates a histone code that mediates transcription elongation," *Cell*, vol. 138, no. 6, pp. 1122–1136, 2009.

[463] J. Zhong, L. Ji, H. Chen, X. Li, J. Zhang, X. Wang, W. Wu, Y. Xu, F. Huang, W. Cai, *et al.*, "Acetylation of hmof modulates h4k16ac to regulate dna repair genes in response to oxidative stress," *International journal of biological sciences*, vol. 13, no. 7, p. 923, 2017.

[464] B. Briney, K. Le, J. Zhu, and D. R. Burton, "Clonify: unseeded antibody lineage assignment from next-generation sequencing data," *Scientific reports*, vol. 6, no. 1, pp. 1–10, 2016.

[465] S. G. Fonseca, F. Urano, G. C. Weir, J. Gromada, and M. Burcin, "Retraction note: Wolfram syndrome 1 and adenylyl cyclase 8 interact at the plasma membrane to regulate insulin production and secretion," *Nature Cell Biology*, vol. 17, no. 1, pp. 105–105, 2015.

[466] D. Schmidt, M. D. Wilson, C. Spyrou, G. D. Brown, J. Hadfield, and D. T. Odom, "ChIP-seq: Using high-throughput sequencing to discover protein-DNA interactions," *Methods*, vol. 48, pp. 240–248, jul 2009.

[467] R. Ng and J. Han, "Efficient and Effective Clustering Methods for Spatial Data Mining," in *Proceedings of the 20th VLDB Conference*, pp. 144–155, 1994.

[468] Y. Tu, G. Stolovitzky, and U. Klein, "Quantitative noise analysis for gene expression microarray experiments," *Proceedings of the National Academy of Sciences*, vol. 99, pp. 14031–14036, oct 2002.

[469] J. Li, P. R. Bushel, T.-M. Chu, and R. D. Wolfinger, *Batch Effects and Noise in Microarray Experiments*. J. Wiley, 2009.

[470] C. Trapnell, D. G. Hendrickson, M. Sauvageau, L. Goff, J. L. Rinn, and L. Pachter, "Differential analysis of gene regulation at transcript resolution with RNA-seq," *Nature Biotechnology*, vol. 31, pp. 46–53, jan 2013.

[471] P. Brennecke, S. Anders, J. K. Kim, A. A. Kołodziejczyk, X. Zhang, V. Proserpio, B. Baying, V. Benes, S. A. Teichmann, J. C. Marioni, and M. G. Heisler, "Accounting for technical noise in single-cell RNA-seq experiments," *Nature Methods*, vol. 10, pp. 1093–1098, nov 2013.

[472] A. R. Rao and M. Pellegrini, "Regulation of the yeast metabolic cycle by transcription factors with periodic activities," *BMC systems biology*, vol. 5, no. 1, p. 160, 2011.

[473] M. M. Levy, "Fourier Transform Analysis," *Journal of the British Institution of Radio Engineers*, pp. 239–257, 1946.

[474] J. McClellan and T. Parks, "Eigenvalue and eigenvector decomposition of the discrete fourier transform," *IEEE Transactions on Audio and Electroacoustics*, vol. 20, no. 1, pp. 66–74, 1972.

[475] P. Duhamel, B. Piron, and J. M. Etcheto, "On computing the inverse dft," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 36, no. 2, pp. 285–286, 1988.

[476] U. De Lichtenberg, L. J. Jensen, A. Fausbøll, T. S. Jensen, P. Bork, and S. Brunak, "Comparison of computational methods for the identification of cell cycle-regulated genes," *Bioinformatics*, vol. 21, no. 7, pp. 1164–1171, 2005.

[477] X. Peng, R. K. M. Karuturi, L. D. Miller, K. Lin, Y. Jia, P. Kondu, L. Wang, L.-S. Wong, E. T. Liu, M. K. Balasubramanian, *et al.*, "Identification of cell cycle-regulated genes in fission yeast," *Molecular biology of the cell*, vol. 16, no. 3, pp. 1026–1042, 2005.

[478] T. Chen, H. L. He, and G. M. Church, "Modeling gene expression with differential equations," in *Biocomputing'99*, pp. 29–40, World Scientific, 1999.

[479] A. A. Ptitsyn, S. Zvonic, and J. M. Gimble, "Permutation test for periodicity in short time series data," in *BMC bioinformatics*, vol. 7, p. S10, Springer, 2006.

[480] K. W. Choo and W. Kong, "Identification of differentially expressed genes in multiple microarray experiments using discrete fourier transform," *Front Biosci*, vol. 12, pp. 1845–1851, 2007.

[481] P. Brockwell and R. Davis, *Time Series: Theory and Methods*. Springer, 1987.

[482] M. Frigo and S. G. Johnson, "The Fastest Fourier Transform in the West," tech. rep., MIT Laboratory for Computer Science, 1997.

[483] M. Frigo and S. G. Johnson, "FFTW: An adaptive software architecture for the FFT," in *IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, vol. 3, pp. 1381–1384, IEEE, 1998.

[484] S. John, L. A. Howe, S. T. Tafrov, P. A. Grant, R. Sternglanz, and J. L. Workman, "The something about silencing protein, Sas3, is the catalytic subunit of NuA3, a yTAF(II)30-containing HAT complex that interacts with the Spt16 subunit of the yeast CP (Cdc68/Pob3)-FACT complex," *Genes and Development*, vol. 14, pp. 1196–1208, may 2000.

[485] Y. Doyon, W. Selleck, W. S. Lane, S. Tan, and J. Côté, "Structural and functional conservation of the NuA4 histone acetyltransferase complex from yeast to humans.," *Molecular and cellular biology*, vol. 24, pp. 1884–96, mar 2004.

[486] H. Steinhaus, "Sur la division des corps materiels en parties," *Bulletin of the Polish Academy of Sciences*, vol. 4, no. 3, pp. 801–804, 1956.

[487] A. Likas, N. Vlassis, and J. J. Verbeek, "The global k-means clustering algorithm," *Pattern Recognition*, vol. 36, pp. 451–461, feb 2003.

[488] A. K. Jain, "Data clustering: 50 years beyond K-means," *Pattern Recognition Letters*, vol. 31, pp. 651–666, jun 2010.

[489] M. R. Van Dijk, A. P. Waters, and C. J. Janse, "Stable transfection of malaria parasite blood stages," *Science*, vol. 268, pp. 1358–1362, jun 1995.

[490] S. Tavazoie, J. D. Hughes, M. J. Campbell, R. J. Cho, and G. M. Church, "Systematic determination of genetic network architecture," *Nature Genetics*, vol. 22, pp. 281–285, jul 1999.

[491] A. McCallum, K. Nigam, and L. L. Ungar, "Efficient clustering of high-dimensional data sets with application to reference matching," *Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 169–178, 2000.

[492] J. Quackenbush, "Computational analysis of microarray data," *Nature Reviews Genetics*, vol. 2, pp. 418–427, jun 2001.

[493] O. G. Troyanskaya, K. Dolinski, A. B. Owen, R. B. Altman, and D. Botstein, "A Bayesian framework for combining heterogeneous data sources for gene function prediction (in Saccharomyces cerevisiae).," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 100, pp. 8348–53, jul 2003.

[494] K. Krishna, K. Ramakrishnan, and M. Thathachar, "Vector quantization using genetic k-means algorithm for image compression," in *Proceedings of ICICS, 1997 International Conference on Information, Communications and Signal Processing. Theme: Trends in Information Systems Engineering and Wireless Multimedia Communications (Cat.*, vol. 3, pp. 1585–1587, IEEE, 1997.

[495] D. Venet, "Matarray: a matlab toolbox for microarray data," *Bioinformatics*, vol. 19, no. 5, pp. 659–660, 2003.

[496] V. Franc and V. Hlavác, "Statistical pattern recognition toolbox for matlab," *Prague, Czech: Center for Machine Perception, Czech Technical University*, 2004.

[497] D. Arthur and S. Vassilvitskii, "K-Means++: the Advantages of Careful Seeding," *Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms*, vol. 8, pp. 1027–1025, 2007.

[498] G. Schwarz, "Estimating the dimension of a model," *The Annals of Statistics*, vol. 6, no. 2, pp. 461–464, 1978.

[499] S. P. Lloyd, "Least Squares Quantization in PCM," *IEEE Transactions on Information Theory*, vol. 28, no. 2, pp. 129–137, 1982.

[500] C. Fraley, "How Many Clusters? Which Clustering Method? Answers Via Model-Based Cluster Analysis," *The Computer Journal*, vol. 41, no. 8, pp. 578–588, 1998.

[501] F. Corpet, "Multiple sequence alignment with hierarchical clustering," *Nucleic Acids Research*, vol. 16, pp. 449–456, nov 1988.

[502] S. C. Johnson, "Hierarchical clustering schemes," *Psychometrika*, vol. 32, no. 3, pp. 241–254, 1967.

[503] R. Xu and D. Wunsch, "Survey of clustering algorithms," *IEEE Transactions on Neural Networks*, vol. 16, pp. 645–678, may 2005.

[504] A. Sturn, J. Quackenbush, and Z. Trajanoski, "Genesis: cluster analysis of microarray data," *Bioinformatics*, vol. 18, pp. 207–8, jan 2002.

[505] Y. Loewenstein, E. Portugaly, M. Fromer, and M. Linial, "Efficient algorithms for accurate hierarchical clustering of huge datasets: Tackling the entire protein space," *Bioinformatics*, vol. 24, pp. i41–i49, jul 2008.

[506] F. Rapaport, R. Khanin, Y. Liang, M. Pirun, A. Krek, P. Zumbo, C. E. Mason, N. D. Socci, and D. Betel, "Comprehensive evaluation of differential gene expression analysis methods for RNA-seq data.," *Genome Biology*, vol. 14, p. R95, sep 2013.

[507] X. X. Martin Ester, Hans-Peter Kriegel, Jorg Sander, "A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise," *KDD-96 Proceedings*, pp. 226–231, 1996.

[508] J. Macqueen, "Some methods for classification and analysis of multivariate observations," *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, vol. 1, no. 233, pp. 281–297, 1967.

[509] R. M. Cormack, "A review of classification," *Journal of the Royal Statistical Society: Series A (General)*, vol. 134, no. 3, pp. 321–353, 1971.

[510] G. Hamerly and C. Elkan, "Learning the k in kmeans," *Advances in neural information processing*, vol. 17, pp. 1–8, 2004.

[511] D. T. Pham, S. S. Dimov, and C. D. Nguyen, "Selection of K in K-means clustering," *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, vol. 219, no. 1, pp. 103–119, 2005.

[512] T. Ishioka, "An Expansion of x -Means for Automatically Determining the Optimal Number of Clusters — Progressive Iterations of Ã -Means and Merging of the Clusters —," *Proceedings of fourth IASTED International conference Computational Intelligence,2005 Canada*, pp. 91–96, 2005.

[513] M. Yan and K. Ye, "Determining the number of clusters using the weighted gap statistic," *Biometrics*, vol. 63, no. 4, pp. 1031–1037, 2007.

[514] K. Park and D. Kim, "Localized network centrality and essentiality in the yeast-protein interaction network," *Proteomics*, vol. 9, no. 22, pp. 5143–5154, 2009.

[515] M. J. Hubisz, D. Falush, M. Stephens, and J. K. Pritchard, "Inferring weak population structure with the assistance of sample group information," *Molecular Ecology Resources*, vol. 9, no. 5, pp. 1322–1332, 2009.

[516] P. McCullagh and J. Yang, "How many clusters?," *Bayesian Anal.*, vol. 3, pp. 101–120, 03 2008.

[517] S. Omatu, J. Neves, J. M. C. Rodríguez, J. F. D. P. Santana, and S. R. González, *Distributed computing and artificial intelligence*. Springer, 2014.

[518] R. C. Dubes, "How many clusters are best? - an experiment," *Pattern Recognition*, vol. 20, no. 6, pp. 645 – 663, 1987.

[519] S. Still and W. Bialek, "How many clusters? an information-theoretic perspective," *Neural Computation*, vol. 16, no. 12, pp. 2483–2506, 2004.

[520] T. M. Kodinariya and P. R. Makwana, "Review on determining number of cluster in k-means clustering," *International Journal*, vol. 1, no. 6, pp. 90–95, 2013.

[521] H. Akaike, "A New Look at the Statistical Model Identification," *IEEE Transactions on Automatic Control*, vol. 19, no. 6, pp. 716–723, 1974.

[522] P. J. Rousseeuw, "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis," *Journal of Computational and Applied Mathematics*, vol. 20, pp. 53–65, 1987.

[523] L. Kaufman and P. J. Rousseuw, "Finding Groups in Data: An Introduction to Cluster Analysis," *Biometrics*, vol. 47, no. 2, p. 788, 1991.

[524] E. R. Hruschka and T. F. Covoes, "Feature selection for cluster analysis: an approach based on the simplified silhouette criterion," in *International Conference on Computational Intelligence for Modelling, Control and Automation and International Conference on Intelligent Agents, Web Technologies and Internet Commerce (CIMCA-IAWTIC'06)*, vol. 1, pp. 32–38, IEEE, 2005.

[525] A. F. Famili, G. Liu, and Z. Liu, "Evaluation and optimization of clustering in gene expression data analysis," *Bioinformatics*, vol. 20, no. 10, pp. 1535–1545, 2004.

[526] R. Lletıé, M. Ortiz, L. Sarabia, and M. Sánchez, "Selecting variables for k-means cluster analysis by using a genetic algorithm that optimises the silhouettes," *Analytica Chimica Acta*, vol. 515, no. 1, pp. 87 – 100, 2004. Papers presented at the 5th Colloquium Chemiometricum Mediterraneum.

[527] S. Bandyopadhyay, A. Mukhopadhyay, and U. Maulik, "An improved algorithm for clustering gene expression data," *Bioinformatics*, vol. 23, no. 21, pp. 2859–2865, 2007.

[528] R. C. De Amorim and C. Hennig, "Recovering the number of clusters in data sets with noise features using feature rescaling factors," *Information Sciences*, vol. 324, pp. 126–145, 2015.

[529] R. L. Thorndike, "Who belongs in the family?," *Psychometrika*, vol. 18, no. 4, pp. 267–276, 1953.

[530] D. C. Hambrick and S. M. Schecter, "Turnaround strategies for mature industrial-product business units," *Academy of management Journal*, vol. 26, no. 2, pp. 231–248, 1983.

[531] D. J. Ketchen and C. L. Shook, "The application of cluster analysis in strategic management research: An analysis and critique," *Strategic Management Journal*, vol. 17, no. 6, pp. 441–458, 1996.

[532] M. S. Aldenderfer and R. K. Blashfield, "A review of clustering methods," 1984.

[533] E. B. Fowlkes and C. L. Mallows, "A method for comparing two hierarchical clusterings," *Journal of the American statistical association*, vol. 78, no. 383, pp. 553–569, 1983.

[534] L. Parsons, E. Haque, and H. Liu, "Subspace clustering for high dimensional data: a review," *Acm Sigkdd Explorations Newsletter*, vol. 6, no. 1, pp. 90–105, 2004.

[535] L. Denœud and A. Guénoche, "Comparison of distance indices between partitions," in *Data Science and Classification*, pp. 21–28, Springer, 2006.

[536] M. Meila, "Comparing clusterings - an information based distance," *Journal of Multivariate Analysis*, vol. 98, no. 5, pp. 873–895, 2007.

[537] D. Zhou, J. Li, and H. Zha, "A new Mallows distance based metric for comparing clusterings," *Proceedings of the 22nd international conference on Machine learning - ICML '05*, pp. 1028–1035, 2005.

[538] M. K. Goldberg, M. Hayvanovych, and M. Magdon-Ismail, "Measuring similarity between sets of overlapping clusters," *Proceedings - SocialCom 2010: 2nd IEEE International Conference on Social Computing*, pp. 303–308, 2010.

[539] S. K. Popat and M. Emmanuel, "Review and comparative study of clustering techniques," *International journal of computer science and information technologies*, vol. 5, no. 1, pp. 805–812, 2014.

[540] W. M. Rand, "Objective criteria for the evaluation of clustering methods," *Journal of the American Statistical association*, vol. 66, no. 336, pp. 846–850, 1971.

[541] S. Van Dongen, "Performance criteria for graph clustering and markov cluster experiments," in *CWI Centre for Mathematics and Computer Science*, Citeseer, 2000.

[542] E. Bae, J. Bailey, and G. Dong, "Clustering similarity comparison using density profiles," in *Australasian Joint Conference on Artificial Intelligence*, pp. 342–351, Springer, 2006.

[543] M. Rezaei and P. Fränti, "Set matching measures for external cluster validity," *IEEE Transactions on Knowledge and Data Engineering*, vol. 28, no. 8, pp. 2173–2186, 2016.

[544] R. Rabbany and O. R. Zaïane, "A general clustering agreement index: For comparing disjoint and overlapping clusters," in *Thirty-First AAAI Conference on Artificial Intelligence*, 2017.

[545] M. Meila, "Comparing clusterings," *University of Washington Department of Statistics*, 2002.

[546] M. Meilă, "Comparing clusterings by the variation of information," in *Learning theory and kernel machines*, pp. 173–187, Springer, 2003.

[547] S. Navlakha, J. White, N. Nagarajan, M. Pop, and C. Kingsford, "Finding biologically accurate clusterings in hierarchical tree decompositions using the variation of information," in *Annual International Conference on Research in Computational Molecular Biology*, pp. 400–417, Springer, 2009.

[548] A. Lancichinetti, S. Fortunato, and J. Kertész, "Detecting the overlapping and hierarchical community structure in complex networks," *New journal of physics*, vol. 11, no. 3, p. 033015, 2009.

[549] M. Meilă, "Comparing clusterings," *Proceedings of the 22nd international conference on Machine learning - ICML '05*, pp. 577–584, 2005.

[550] L. Hubert and P. Arabie, "Comparing partitions," *Journal of classification*, vol. 2, no. 1, pp. 193–218, 1985.

[551] J. Han, J. Pei, and M. Kamber, *Data mining: concepts and techniques*. Elsevier, 2011.

[552] W. De Mulder, "Optimal clustering in the context of overlapping cluster analysis," *Information Sciences*, vol. 223, pp. 56–74, 2013.

[553] U. Feige, "A threshold of ln n for approximating set cover," *Journal of the ACM (JACM)*, vol. 45, no. 4, pp. 634–652, 1998.

[554] P. Jaccard, "Distribution de la flore alpine dans le bassin des dranses et dans quelques régions voisines," *Bull Soc Vaudoise Sci Nat*, vol. 37, pp. 241–272, 1901.

[555] D. L. Wallace, "Comment," *Journal of the American Statistical Association*, vol. 78, pp. 569–576, 1983.

[556] J. Lerman, "Comparing partitions. mathematical and statistical aspects," in *Classification and Related Methods of Data Analysis*, pp. 121—-132, Elsevier, Amsterdam, 1988.

[557] R. W. Hamming, "Error detecting and error correcting codes," *The Bell system technical journal*, vol. 29, no. 2, pp. 147–160, 1950.

[558] M. M. Fréchet, "Sur quelques points du calcul fonctionnel," *Rendiconti del Circolo Matematico di Palermo (1884-1940)*, vol. 22, no. 1, pp. 1–72, 1906.

[559] F. Hausdorff, "Grundzüge einer theorie der geordneten mengen," *Mathematische Annalen*, vol. 65, no. 4, pp. 435–505, 1908.

[560] F. Hausdorff, "Dimension und äußeres maß," *Mathematische Annalen*, vol. 79, no. 1-2, pp. 157–179, 1918.

[561] W. Rucklidge, *Efficient visual recognition using the Hausdorff distance*. Springer-Verlag, 1996.

[562] P. C. Mahalanobis, "On the generalized distance in statistics," *Proceedings of the National Institute of Science of India*, vol. 12, pp. 49–55, 1936.

[563] R. De Maesschalck, D. Jouan-Rimbaud, and D. L. Massart, "The mahalanobis distance," *Chemometrics and intelligent laboratory systems*, vol. 50, no. 1, pp. 1–18, 2000.

[564] O. Fujita, "Metrics based on average distance between sets," *Japan Journal of Industrial and Applied Mathematics*, vol. 30, no. 1, pp. 1–19, 2013.

[565] L. Denœud, "Transfer distance between partitions," *Advances in Data Analysis and Classification*, vol. 2, no. 3, pp. 279–294, 2008.

[566] R. Mao, W. Xu, S. Ramakrishnan, G. Nuckolls, and D. P. Miranker, "On optimizing distance-based similarity search for biological databases," in *2005 IEEE Computational Systems Bioinformatics Conference (CSB'05)*, pp. 351–361, IEEE, 2005.

[567] G. Rossi, "Hamming distance between partitions, clustering comparison and information," in *International Conference on Pure Mathematics, Applied Mathematics and Computational Methods (PMAMCM 2015), Zakynthos Island, Greece*, 2015.

[568] M. Woodhams, D. A. Steane, R. C. Jones, D. Nicolle, V. Moulton, and B. R. Holland, "Novel distances for dollo data," *Systematic biology*, vol. 62, no. 1, pp. 62–77, 2013.

[569] S. Gog and R. Venturini, "Fast and compact hamming distance index," in *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval*, pp. 285–294, 2016.

[570] L. A. García-Escudero, A. Gordaliza, R. San Martín, S. Van Aelst, and R. Zamar, "Robust linear clustering," *Journal of the Royal Statistical Society. Series B*, vol. 71, pp. 301–318, jan 2009.

[571] X. Fu, N. Fu, S. Guo, Z. Yan, Y. Xu, H. Hu, C. Menzel, W. Chen, Y. Li, R. Zeng, and P. Khaitovich, "Estimating accuracy of RNA-Seq and microarrays with proteomics," *BMC Genomics*, vol. 10, p. 161, apr 2009.

[572] A. Oshlack, M. D. Robinson, and M. D. Young, "From RNA-seq reads to differential expression results," *Genome Biology*, vol. 11, p. 220, dec 2010.

[573] SEQC/MAQC-III Consortium, "A comprehensive assessment of RNA-seq accuracy, reproducibility and information content by the Sequencing Quality Control Consortium," *Nature Biotechnology*, vol. 32, no. 9, pp. 903–914, 2014.

[574] A. Conesa, P. Madrigal, S. Tarazona, D. Gomez-Cabrero, A. Cervera, A. McPherson, M. W. Szcześniak, D. J. Gaffney, L. L. Elo, X. Zhang, and A. Mortazavi, "A survey of best practices for RNA-seq data analysis," *Genome Biology*, vol. 17, no. 1, pp. 1–19, 2016.

[575] L. S. Churchman, "Not Just Noise: Genomics and Genetics Bring Long Noncoding RNAs into Focus," *Molecular Cell*, vol. 65, no. 1, pp. 1–2, 2017.

[576] P. Lefrançois, G. M. Euskirchen, R. K. Auerbach, J. Rozowsky, T. Gibson, C. M. Yellman, M. Gerstein, and M. Snyder, "Efficient yeast ChIP-Seq using multiplex short-read DNA sequencing," *BMC Genomics*, vol. 10, pp. 1–18, 2009.

[577] Sonia Tarazona, Fernando Garcia-Alcalde, S. Tarazona, F. Garcia-Alcalde, A. Ferrer, and A. Conesa, "Differential expression in RNA-seq: a matter of depth," *Genome Research*, vol. 21, pp. 2213–2223, 2011.

[578] M. I. Love, W. Huber, and S. Anders, "Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2," *Genome Biology*, vol. 15, no. 12, pp. 1–21, 2014.

[579] M. J. Nueda, A. Conesa, J. A. Westerhuis, H. C. J. Hoefsloot, A. K. Smilde, M. Talón, and A. Ferrer, "Discovering gene expression patterns in time course microarray experiments by ANOVA-SCA," *Bioinformatics*, vol. 23, no. 14, pp. 1792–1800, 2007.

[580] M. Love, S. Anders, and W. Huber, "Differential analysis of RNA-Seq data at the gene level using the DESeq package," *European Molecular Biology Laboratory*, vol. 10, pp. 1–32, 2013.

[581] J. Feng, C. A. Meyer, Q. Wang, J. S. Liu, X. S. Liu, and Y. Zhang, "GFOLD: A generalized fold change for ranking differentially expressed genes from RNA-seq data," *Bioinformatics*, vol. 28, no. 21, pp. 2782–2788, 2012.

[582] A. Mayer, J. Di Iulio, S. Maleri, U. Eser, J. Vierstra, A. Reynolds, R. Sandstrom, J. A. Stamatoyannopoulos, and L. S. Churchman, "Native elongating transcript sequencing reveals human transcriptional activity at nucleotide resolution," *Cell*, vol. 161, pp. 541–544, apr 2015.

[583] K. Y. Yeung, C. Fraley, A. Murua, A. E. Raftery, and W. L. Ruzzo, "Model-based clustering and data transformations for gene expression data," *Bioinformatics*, vol. 17, no. 10, pp. 977–987, 2001.

[584] F. de Macedo Dossin and S. Schenkman, "Actively transcribing rna polymerase ii concentrates on spliced leader genes in the nucleus of trypanosoma cruzi," *Eukaryotic cell*, vol. 4, no. 5, pp. 960–970, 2005.

[585] J. Quackenbush, "Microarray data normalization and transformation," *Nature Genetics*, vol. 32, no. 4S, pp. 496–501, 2002.

[586] D. Edwards, "Non-linear normalization and background correction in one-channel cDNA microarray studies," *Bioinformatics*, vol. 19, pp. 825–833, may 2003.

[587] S. C. Geller, J. P. Gregg, P. Hagerman, and D. M. Rocke, "Transformation and normalization of oligonucleotide microarray data," *Bioinformatics*, vol. 19, pp. 1817–1823, sep 2003.

[588] S. J. Callister, R. C. Barry, J. N. Adkins, E. T. Johnson, W. J. Qian, B. J. M. Webb-Robertson, R. D. Smith, and M. S. Lipton, "Normalization approaches for removing systematic biases associated with mass spectrometry and label-free proteomics," *Journal of Proteome Research*, vol. 5, no. 2, pp. 277–286, 2006.

[589] J. H. Bullard, K. Hansen, and S. Dudoit, "Evaluation of Statistical Methods for Normalization and Differential Expression in mRNA-Seq Experiments Evaluation of Statistical Methods for Normalization and Differential Expression in mRNA-Seq Experiments," *BMC Bioinformatics*, vol. 11, no. 1, p. 94, 2010.

[590] M. A. Dillies, A. Rau, J. Aubert, C. Hennequet-Antier, M. Jeanmougin, N. Servant, C. Keime, N. S. Marot, D. Castel, J. Estelle, G. Guernec, B. Jagla, L. Jouneau, D. Laloë, C. Le Gall, B. Schaëffer, S. Le Crom, M. Guedj, and F. Jaffrézic, "A comprehensive evaluation of normalization methods for Illumina high-throughput RNA sequencing data analysis," *Briefings in Bioinformatics*, vol. 14, no. 6, pp. 671–683, 2013.

[591] C. Zang, D. E. Schones, C. Zeng, K. Cui, K. Zhao, and W. Peng, "A clustering approach for identification of enriched domains from histone modification ChIP-Seq data," *Bioinformatics*, vol. 25, no. 15, pp. 1952–1958, 2009.

[592] M. Stonebraker, D. J. Abadi, A. Batkin, X. Chen, M. Cherniack, M. Ferreira, E. Lau, A. Lin, S. Madden, E. O'Neil, P. O'Neil, A. Rasin, N. Tran, and S. Zdonik, "C-store: a column-oriented DBMS," *VLDB Conference*, pp. 553–564, 2005.

[593] D. J. Abadi, P. A. Boncz, and S. Harizopoulos, "Column-oriented database systems," in *Proceedings of the VLDB*, vol. 2, (Lyon, France), pp. 1664–1665, ACM, 2009.

[594] H. Ding, Y. Liu, L. Huang, and J. Li, "K-Means Clustering with Distributed Dimensions," in *Proceedings of the 33rd International Conference on Machine Learning*, vol. 48, (New York, NY, USA), 2016.

[595] M. Boussahoua, O. Boussaid, and F. Bentayeb, "Logical schema for data warehouse on column-oriented nosql databases," in *International Conference on Database and Expert Systems Applications*, pp. 247–256, Springer, 2017.

[596] H. Ku, "Notes on the use of propagation of error formulas," *Journal of Research of the National Bureau of Standards, Section C: Engineering and Instrumentation*, vol. 70C, p. 263, oct 1966.

[597] L. A. Goodman, "On the Exact Variance of Products," *Journal of the American Statistical Association*, vol. 55, pp. 708–713, dec 1960.

[598] C. Lecomte, "Exact statistics of systems with uncertainties: An analytical theory of rank-one stochastic dynamic systems," *Journal of Sound and Vibration*, vol. 332, pp. 2750–2776, may 2013.

[599] Z. Xu, W. Wei, J. Gagneur, F. Perocchi, S. Clauder-Münster, J. Camblong, E. Guffanti, F. Stutz, W. Huber, and L. M. Steinmetz, "Bidirectional promoters generate pervasive transcription in yeast," *Nature*, vol. 457, no. 7232, pp. 1033–1037, 2009.

[600] A. J. Goodman, E. R. Daugharthy, and J. Kim, "Pervasive antisense transcription is evolutionarily conserved in budding yeast," *Molecular biology and evolution*, vol. 30, no. 2, pp. 409–421, 2012.

[601] H. Thorvaldsdóttir, J. T. Robinson, and J. P. Mesirov, "Integrative Genomics Viewer (IGV): High-performance genomics data visualization and exploration," *Briefings in Bioinformatics*, vol. 14, no. 2, pp. 178–192, 2013.

[602] James T Robinson, H. Thorvaldsdóttir, W. Winckler, M. Guttman, E. S. Lander, G. Getz, J. P. Mesirov, and IGV (Integrative Genomic Viewer), "Integrative Genomics Viewer," *Broad Institute*, vol. 29, no. 1, pp. 24–26, 2013.

[603] E. R. Gansner and S. C. North, "An open graph visualization system and its applications to software engineering," *Software: practice and experience*, vol. 30, no. 11, pp. 1203–1233, 2000.

[604] F. A. Haight, *Hanbook of the Poisson Distribution*, vol. 18. Taylor & Francis, dec 1967.

[605] I. Barta and R. Iggo, "Autoregulation of expression of the yeast Dbp2p 'DEAD-box' protein is mediated by sequences in the conserved DBP2 intron.," *The EMBO journal*, vol. 14, no. 15, pp. 3800–8, 1995.

[606] W. K. Ma, S. C. Cloutier, and E. J. Tran, "The DEAD-box protein Dbp2 functions with the RNA-binding protein Yra1 to promote mRNP assembly," *Journal of Molecular Biology*, vol. 425, no. 20, pp. 3824–3838, 2013.

[607] Z. Xing, S. Wang, and E. J. Tran, "Characterization of the mammalian DEAD-box protein DDX5 reveals functional conservation with {\it S. cerevisiae} ortholog Dbp2 in transcriptional control and glucose metabolism," *Rna*, vol. 23, no. 7, pp. 1125–1138, 2017.

[608] S. Wang, Z. Xing, P. E. Pascuzzi, and E. J. Tran, "Metabolic Adaptation to Nutrients Involves Coregulation of Gene Expression by the RNA Helicase Dbp2 and the Cyc8 Corepressor in Saccharomyces cerevisiae," *G3: Genes—Genomes—Genetics*, vol. 7, no. 7, pp. 2235–2247, 2017.

[609] S. C. Cloutier, S. Wang, W. K. Ma, N. Al Husini, Z. Dhoondia, A. Ansari, P. E. Pascuzzi, and E. J. Tran, "Regulated Formation of lncRNA-DNA Hybrids Enables Faster Transcriptional Induction and Environmental Adaptation," *Molecular Cell*, vol. 61, no. 3, pp. 393–404, 2016.

[610] R. A. Cartwright and D. Graur, "The multiple personalities of Watson and Crick strands," *Biology Direct*, vol. 6, p. 7, feb 2011.

[611] P. Grandi, V. Rybin, J. Baßler, E. Petfalski, D. Strauß, M. Marzioch, T. Schäfer, B. Kuster, H. Tschochner, D. Tollervey, A. C. Gavin, and E. Hurt, "90S pre-ribosomes include the 35S pre-rRNA, the U3 snoRNP, and 40S subunit processing factors but predominantly lack 60S synthesis factors," *Molecular Cell*, vol. 10, no. 1, pp. 105–115, 2002.

[612] E. Thomson, J. Rappsilber, and D. Tollervey, "Nop9 is an RNA binding protein present in pre-40S ribosomes and required for 18S rRNA synthesis in yeast," *Rna*, vol. 13, pp. 2165–2174, 2007.

[613] J. E. G. Gallagher, D. A. Dunbar, S. Granneman, B. M. Mitchell, Y. Osheim, A. L. Beyer, and S. J. Baserga, "RNA polymerase I transcription and pre-rRNA processing are linked by specific SSU processome components," *Genes and Development*, vol. 18, no. 20, pp. 2507–2517, 2004.

[614] J. Perez-Fernandez, A. Roman, J. De Las Rivas, X. R. Bustelo, and M. Dosil, "The 90S Preribosome Is a Multimodular Structure That Is Assembled through a Hierarchical Mechanism," *Molecular and Cellular Biology*, vol. 27, no. 15, pp. 5414–5429, 2007.

[615] K. A. Bernstein, J. E. G. Gallagher, B. M. Mitchell, S. Granneman, and S. J. Baserga, "The small-subunit processome is a ribosome assembly intermediate," *Eukaryotic Cell*, vol. 3, no. 6, pp. 1619–1626, 2004.

[616] V. Bevilacqua, G. Mastronardi, F. Menolascina, A. Paradiso, and S. Tommasi, "Genetic algorithms and artificial neural networks in microarray data analysis: A distributed approach.," *Engineering Letters*, vol. 13, no. 4, 2006.

[617] B. Pontes, R. Giraldez, and J. S. Aguilar-Ruiz, "Biclustering on expression data: A review," *Journal of biomedical informatics*, vol. 57, pp. 163–180, 2015.

[618] E. Weitschek, G. Fiscon, V. Fustaino, G. Felici, and P. Bertolazzi, "Clustering and classification techniques for gene expression profile pattern analysis," *Pattern Recognition in Computational Molecular Biology: Techniques and Approaches. Hoboken: Wiley*, p. 347, 2015.

[619] Y. L. Chen and H. L. Hu, "An overlapping cluster algorithm to provide non-exhaustive clustering," *European Journal of Operational Research*, vol. 173, no. 3, pp. 762–780, 2006.

[620] J. J. Whang, I. Dhillon, and D. Gleich, "Non-exhaustive, Overlapping k - means," *Proceedings of the SIAM International Conference on Data Mining*, pp. 936–944, 2015.

[621] C. Bonferroni, "Sulle medie multiple di potenze," *Bollettino dell'Unione Matematica Italiana*, vol. 5, no. 3-4, pp. 267–270, 1950.

[622] Y. Benjamini and Y. Hochberg, "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing," *Journal of the Royal Statistical Society Series B*, vol. 57, no. 1, pp. 289–300, 1995.

[623] J. Green and N. Britten, "Qualitative research and evidence based medicine," *British Medical Journal*, vol. 316, no. April, pp. 1230–1232, 1998.

[624] F. Estruch and M. Carlson, "Two homologous zinc finger genes identified by multicopy suppression in a SNF1 protein kinase mutant of Saccharomyces cerevisiae.," *Molecular and cellular biology*, vol. 13, no. 7, pp. 3872–81, 1993.

[625] F. Estruch, "Stress-controlled transcription factors, stress-induced genes and stress tolerance in budding yeast," *FEMS Microbiology Reviews*, vol. 24, no. 4, pp. 469–486, 2000.

[626] D. Thomas and Y. Surdin-Kerjan, "Metabolism of sulfur amino acids in Saccharomyces cerevisiae," *Microbiology and molecular biology reviews*, vol. 61, no. 4, pp. 503–532, 1997.

[627] S. L. Chin, I. M. Marcus, R. R. Klevecz, and C. M. Li, "Dynamics of oscillatory phenotypes in Saccharomyces cerevisiae reveal a network of genome-wide transcriptional oscillators," *FEBS Journal*, vol. 279, no. 6, pp. 1119–1130, 2012.

[628] J. Wu, N. Zhang, A. Hayes, K. Panoutsopoulou, and S. G. Oliver, "Global analysis of nutrient control of gene expression in Saccharomyces cerevisiae during growth and starvation.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 9, pp. 3148–53, 2004.

[629] A. P. Gasch, P. T. Spellman, C. M. Kao, Carmel-Harel, M. B. Eisen, G. Storz, D. Botstein, and P. O. Brown, "Genomic Expression Programs in the Response of Yeast Cells to Environmental Changes," *Mol Biol Cell*, vol. 11(12), no. December, pp. 4241–4257, 2000.

[630] M. Huang, J. Bao, B. M. Hallström, D. Petranovic, and J. Nielsen, "Efficient protein production by yeast requires global tuning of metabolism," *Nature Communications*, vol. 8, no. 1, 2017.

[631] T. I. Lee, N. M. Hannett, C. T. Harbison, C. M. Thompson, I. Simon, J. Zeitlinger, E. G. Jennings, H. L. Murray, D. B. Gordon, B. Ren, J. J. Wyrick, J.-b. Tagne, and R. A. Young, "Transcriptional Regulatory Networks in Saccharomyces cerevisiae," *Science*, vol. 298, no. 5594, pp. 799–804, 2002.

[632] S. Hohmann and P. A. Meacock, "Thiamin metabolism and thiamin diphosphate-dependent enzymes in the yeast Saccharomyces cerevisiae: genetic regulation," *Biochimica et Biophysica Acta - Protein Structure and Molecular Enzymology*, vol. 1385, no. 2, pp. 201–219, 1998.

[633] M. J. Brauer, C. Huttenhower, E. M. Airoldi, R. Rosenstein, J. C. Matese, D. Gresham, V. M. Boer, O. G. Troyanskaya, and D. Botstein, "Coordination of Growth Rate, Cell Cycle, Stress Response, and Metabolic Activity in Yeast," *Molecular biology of the cell*, vol. 19, no. January, pp. 352–367, 2008.

[634] E. E. Lilja and D. R. Johnson, "Segregating metabolic processes into different microbial cells accelerates the consumption of inhibitory substrates," *ISME Journal*, vol. 10, pp. 1568–1578, jul 2016.

[635] S. W. Edwards and D. Lloyd, "Oscillations in protein and RNA content during synchronous growth of Acanthamoeba castellanii. Evidence for periodic turnover of macromolecules during the cell cycle," *FEBS Letters*, vol. 109, pp. 21–26, jan 1980.

[636] K. B. Hallberg and D. B. Johnson, "Biodiversity of acidophilic prokaryotes," *Advances in applied microbiology*, vol. 49, pp. 37–84, 2001.

[637] P. D. Cotter and C. Hill, "Surviving the acid test: responses of gram-positive bacteria to low ph," *Microbiol. Mol. Biol. Rev.*, vol. 67, no. 3, pp. 429–453, 2003.

[638] N. Beales, "Adaptation of Microorganisms to Cold Temperatures, Weak Acid Preservatives, Low pH, and Osmotic Stress: A Review," *Comprehensive Reviews in Food Science and Food Safety*, vol. 3, pp. 1–20, jan 2004.

[639] O. V. Golyshina and K. N. Timmis, "Ferroplasma and relatives, recently discovered cell wall-lacking archaea making a living in extremely acid, heavy metal-rich environments," *Environmental Microbiology*, vol. 7, pp. 1277–1288, sep 2005.

[640] L. S. Robertson and G. R. Fink, "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*, vol. 95, pp. 13783–7, nov 1998.

[641] J. P. Gonçalves, A. P. Francisco, N. P. Mira, M. C. Teixeira, I. Sá-Correia, A. L. Oliveira, and S. C. Madeira, "TFRank: Network-based prioritization of regulatory associations underlying transcriptional responses," *Bioinformatics*, vol. 27, no. 22, pp. 3149–3157, 2011.

[642] K. Nasmyth, "At the heart of the budding yeast cell cycle," *TIG*, vol. 10, no. October, pp. 405–412, 1996.

[643] N. L. Maas, K. M. Miller, L. G. DeFazio, and D. P. Toczyski, "Cell Cycle and Checkpoint Regulation of Histone H3 K56 Acetylation by Hst3 and Hst4," *Molecular Cell*, vol. 23, pp. 109–119, jul 2006.

[644] N. Zhang, J. Wu, and S. G. Oliver, "Gis1 is required for transcriptional reprogramming of carbon metabolism and the stress response during Transition into stationary phase in yeast," *Microbiology*, vol. 155, pp. 1690–1698, may 2009.

[645] Y. Pan, E. A. Schroeder, A. Ocampo, A. Barrientos, and G. S. Shadel, "Regulation of yeast chronological life span by TORC1 via adaptive mitochondrial ROS signaling," *Cell Metabolism*, vol. 13, pp. 668–678, jun 2011.

[646] N. Soranzo, M. Zampieri, L. Farina, and C. Altafini, "mRNA stability and the unfolding of gene expression in the long-period yeast metabolic cycle," *BMC Systems Biology*, vol. 3, p. 18, feb 2009.

[647] N. Soontorngun, "Reprogramming of nonfermentative metabolism by stress-responsive transcription factors in the yeast Saccharomyces cerevisiae," *Current Genetics*, vol. 63, pp. 1–7, feb 2017.

[648] B. Mai and L. Breeden, "Xbp1, a stress-induced transcriptional repressor of the Saccharomyces cerevisiae Swi4/Mbp1 family.," *Molecular and cellular biology*, vol. 17, pp. 6491–6501, nov 1997.

[649] Y. Pan, C. C. Dacso, B. W. O. Malley, B. Zhu, Q. Zhang, Y. Pan, E. M. Mace, B. York, and A. C. Antoulas, "A Cell-Autonomous Mammalian 12 hr Clock Coordinates Metabolic and Stress Rhythms Article A Cell-Autonomous Mammalian 12 hr Clock Coordinates Metabolic and Stress Rhythms," *Cell Metabolism*, vol. 25, no. 6, pp. 1305–1319.e9, 2017.

[650] F. S. Howe, H. Fischl, S. C. Murray, and J. Mellor, "Is h3k4me3 instructive for transcription activation?," *Bioessays*, vol. 39, no. 1, pp. 1–12, 2017.

[651] S. C. Murray, P. Lorenz, F. Howe, M. Wouters, T. Brown, S. Xi, H. Fischl, W. Khushaim, J. R. Rayappu, A. Angel, *et al.*, "H3k4me3 is neither instructive for, nor informed by, transcription.," *bioRxiv*, p. 709014, 2019.

[652] L. A. Gates, C. E. Foulds, and B. W. O'Malley, "Histone marks in the 'driver's seat': functional roles in steering the transcription cycle," *Trends in biochemical sciences*, vol. 42, no. 12, pp. 977–989, 2017.

[653] J. D. Murray, *Mathematical biology*, vol. 17 of *Interdisciplinary Applied Mathematics*. New York, NY: Springer-Verlag New York, 2002.

[654] S. Strogatz, *Nonlinear Dynamics and Chaos*. Boca Raton: CRC Press, 2nd ed., may 2018.

[655] M. A. Savageau, "Biochemical systems analysis: Ii. the steady-state solutions for an n-pool system using a power-law approximation," *Journal of theoretical biology*, vol. 25, no. 3, pp. 370–379, 1969.

[656] R. Kronauer, C. Czeisler, S. Pilato, D. Elliot, M. Moore-Ede, E. Weitzman, and R. Weber, "Mathematical model of the human circadian system with two interacting oscillators," *Invited Opinions*, pp. 1–17, 1982.

[657] A. Goldbeter, "A model for circadian oscillations in the Drosophila period protein (PER)," *Proceedings of the Royal Society B: Biological Sciences*, vol. 261, pp. 319–324, sep 1995.

[658] T. Scheper, D. Klinkenberg, C. Pennartz, and J. van Pelt, "A mathematical model for the intracellular circadian rhythm generator," *The Journal of neuroscience*, vol. 19, no. 1, pp. 40–47, 1999.

[659] C. Oguz, T. Laomettachit, K. C. Chen, L. T. Watson, W. T. Baumann, and J. J. Tyson, "Optimization and model reduction in the high dimensional parameter space of a budding yeast cell cycle model.," *BMC systems biology*, vol. 7, p. 53, 2013.

[660] B. Novák and J. J. Tyson, "Design principles of biochemical oscillators," *Nature Reviews Molecular Cell Biology*, vol. 9, no. December, pp. 981–991, 2008.

[661] V. Černý, "Thermodynamical approach to the traveling salesman problem: An efficient simulation algorithm," *Journal of Optimization Theory and Applications*, vol. 45, no. 1, pp. 41–51, 1985.

[662] N. A. W. van Riel, "Dynamic modelling and analysis of biochemical networks: Mechanism-based models and model-based experiments," *Briefings in Bioinformatics*, vol. 7, no. 4, pp. 364–374, 2006.

[663] M. Rizzi, M. Baltes, U. Theobald, and M. Reuss, "In vivo analysis of metabolic dynamics in Saccharomyces cerevisiae: II. Mathematical model," *Biotechnology and Bioengineering*, vol. 55, no. 4, pp. 592–608, 1997.

[664] F. Crick, "Central dogma of molecular biology," *Nature*, vol. 227, pp. 561–3, aug 1970.

[665] P. Tapaswi and P. Bhattacharya, "An extended mathematical model of transcription and translation during embryogenesis," *Revue Cybernetica*, vol. 24, no. 1, 1981.

[666] H. Smith, "Oscillations and multiple steady states in a cyclic gene model with repression," *Journal of mathematical biology*, vol. 25, no. 2, pp. 169–190, 1987.

[667] E. L. Weitzke and P. J. Ortoleva, "Simulating cellular dynamics through a coupled transcription, translation, metabolic model," *Computational Biology and Chemistry*, vol. 27, no. 4-5, pp. 469–480, 2003.

[668] G. V. Demidenko and V. A. Likhoshvai, "On differential equations with retarded argument," *Siberian Mathematical Journal*, vol. 46, no. 3, pp. 417–430, 2005.

[669] K. Rounseville and P. Chau, "Three-dimensional cell cycle model with distributed transcription and translation," *Medical and Biological Engineering and Computing*, vol. 43, no. 1, pp. 155–161, 2005.

[670] S. Kosuri, D. B. Goodman, G. Cambray, V. K. Mutalik, Y. Gao, A. P. Arkin, D. Endy, and G. M. Church, "Composability of regulatory sequences controlling transcription and translation in escherichia coli," *Proceedings of the National Academy of Sciences*, vol. 110, no. 34, pp. 14024–14029, 2013.

[671] M. R. Roussel and R. Zhu, "Validation of an algorithm for delay stochastic simulation of transcription and translation in prokaryotic gene expression," *Physical biology*, vol. 3, no. 4, p. 274, 2006.

[672] L. Qian, H. Wang, and E. R. Dougherty, "Inference of noisy nonlinear differential equation models for gene regulatory networks using genetic programming and kalman filtering," *IEEE Transactions on Signal Processing*, vol. 56, no. 7, pp. 3327–3339, 2008.

[673] S. B.-T. de Leon and E. H. Davidson, "Modeling the dynamics of transcriptional gene regulatory networks for animal development," *Developmental biology*, vol. 325, no. 2, pp. 317–328, 2009.

[674] R. Edwards, A. Machina, G. McGregor, and P. Van den Driessche, "A modelling framework for gene regulatory networks including transcription and translation," *Bulletin of mathematical biology*, vol. 77, no. 6, pp. 953–983, 2015.

[675] M. P. Washburn, D. Wolters, and J. R. Yates, "Large-scale analysis of the yeast proteome by multidimensional protein identification technology," *Nature Biotechnology*, vol. 19, pp. 242–247, mar 2001.

[676] P. Picotti, B. Bodenmiller, L. N. Mueller, B. Domon, and R. Aebersold, "Full Dynamic Range Proteome Analysis of S. cerevisiae by Targeted Proteomics," *Cell*, vol. 138, pp. 795–806, aug 2009.

[677] A. Whitmarsh and R. Davis, "Regulation of transcription factor function by phosphorylation," *Cellular and Molecular Life Sciences CMLS*, vol. 57, no. 8-9, pp. 1172–1183, 2000.

[678] M. J. Buck and J. D. Lieb, "A chromatin-mediated mechanism for specification of conditional transcription factor targets," *Nature genetics*, vol. 38, no. 12, pp. 1446–1451, 2006.

[679] Z. Darieva, A. Clancy, R. Bulmer, E. Williams, A. Pic-Taylor, B. A. Morgan, and A. D. Sharrocks, "A competitive transcription factor binding mechanism determines the timing of late cell cycle-dependent gene expression," *Molecular cell*, vol. 38, no. 1, pp. 29–40, 2010.

[680] S. Hahn and E. T. Young, "Transcriptional regulation in saccharomyces cerevisiae: transcription factor regulation and function, mechanisms of initiation, and roles of activators and coactivators," *Genetics*, vol. 189, no. 3, pp. 705–736, 2011.

[681] E. Aurell, A. F. D'Hérouël, C. Malmnäs, and M. Vergassola, "Transcription factor concentrations versus binding site affinities in the yeast S. cerevisiae," *Physical Biology*, vol. 4, pp. 134–143, jun 2007.

[682] M. D. Biggin, "Animal Transcription Networks as Highly Connected, Quantitative Continua," *Developmental Cell*, vol. 21, pp. 611–626, oct 2011.

[683] B. Alberts, "Cell biology: The endless frontier," *Molecular Biology of the Cell*, vol. 21, no. 22, pp. 3785–3785, 2010. PMID: 21079009.

[684] C. Kurtzman, J. W. Fell, and T. Boekhout, *The yeasts: a taxonomic study*. Elsevier, 2011.

[685] K. Chen, A. Csikasz-Nagy, B. Gyorffy, J. Val, B. Novak, and J. Tyson, "Kinetic analysis of a molecular model of the budding yeast cell cycle," *Molecular Biology of the Cell*, vol. 11, no. 1, pp. 369–391, 2000.

[686] K. C. Chen, L. Calzone, A. Csikasz-Nagy, F. R. Cross, B. Novak, and J. J. Tyson, "Integrative analysis of cell cycle control in budding yeast," *Molecular biology of the cell*, vol. 15, no. 8, pp. 3841–3862, 2004.

[687] F. Montefusco, O. E. Akman, O. S. Soyer, and D. G. Bates, "Modelling and analysis of feedback control mechanisms underlying osmoregulation in yeast," in *A Systems Theoretic Approach to Systems and Synthetic Biology II: Analysis and Design of Cellular Systems*, pp. 83–116, Springer, 2014.

[688] E. E. Selkov, "Self-Oscillations in Glycolysis: A Simple Kinetic Model," *European Journal of Biochemistry*, vol. 4, pp. 79–86, mar 1968.

[689] L. Edelstein and L. A. Segel, "Growth and metabolism in mycelial fungi," *Journal of Theoretical Biology*, vol. 104, pp. 187–210, sep 1983.

[690] G. Boswell, H. Jacobs, F. Davidson, G. Gadd, and K. Ritz, "A positive numerical scheme for a mixed-type partial differential equation model for fungal growth," *Applied Mathematics and Computation*, vol. 138, pp. 321–340, jun 2003.

[691] M. A. Henson, "Dynamic modeling of microbial cell populations," *Current Opinion in Biotechnology*, vol. 14, pp. 460–467, oct 2003.

[692] J. Amenta, M. Sargus, and S. Brocher, "Protein synthesis and degradation in growth regulation in rat embryo fibroblasts: Role of fast-turnover and slow-turnover protein," *Journal of cellular physiology*, vol. 105, no. 1, pp. 51–61, 1980.

[693] R. S. Pillai, S. N. Bhattacharyya, and W. Filipowicz, "Repression of protein synthesis by mrnas: how many mechanisms?," *Trends in cell biology*, vol. 17, no. 3, pp. 118–126, 2007.

[694] Y. Zhang, J. Nicholatos, J. R. Dreier, S. J. Ricoult, S. B. Widenmaier, G. S. Hotamisligil, D. J. Kwiatkowski, and B. D. Manning, "Coordinated regulation of protein synthesis and degradation by mtorc1," *Nature*, vol. 513, no. 7518, pp. 440–443, 2014.

[695] T. B. Kepler and T. C. Elston, "Stochasticity in transcriptional regulation: origins, consequences, and mathematical representations," *Biophysical journal*, vol. 81, no. 6, pp. 3116–3136, 2001.

[696] J. E. Hornos, D. Schultz, G. C. Innocentini, J. Wang, A. M. Walczak, J. N. Onuchic, and P. G. Wolynes, "Self-regulating gene: an exact solution," *Physical Review E*, vol. 72, no. 5, p. 051907, 2005.

[697] R. Alves, F. Antunes, and A. Salvador, "Tools for kinetic modeling of biochemical networks," *Nature Biotechnology*, vol. 24, pp. 667–672, jun 2006.

[698] T. Patarinska, D. Dochain, S. N. Agathos, and L. Ganovski, "Modelling of continuous microbial cultivation taking into account the memory effects," *Bioprocess Engineering*, vol. 22, pp. 517–527, jun 2000.

[699] R. Grosz and G. Stephanopoulos, "Physiological, biochemical, and mathematical studies of micro-aerobic continuous ethanol fermentation by Saccharomyces cerevisiae. I: Hysteresis, oscillations, and maximum specific ethanol productivities in chemostat culture," *Biotechnology and Bioengineering*, vol. 36, pp. 1006–1019, dec 1990.

[700] I. J. Ablowitz, A. Ramani, and H. Segur, "Nonlinear Evolution Equations and Ordinary Differential Equations of Painleve Type," *Lettere al Nuovo Cimento*, vol. 23, no. 1977, pp. 333–338, 1978.

[701] L. H. Erbe and H. Wang, "On the existence of positive solutions of ordinary differential equations," *Proceedings of the American Mathematical Society*, vol. 120, no. 3, pp. 743–748, 1994.

[702] W. Malfliet and W. Hereman, "The tanh method : Exact solutions of nonlinear evolution and wave equations," *Physica Scripta*, vol. 54, pp. 563–568, 1996.

[703] B. S. Desale and K. D. Patil, "Weak Singular Solution of Six Coupled Nonlinear ODEs," *Nonlinear Dynamics and Systems Theory*, vol. 16, no. 1, p. 8353, 2016.

[704] V. Zaitsev and A. Polyanin, *Handbook of exact solutions for ordinary differential equations*. CRC press, 2002.

[705] I. E. Lagaris, A. Likas, and D. I. Fotiadis, "Artificial Neural Networks for Solving Ordinary and Partial Differential Equations," *IEEE Transactions on Neural Networks*, vol. 9, no. 5, pp. 987–1000, 1998.

[706] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical recipes 3rd edition: The art of scientific computing*. Cambridge university press, 2007.

[707] M. S. Petković, B. Neta, L. D. Petković, and J. Džunić, "Multipoint methods for solving nonlinear equations: A survey," *Applied Mathematics and Computation*, vol. 226, pp. 635 – 660, 2014.

[708] H. Schmidt and M. Jirstrand, "Systems biology toolbox for matlab: a computational platform for research in systems biology," *Bioinformatics*, vol. 22, no. 4, pp. 514–515, 2006.

[709] S. M. Keating, B. J. Bornstein, A. Finney, and M. Hucka, "Sbmltoolbox: an sbml toolbox for matlab users," *Bioinformatics*, vol. 22, no. 10, pp. 1275–1277, 2006.

[710] P. Bogacki and L. Shampine, "A 3(2) Pair of Runge-Kutta Formulas," *Appl Math Lett*, vol. 2, no. 4, pp. 321–325, 1989.

[711] L. F. Shampine and M. W. Reichelt, "The MATLAB ODE Suite," *SIAM Journal on Scientific Computing*, vol. 18, no. 1, pp. 1–22, 1997.

[712] M. J. de Hoon, S. Imoto, J. Nolan, and S. Miyano, "Open source clustering software," *Bioinformatics*, vol. 20, no. 9, pp. 1453–1454, 2004.

[713] J. C. Butcher, "On the implementation of implicit runge-kutta methods," *BIT Numerical Mathematics*, vol. 16, no. 3, pp. 237–240, 1976.

[714] Z. Anastassi and T. Simos, "An optimized runge–kutta method for the solution of orbital problems," *Journal of Computational and Applied Mathematics*, vol. 175, no. 1, pp. 1–9, 2005.

[715] R. J. Britten and E. H. Davidson, "Gene regulation for higher cells: a theory," *Science*, vol. 165, no. 3891, pp. 349–357, 1969.

[716] R. Santer, "Chapter 8 - cellular mechanisms of aging," in *Brocklehurst's Textbook of Geriatric Medicine and Gerontology (Seventh Edition)* (H. M. Fillit, K. Rockwood, and K. Woodhouse, eds.), pp. 42 – 50, Philadelphia: W.B. Saunders, seventh edition ed., 2010.

[717] M. Ptashne, "Gene regulation by proteins acting nearby and at a distance," *Nature*, vol. 322, no. 6081, pp. 697–701, 1986.

[718] M. Inouye, "Antisense rna: its functions and applications in gene regulation—a review," *Gene*, vol. 72, no. 1-2, pp. 25–34, 1988.

[719] M. Mandal and R. R. Breaker, "Gene regulation by riboswitches," *Nature reviews Molecular cell biology*, vol. 5, no. 6, pp. 451–463, 2004.

[720] A. G. Hinnebusch, "Mechanisms of gene regulation in the general control of amino acid biosynthesis in saccharomyces cerevisiae.," *Microbiological reviews*, vol. 52, no. 2, p. 248, 1988.

[721] E. Stolovicki, T. Dror, N. Brenner, and E. Braun, "Synthetic gene recruitment reveals adaptive reprogramming of gene regulation in yeast," *Genetics*, vol. 173, no. 1, pp. 75–85, 2006.

[722] H. Yurimoto, M. Oku, and Y. Sakai, "Yeast methylotrophy: metabolism, gene regulation and peroxisome homeostasis," *International journal of microbiology*, vol. 2011, 2011.

[723] K. L. Mack and M. W. Nachman, "Gene regulation and speciation," *Trends in Genetics*, vol. 33, no. 1, pp. 68–80, 2017.

[724] E. H. Snoussi and R. Thomas, "Logical identification of all steady states: the concept of feedback loop characteristic states," *Bulletin of Mathematical Biology*, vol. 55, no. 5, pp. 973–991, 1993.

[725] R. Thomas, D. Thieffry, and M. Kaufman, "Dynamical behaviour of biological regulatory networks—i. biological role of feedback loops and practical use of the concept of the loop-characteristic state," *Bulletin of mathematical biology*, vol. 57, no. 2, pp. 247–276, 1995.

[726] T. Y.-C. Tsai, Y. S. Choi, W. Ma, J. R. Pomerening, C. Tang, and J. E. Ferrell, "Robust, tunable biological oscillations from interlinked positive and negative feedback loops.," *Science*, vol. 321, no. 5885, pp. 126–9, 2008.

[727] T. D. Pollard, W. C. Earnshaw, J. Lippincott-Schwartz, and G. Johnson, *Cell Biology*. Elsevier Health Sciences, 2016.

[728] O. Brandman, J. E. Ferrell, R. Li, and T. Meyer, "Interlinked Fast and Slow Positive Feedback Loops Drive Reliable Cell Decisions," *Science*, vol. 310, no. 5747, pp. 496–498, 2005.

[729] E. Karsenti, "Self-organization in cell biology: a brief history," *Nature reviews Molecular cell biology*, vol. 9, no. 3, pp. 255–262, 2008.

[730] L. Chen and K. Aihara, "Stability of genetic regulatory networks with time delay," *IEEE Transactions on circuits and systems I: Fundamental Theory and Applications*, vol. 49, no. 5, pp. 602–608, 2002.

[731] A. Y. Mitrophanov and E. A. Groisman, "Positive feedback in cellular control systems," *Bioessays*, vol. 30, no. 6, pp. 542–555, 2008.

[732] C. Tan, P. Marguet, and L. You, "Emergent bistability by a growth-modulating positive feedback circuit," *Nature chemical biology*, vol. 5, no. 11, p. 842, 2009.

[733] B. C. Goodwin, "Oscillatory behavior in enzymatic control processes," *Advances in enzyme regulation*, vol. 3, pp. 425–438, 1965.

[734] D. Greenbaum, R. Jansen, and M. Gerstein, "Analysis of mRNA expression and protein abundance data: An approach for the comparison of the enrichment of features in the cellular population of proteins and transcripts," *Bioinformatics*, vol. 18, no. 4, pp. 585–596, 2002.

[735] D. Angeli, J. E. Ferrell, and E. D. Sontag, "Detection of multistability, bifurcations, and hysteresis in a large class of biological positive-feedback systems," *Proceedings of the National Academy of Sciences*, vol. 101, no. 7, pp. 1822–1827, 2004.

[736] A. Singh, "Negative feedback through mrna provides the best control of gene-expression noise," *IEEE transactions on nanobioscience*, vol. 10, no. 3, pp. 194–200, 2011.

[737] S. Krishna, A. M. Andersson, S. Semsey, and K. Sneppen, "Structure and function of negative feedback loops at the interface of genetic and metabolic networks," *Nucleic acids research*, vol. 34, no. 8, pp. 2455–2462, 2006.

[738] D. P. Bazett-Jones, B. Leblanc, M. Herfort, and T. Moss, "Short-range dna looping by the xenopus hmg-box transcription factor, xubf," *Science*, vol. 264, no. 5162, pp. 1134–1137, 1994.

[739] J. Lu, T. A. McKinsey, R. L. Nicol, and E. N. Olson, "Signal-dependent activation of the mef2 transcription factor by dissociation from histone deacetylases," *Proceedings of the National Academy of Sciences*, vol. 97, no. 8, pp. 4070–4075, 2000.

[740] S. M. Parkhurst, D. Bopp, and D. Ish-Horowicz, "X: A ratio, the primary sex-determining signal in drosophila, is transduced by helix-loop-helix proteins," *Cell*, vol. 63, no. 6, pp. 1179–1191, 1990.

[741] P. J. O'Toole, T. Inoue, L. Emerson, I. E. Morrison, A. R. Mackie, R. J. Cherry, and J. D. Norton, "Id proteins negatively regulate basic helix-loop-helix transcription factor function by disrupting subnuclear compartmentalization," *Journal of Biological Chemistry*, vol. 278, no. 46, pp. 45770–45776, 2003.

[742] F. Jacob and J. Monod, "Genetic regulatory mechanisms in the synthesis of proteins," *Journal of Molecular Biology*, vol. 3, pp. 318–356, jun 1961.

[743] J. Monod, J. Wyman, and J.-P. Changeux, "On the nature of allosteric transitions: A plausible model," *Journal of Molecular Biology*, vol. 12, no. 1, pp. 88 – 118, 1965.

[744] M. Morales and D. McKay, "Biochemical oscillations in" controlled" systems," *Biophysical journal*, vol. 7, no. 5, pp. 621–625, 1967.

[745] S. Hastings, J. Tyson, and D. Webster, "Existence of periodic solutions for negative feedback cellular control systems," *Journal of Differential Equations*, vol. 25, no. 1, pp. 39–64, 1977.

[746] J. E. Ferrell, "Self-perpetuating states in signal transduction: Positive feedback, double-negative feedback and bistability," *Current Opinion in Cell Biology*, vol. 14, no. 2, pp. 140–148, 2002.

[747] S. Rogers, R. Khanin, and M. Girolami, "Bayesian model-based inference of transcription factor activity," *BMC Bioinformatics*, vol. 8, p. S2, may 2007.

[748] J. R. Williamson, "Cooperativity in macromolecular assembly," *Nature Chemical Biology*, vol. 4, pp. 458–465, aug 2008.

[749] A. Redei, *Protein Synthesis*, pp. 1580–1585. Dordrecht: Springer Netherlands, 2008.

[750] S. Y. Lee *et al.*, *Systems biology and biotechnology of Escherichia coli*. Springer, 2009.

[751] Y. Takahashi, M. Nishikawa, N. Takiguchi, T. Suehara, and Y. Takakura, "Saturation of transgene protein synthesis from mrna in cells producing a large number of transgene mrna," *Biotechnology and bioengineering*, vol. 108, no. 10, pp. 2380–2389, 2011.

[752] Y. Liu, A. Beyer, and R. Aebersold, "On the dependency of cellular protein levels on mrna abundance," *Cell*, vol. 165, no. 3, pp. 535–550, 2016.

[753] J. McDougall and E. C. Stoner, "The Computation of Fermi-Dirac Functions," *Philosophical Transactions of the Royal Society of London*, vol. 237, pp. 67–104, feb 1938.

[754] P. Rhodes, "Fermi-Dirac Functions of Integral Order," *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 204, pp. 396–405, dec 1950.

[755] A. Lipshtat, G. Jayaraman, J. C. He, and R. Iyengar, "Design of versatile biochemical switches that respond to amplitude, duration, and spatial cues," *Proceedings of the National Academy of Sciences*, vol. 107, no. 3, pp. 1247–1252, 2010.

[756] J. J. Tyson and H. G. Othmer, "The dynamics of feedback control circuits in biochemical pathways," *Prog. Theor. Biol*, vol. 5, no. 1, p. 62, 1978.

[757] J. Higgins, "The theory of oscillating reactions," *Industrial and Engineering Chemistry*, vol. 59, no. 5, pp. 18–62, 1967.

[758] B. Hess and A. Boiteux, "Oscillatory phenomena in biochemistry," *Annual review of biochemistry*, vol. 40, no. October 1970, pp. 237–258, 1971.

[759] J. Stelling and E. D. Gilles, "Mathematical modeling of complex regulatory networks," *IEEE transactions on nanobioscience*, vol. 3, no. 3, pp. 172–179, 2004.

[760] J. Paulsson, "Models of stochastic gene expression," *Physics of life reviews*, vol. 2, no. 2, pp. 157–175, 2005.

[761] C. Garmendia-Torres, A. Goldbeter, and M. Jacquet, "Nucleocytoplasmic oscillations of the yeast transcription factor msn2: evidence for periodic pka activation," *Current Biology*, vol. 17, no. 12, pp. 1044–1049, 2007.

[762] M. Mincheva and G. Craciun, "Multigraph conditions for multistability, oscillations and pattern formation in biochemical reaction networks," *Proceedings of the IEEE*, vol. 96, no. 8, pp. 1281–1291, 2008.

[763] H. Szu, "Fast simulated annealing," in *AIP conference proceedings*, vol. 151, pp. 420–425, American Institute of Physics, 1986.

[764] H. Szu and R. Hartley, "Fast simulated annealing," *Physics Letters A*, vol. 122, pp. 157–162, jun 1987.

[765] W. Bober, *Introduction to Numerical and Analytical Methods with MATLAB for Engineers and Scientists*, vol. 2013. CRC Press, nov 2014.

[766] J. J. Tyson, "On the existence of oscillatory solutions in negative feedback cellular control processes," *Journal of Mathematical Biology*, vol. 1, no. 4, pp. 311–315, 1975.

[767] G. Honig and M. Rabinovitz, "Inhibition of Protein Synthesis Unrelated to Effect on Template RNA Synthesis," *Science*, vol. 149, no. 1, pp. 1504–1506, 1965.

[768] T. Tuschl, P. D. Zamore, R. Lehmann, D. P. Bartel, and P. Sharp, "Targeted mRNA degradation by double-stranded RNA in vitro," *Genes and Development*, vol. 13, pp. 3191–3197, 1999.

[769] V. Sosunov, E. Sosunova, A. Mustaev, I. Bass, V. Nikiforov, and A. Goldfarb, "Unified two-metal mechanism of RNA synthesis and degradation by RNA polymerase," *EMBO Journal*, vol. 22, no. 9, pp. 2234–2244, 2003.

[770] J. DiStefano III, "A model of the normal thyroid hormone glandular secretion mechanism," *Journal of theoretical biology*, vol. 22, no. 3, pp. 412–417, 1969.

[771] L. Olsen, J. A. Sherratt, P. K. Maini, and F. Arnold, "A mathematical model for the capillary endothelial cell-extracellular matrix interactions in wound-healing angiogenesis," *Mathematical Medicine and Biology: A Journal of the IMA*, vol. 14, no. 4, pp. 261–281, 1997.

[772] H. A. Levine, B. D. Sleeman, and M. Nilsen-Hamilton, "A mathematical model for the roles of pericytes and macrophages in the initiation of angiogenesis. i. the role of protease inhibitors in preventing angiogenesis," *Mathematical biosciences*, vol. 168, no. 1, pp. 77–115, 2000.

[773] A. Hershko and A. Ciechanover, "Mechanisms of intracellular protein breakdown," *Annual review of biochemistry*, vol. 51, no. 1, pp. 335–364, 1982.

[774] W. C. Merrick, "Mechanism and regulation of eukaryotic protein synthesis.," *Microbiology and Molecular Biology Reviews*, vol. 56, no. 2, pp. 291–315, 1992.

[775] H. Koch and J. Friesen, "Individual messenger rna half lives in saccharomyces cerevisiae," *Molecular and General Genetics MGG*, vol. 170, no. 2, pp. 129–135, 1979.

[776] D. S. Pederson and R. Morse, "Effect of transcription of yeast chromatin on dna topology in vivo.," *The EMBO journal*, vol. 9, no. 6, pp. 1873–1881, 1990.

[777] S. A. Uhler, D. Cai, Y. Man, C. Figge, and N. G. Walter, "Rna degradation in cell extracts: real-time monitoring by fluorescence resonance energy transfer," *Journal of the American Chemical Society*, vol. 125, no. 47, pp. 14230–14231, 2003.

[778] C. Miller, B. Schwalb, K. Maier, D. Schulz, S. Dümcke, B. Zacher, A. Mayer, J. Sydow, L. Marcinowski, L. Dölken, D. E. Martin, A. Tresch, and P. Cramer, "Dynamic transcriptome analysis measures rates of mRNA synthesis and decay in yeast," *Molecular Systems Biology*, vol. 7, no. 458, 2011.

[779] I. G. Romero, A. A. Pai, J. Tung, and Y. Gilad, "Rna-seq: impact of rna degradation on transcript quantification," *BMC biology*, vol. 12, no. 1, p. 42, 2014.

[780] A. Baudrimont, V. Jaquet, S. Wallerich, S. Voegeli, and A. Becskei, "Contribution of rna degradation to intrinsic and extrinsic noise in gene expression," *Cell reports*, vol. 26, no. 13, pp. 3752–3761, 2019.

[781] G. Johnston and R. Singer, "Rna synthesis and control of cell division in the yeast s. cerevisiae," *Cell*, vol. 14, no. 4, pp. 951–958, 1978.

[782] A. Schwabe and F. J. Bruggeman, "Single yeast cells vary in transcription activity not in delay time after a metabolic shift," *Nature communications*, vol. 5, no. 1, pp. 1–10, 2014.

[783] D. G. Dalgleish, "Proteolysis and aggregation of casein micelles treated with immobilized or soluble chymosin," *Journal of Dairy Research*, vol. 46, no. 4, pp. 653–661, 1979.

[784] U. Krishnamoorthy, C. Sniffen, M. D. Stern, and P. Van Soest, "Evaluation of a mathematical model of rumen digestion and an in vitro simulation of rumen proteolysis to estimate the rumen-undegraded nitrogen content of feedstuffs," *British Journal of Nutrition*, vol. 50, no. 3, pp. 555–568, 1983.

[785] B. Novak, A. Csikasz-Nagy, B. Gyorffy, K. Chen, and J. J. Tyson, "Mathematical model of the fission yeast cell cycle with checkpoint controls at the g1/s, g2/m and metaphase/anaphase transitions," *Biophysical chemistry*, vol. 72, no. 1-2, pp. 185–200, 1998.

[786] C. K. Biebricher, M. Eigen, and R. Luce, "Kinetic analysis of template-instructed and de novo rna synthesis by qβ replicase," *Journal of molecular biology*, vol. 148, no. 4, pp. 391–410, 1981.

[787] S. J. Coppella and P. Dhurjati, "A mathematical description of recombinant yeast," *Biotechnology and bioengineering*, vol. 35, no. 4, pp. 356–374, 1990.

[788] D. A. Drew, "A mathematical model for prokaryotic protein synthesis," *Bulletin of mathematical biology*, vol. 63, no. 2, pp. 329–351, 2001.

[789] H. Bremer, P. Dennis, and M. Ehrenberg, "Free rna polymerase and modeling global transcription in escherichia coli," *Biochimie*, vol. 85, no. 6, pp. 597–609, 2003.

[790] A. Tonnelier, "Cyclic negative feedback systems: what is the chance of oscillation?," *Bulletin of mathematical biology*, vol. 76, no. 5, pp. 1155–1193, 2014.

[791] J. Milton and T. Ohira, "Feedback and control systems," in *Mathematics as a Laboratory Tool*, pp. 219–257, Springer, 2014.

[792] M. B. Elowitz, A. J. Levine, E. D. Siggia, and P. S. Swain, "Stochastic gene expression in a single cell," *Science*, vol. 297, pp. 1183–1186, aug 2002.

[793] J. M. Raser and E. K. O'Shea, "Control of stochasticity in eukaryotic gene expression," *science*, vol. 304, no. 5678, pp. 1811–1814, 2004.

[794] I. Golding, J. Paulsson, S. M. Zawilski, and E. C. Cox, "Real-time kinetics of gene activity in individual bacteria," *Cell*, vol. 123, no. 6, pp. 1025–1036, 2005.

[795] A. Raj, C. S. Peskin, D. Tranchina, D. Y. Vargas, and S. Tyagi, "Stochastic mrna synthesis in mammalian cells," *PLoS biology*, vol. 4, no. 10, 2006.

[796] S. Wang and S. Raghavachari, "Quantifying negative feedback regulation by micro-rnas," *Physical biology*, vol. 8, no. 5, p. 055002, 2011.

[797] N. Friedman, M. Linial, I. Nachman, and D. Pe'er, "Using bayesian networks to analyze expression data," *Journal of computational biology*, vol. 7, no. 3-4, pp. 601–620, 2000.

[798] K. Murphy, S. Mian, *et al.*, "Modelling gene expression data using dynamic bayesian networks," tech. rep., Technical report, Computer Science Division, University of California ..., 1999.

[799] D. Pe'er, A. Regev, G. Elidan, and N. Friedman, "Inferring subnetworks from perturbed expression profiles," *Bioinformatics*, vol. 17, no. suppl_1, pp. S215–S224, 2001.

[800] H. Hirata, S. Yoshiura, T. Ohtsuka, Y. Bessho, T. Harada, K. Yoshikawa, and R. Kageyama, "Oscillatory expression of the bhlh factor hes1 regulated by a negative feedback loop," *Science*, vol. 298, no. 5594, pp. 840–843, 2002.

[801] N. A. Monk, "Oscillatory expression of hes1, p53, and nf-κb driven by transcriptional time delays," *Current Biology*, vol. 13, no. 16, pp. 1409–1413, 2003.

[802] J. De Caluwé, Q. Xiao, C. Hermans, N. Verbruggen, J.-C. Leloup, and D. Gonze, "A compact model for the complex plant circadian clock," *Frontiers in plant science*, vol. 7, p. 74, 2016.

[803] K. Kaufmann, M. Nagasaki, and R. Jáuregui, "Modelling the molecular interactions in the flower developmental network of arabidopsis thaliana," *In silico biology*, vol. 10, no. 1, 2, pp. 125–143, 2010.

[804] G. Yagil, "Quantitative aspects of protein induction," in *Current topics in cellular regulation*, vol. 9, pp. 183–236, Elsevier, 1975.

[805] G. Yagil and E. Yagil, "On the relation between effector concentration and the rate of induced enzyme synthesis," *Biophysical journal*, vol. 11, no. 1, p. 11, 1971.

[806] M. W. Pfaffl, "A new mathematical model for relative quantification in real-time RT-PCR," *Nucleic Acids Research*, vol. 29, no. 9, pp. 45e–45, 2001.

[807] K. Glover-Cutter, S. Kim, J. Espinosa, and D. L. Bentley, "RNA polymerase II pauses and associates with pre-mRNA processing factors at both ends of genes," *Nature Structural and Molecular Biology*, vol. 15, no. 1, pp. 71–78, 2008.

[808] B. Schwanhüusser, D. Busse, N. Li, G. Dittmar, J. Schuchhardt, J. Wolf, W. Chen, and M. Selbach, "Global quantification of mammalian gene expression control," *Nature*, vol. 473, no. 7347, pp. 337–342, 2011.

[809] N. Banerjee and M. Q. Zhang, "Identifying cooperativity among transcription factors controlling the cell cycle in yeast," *Nucleic Acids Research*, vol. 31, pp. 7024–7031, dec 2003.

[810] D. Das, N. Banerjee, and M. Q. Zhang, "Interacting models of cooperative gene regulation," *Proceedings of the National Academy of Sciences*, vol. 101, no. 46, pp. 16234–16239, 2004.

[811] F. Spitz and E. E. M. Furlong, "Transcription factors: from enhancer binding to developmental control," *Nature Reviews Genetics*, vol. 13, pp. 613–626, sep 2012.

[812] L. A. Mirny, "Nucleosome-mediated cooperativity between transcription factors," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 107, pp. 22534–9, dec 2010.

[813] M. P. Deutscher, "Degradation of rna in bacteria: comparison of mrna and stable rna," *Nucleic acids research*, vol. 34, no. 2, pp. 659–666, 2006.

[814] A. J. Carpousis, B. F. Luisi, and K. J. McDowall, "Endonucleolytic initiation of mrna decay in escherichia coli," *Progress in molecular biology and translational science*, vol. 85, pp. 91–135, 2009.

[815] J. E. Lee, J. Y. Lee, J. Wilusz, B. Tian, and C. J. Wilusz, "Systematic analysis of cis-elements in unstable mrnas demonstrates that cugbp1 is a key regulator of mrna decay in muscle cells," *PloS one*, vol. 5, no. 6, 2010.

[816] J. Frank, J. Zhu, P. Penczek, Y. Li, S. Srivastava, A. Verschoor, M. Radermacher, R. Grassucci, R. K. Lata, and R. K. Agrawal, "A model of protein synthesis based on cryo-electron microscopy of the E. coli ribosome," *Nature*, vol. 376, no. 6539, pp. 441–444, 1995.

[817] J. A. Doudna and V. L. Rath, "Structure and function of the eukaryotic ribosome: the next frontier," *Cell*, vol. 109, no. 2, pp. 153–156, 2002.

[818] L. B. Shaw, R. K. Zia, and K. H. Lee, "Totally asymmetric exclusion process with extended objects: A model for protein synthesis," *Physical Review E - Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics*, vol. 68, no. 2, p. 17, 2003.

[819] C. Vogel, R. de Sousa Abreu, D. Ko, S.-Y. Le, B. A. Shapiro, S. C. Burns, D. Sandhu, D. R. Boutz, E. M. Marcotte, and L. O. Penalva, "Sequence signatures and mrna concentration can explain two-thirds of protein abundance variation in a human cell line," *Molecular systems biology*, vol. 6, no. 1, 2010.

[820] C. Vogel and E. M. Marcotte, "Insights into the regulation of protein abundance from proteomic and transcriptomic analyses," *Nature reviews genetics*, vol. 13, no. 4, pp. 227–232, 2012.

[821] A. Belle, A. Tanay, L. Bitincka, R. Shamir, and E. K. O'Shea, "Quantification of protein half-lives in the budding yeast proteome," *Proceedings of the National Academy of Sciences*, vol. 103, no. 35, pp. 13004–13009, 2006.

[822] W. J. Becktel and J. A. Schellman, "Protein stability curves," *Biopolymers: Original Research on Biomolecules*, vol. 26, no. 11, pp. 1859–1877, 1987.

[823] R. Jaenicke, "Protein stability and molecular adaptation to extreme conditions," in *EJB Reviews 1991*, pp. 291–304, Springer, 1991.

[824] B. W. Matthews, "Structural and genetic analysis of protein stability," *Annual review of biochemistry*, vol. 62, no. 1, pp. 139–160, 1993.

[825] J. D. Bloom, S. T. Labthavikul, C. R. Otey, and F. H. Arnold, "Protein stability promotes evolvability," *Proceedings of the National Academy of Sciences*, vol. 103, no. 15, pp. 5869–5874, 2006.

[826] M. L. Bender and F. J. Kezdy, "Mechanism of action of proteolytic enzymes," *Annual review of biochemistry*, vol. 34, no. 1, pp. 49–76, 1965.

[827] H. Neurath and K. A. Walsh, "Role of proteolytic enzymes in biological regulation (a review)," *Proceedings of the National Academy of Sciences*, vol. 73, no. 11, pp. 3825–3832, 1976.

[828] W. Q. Qiu, D. M. Walsh, Z. Ye, K. Vekrellis, J. Zhang, M. B. Podlisny, M. R. Rosner, A. Safavi, L. B. Hersh, and D. J. Selkoe, "Insulin-degrading enzyme regulates extracellular levels of amyloid β-protein by degradation," *Journal of Biological Chemistry*, vol. 273, no. 49, pp. 32730–32738, 1998.

[829] M. Benhar, M. T. Forrester, and J. S. Stamler, "Protein denitrosylation: enzymatic mechanisms and cellular functions," *Nature reviews Molecular cell biology*, vol. 10, no. 10, pp. 721–732, 2009.

[830] K. Linderstrøm-Lang, "Proteolytic enzymes," *Annual Review of Biochemistry*, vol. 8, no. 1, pp. 37–58, 1939.

[831] M. Bley, M. Johnson, and N. Olson, "Factors affecting nonenzymatic browning of process cheese," *Journal of Dairy Science*, vol. 68, no. 3, pp. 555–561, 1985.

[832] A. L. Goldberg, "Protein degration and protection against misfolded proteins," *Nature*, vol. 426, no. December, pp. 895–899, 2003.

[833] G. A. Broderick and M. K. Clayton, "Rumen protein degradation rates estimated by non-linear regression analysis of Michaelis-Menten in vitro data," *The British journal of nutrition*, vol. 67, no. 1992, pp. 27–42, 1992.

[834] L. Menten and M. Michaelis, "Die kinetik der invertinwirkung," *Biochem Z*, vol. 49, no. 333-369, p. 5, 1913.

[835] B. Li, V. Ruotti, R. M. Stewart, J. A. Thomson, and C. N. Dewey, "RNA-Seq gene expression estimation with read mapping uncertainty," *Bioinformatics*, vol. 26, pp. 493–500, feb 2009.

[836] M. Hampsey, "Molecular genetics of the rna polymerase ii general transcriptional machinery," *Microbiol. Mol. Biol. Rev.*, vol. 62, no. 2, pp. 465–503, 1998.

[837] H. Lodish, A. Berk, S. L. Zipursky, P. Matsudaira, D. Baltimore, and J. Darnell, "Rna polymerase ii transcription-initiation complex," in *Molecular Cell Biology. 4th edition*, WH Freeman, 2000.

[838] B. Lewin, "Units of transcription and translation: the relationship between heterogeneous nuclear RNA and messenger RNA," *Cell*, vol. 4, no. 1, pp. 11–20, 1975.

[839] H. B. Fraser, A. E. Hirsh, G. Giaever, J. Kumm, and M. B. Eisen, "Noise minimization in eukaryotic gene expression," *PLoS Biology*, vol. 2, p. e137, apr 2004.

[840] V. Pelechano, S. Chávez, and J. E. Pérez-Ortín, "A complete set of nascent transcription rates for yeast genes," *PLoS ONE*, vol. 5, p. e15442, nov 2010.

[841] K. W. Boehlke and J. D. Friesen, "Cellular content of ribonucleic acid and protein in Saccharomyces cerevisiae as a function of exponential growth rate: calculation of the apparent peptide chain elongation rate," *Journal of Bacteriology*, vol. 121, no. 2, pp. 429–433, 1975.

[842] C. Waldron and F. Lacroute, "Effect of Growth Rate on the Amounts of Ribosomal and Transfer Ribonucleic Acids in Yeast," *Journal of Bacteriology*, vol. 122, no. 3, pp. 855–865, 1975.

[843] T. V. Karpinets, D. J. Greenwood, C. E. Sams, and J. T. Ammons, "RNA:Protein ratio of the unicellular organism as a characteristic of phosphorous and nitrogen stoichiometry and of the cellular requirement of ribosomes for protein synthesis," *BMC Biology*, vol. 4, p. 30, sep 2006.

[844] P. Jorgensen, N. Edgington, B. Schneider, I. Rupes, M. Tyers, and B. Futcher, "The size of the nucleus increases as yeast cells grow," *Mol Biol of the cell*, vol. 18, pp. 3523–3532, sep 2007.

[845] Bureau International des Poids et Mesures, *The International System of Units (SI), 9th edition*. BIPM Press, 2019.

[846] L. M. Hereford and M. Rosbash, "Number and distribution of polyadenylated RNA sequences in yeast," *Cell*, vol. 10, pp. 453–462, mar 1977.

[847] A. Coghlan and K. H. Wolfe, "Relationship of codon bias to mRNA and concentration protein length in Saccharomyces cerevisiae," *Yeast*, vol. 16, pp. 1131–1145, sep 2000.

[848] D. Greenbaum, C. Colangelo, K. Williams, and M. Gerstein, "Comparing protein abundance and mRNA expression levels on a genomic scale," *Genome Biology*, vol. 4, p. 117, aug 2003.

[849] M. L. Fournier, A. Paulson, N. Pavelka, A. L. Mosley, K. Gaudenz, W. D. Bradford, E. Glynn, H. Li, M. E. Sardiu, B. Fleharty, *et al.*, "Delayed correlation of mrna and protein expression in rapamycin-treated cells and a role for ggc1 in cellular sensitivity to rapamycin," *Molecular & Cellular Proteomics*, vol. 9, no. 2, pp. 271–284, 2010.

[850] S. Kojima, E. L. Sher-Chen, and C. B. Green, "Circadian control of mrna polyadenylation dynamics regulates rhythmic protein expression," *Genes & development*, vol. 26, no. 24, pp. 2724–2736, 2012.

[851] S. Hocine, P. Raymond, D. Zenklusen, J. A. Chao, and R. H. Singer, "Single-molecule analysis of gene expression using two-color rna labeling in live yeast," *Nature methods*, vol. 10, no. 2, pp. 119–121, 2013.

[852] J. Sun, J. M. Garibaldi, C. Hodgman, Jianyong Sun, J. M. Garibaldi, and C. Hodgman, "Parameter estimation using meta-heuristics in systems biology: a comprehensive review," *IEEE/ACM transactions on computational biology and bioinformatics*, vol. 9, no. 1, pp. 185–202, 2012.

[853] E. O. Voit, "Biochemical systems theory: a review," *ISRN biomathematics*, vol. 2013, 2013.

[854] S. Mahdavi, M. E. Shiri, and S. Rahnamayan, "Metaheuristics in large-scale global continues optimization: A survey," *Information Sciences*, vol. 295, pp. 407–428, 2015.

[855] W. L. Goffe, G. D. Ferrier, and J. Rogers, "Global optimization of statistical functions with simulated annealing," *Journal of econometrics*, vol. 60, no. 1-2, pp. 65–99, 1994.

[856] J. April, F. Glover, J. K. Kelly, and M. Laguna, "Practical Introduction To Simulation Optimization," *Proceedings of the 2003 Winter Simulation Conference*, pp. 71–78, 2003.

[857] J. A. Snyman, *Practical mathematical optimization*. Springer, 2005.

[858] R. R. Barton and M. Meckesheimer, "Metamodel-based simulation optimization," *Handbooks in operations research and management science*, vol. 13, pp. 535–574, 2006.

[859] A. Dekkers and E. Aarts, "Global optimization and simulated annealing," *Mathematical programming*, vol. 50, no. 1-3, pp. 367–393, 1991.

[860] S. P. Brooks and B. J. Morgan, "Optimization using simulated annealing," *Journal of the Royal Statistical Society: Series D (The Statistician)*, vol. 44, no. 2, pp. 241–257, 1995.

[861] J. Wu and S. Wei, "Time series analysis," *Hunan Science and Technology Press, ChangSha. Available online at: http://www2. geog. ucl. ac. uk/˜ mdisney/teaching/GEOGG121/time_series/GEOGG121_5_TimeSeries_ Wu. pdf. Retrieved: January*, vol. 20, p. 2018, 1989.

[862] T. Obayashi and K. Kinoshita, "Rank of correlation coefficient as a comparable measure for biological significance of gene coexpression," *DNA Research*, vol. 16, pp. 249–260, oct 2009.

[863] M. A. Savageau, "Biochemical systems analysis: Iii. dynamic solutions using a power-law approximation," *Journal of theoretical biology*, vol. 26, no. 2, pp. 215–226, 1970.

[864] W. Yin and E. O. Voit, "Construction and customization of stable oscillation models in biology," *Journal of Biological Systems*, vol. 16, no. 04, pp. 463–478, 2008.

[865] S. Proshkin, A. Rachid Rahmouni, A. Mironov, and E. Nudler, "Cooperation between translating ribosomes and RNA polymerase in transcription elongation," *Science*, vol. 328, pp. 504–508, apr 2010.

[866] Y.-H. Chang, Y.-C. Wang, and B.-S. Chen, "Identification of transcription factor cooperativity via stochastic system model," *Bioinformatics*, vol. 22, no. 18, pp. 2276–2282, 2006.

[867] W. W. Wong, T. Y. Tsai, and J. C. Liao, "Single-cell zeroth-order protein degradation enhances the robustness of synthetic oscillator," *Molecular Systems Biology*, vol. 3, no. 130, pp. 1–8, 2007.

[868] S. Ghaemmaghami, W.-K. Huh, K. Bower, R. W. Howson, A. Belle, N. Dephoure, E. K. O'Shea, and J. S. Weissman, "Global analysis of protein expression in yeast," *Nature*, vol. 425, pp. 737–741, oct 2003.

[869] C. Grilly, J. Stricker, W. L. Pang, M. R. Bennett, and J. Hasty, "A synthetic gene network for tuning protein degradation in saccharomyces cerevisiae," *Molecular systems biology*, vol. 3, no. 1, 2007.

[870] M. Sabouri-Ghomi, A. Ciliberto, S. Kar, B. Novak, and J. J. Tyson, "Antagonism and bistability in protein interaction networks," *Journal of theoretical biology*, vol. 250, no. 1, pp. 209–218, 2008.

[871] R. S. Costa, D. Machado, I. Rocha, and E. C. Ferreira, "Hybrid dynamic modeling of escherichia coli central metabolic network combining michaelis–menten and approximate kinetic equations," *Biosystems*, vol. 100, no. 2, pp. 150–157, 2010.

[872] L. Xu and Z. Qu, "Roles of protein ubiquitination and degradation kinetics in biological oscillations," *PloS one*, vol. 7, no. 4, 2012.

[873] A. Goldbeter, "Oscillatory enzyme reactions and michaelis–menten kinetics," *FEBS letters*, vol. 587, no. 17, pp. 2778–2784, 2013.

[874] R. Grima, N. G. Walter, and S. Schnell, "Single-molecule enzymology à la michaelis–menten," *The FEBS journal*, vol. 281, no. 2, pp. 518–530, 2014.

[875] S. H. Strogatz, *Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry, and Engineering Perseus*. Perseus Books. Reading, Massachusetts, 1994.

[876] M. T. Borisuk and J. J. Tyson, "Bifurcation analysis of a model of mitotic control in frog eggs," *Journal of theoretical biology*, vol. 195, no. 1, pp. 69–85, 1998.

[877] H. De Jong, "Modeling and simulation of genetic regulatory systems: a literature review," *Journal of computational biology*, vol. 9, no. 1, pp. 67–103, 2002.

[878] B. Novak and J. J. Tyson, "Quantitative analysis of a molecular model of mitotic control in fission yeast," *Journal of Theoretical Biology*, vol. 173, no. 3, pp. 283–305, 1995.

[879] M. Tanelli, G. Osorio, M. di Bernardo, S. M. Savaresi, and A. Astolfi, "Existence, stability and robustness analysis of limit cycles in hybrid anti-lock braking systems," *International Journal of Control*, vol. 82, no. 4, pp. 659–678, 2009.

[880] G. Tanaka, K. Morino, H. Daido, and K. Aihara, "Dynamical robustness of coupled heterogeneous oscillators," *Physical Review E*, vol. 89, no. 5, p. 052906, 2014.

[881] D. Barik, W. T. Baumann, M. R. Paul, B. Novak, and J. J. Tyson, "A model of yeast cell-cycle regulation based on multisite phosphorylation," *Molecular Systems Biology*, vol. 6, no. 405, pp. 1–18, 2010.

[882] I. Enting, "The second-neighbour ising chain as a model for spin-phonon interactions: susceptibility," *Journal of Physics A: Mathematical, Nuclear and General*, vol. 6, no. 2, p. 170, 1973.

[883] P. Slotte and P. Hemmer, "The triangular ising antiferromagnet with first-and second-neighbour interactions," *Journal of Physics C: Solid State Physics*, vol. 17, no. 26, p. 4645, 1984.

[884] P. Sen, "Order disorder transitions in ising models in transverse fields with second neighbour interactions," *Zeitschrift für Physik B Condensed Matter*, vol. 98, no. 2, pp. 251–254, 1995.

[885] G. Lelandais, Y. Saint-Georges, C. Geneix, L. Al-Shikhley, G. Dujardin, and C. Jacq, "Spatio-temporal dynamics of yeast mitochondrial biogenesis: Transcriptional and post-transcriptional mRNA oscillatory modules," *PLoS Computational Biology*, vol. 5, no. 6, 2009.

[886] M. Rullan, D. Benzinger, G. W. Schmidt, and A. Milias-argeitis, "An optogenetic platform for real-time , single-cell interrogation of stochastic transcription regulation," *Molecular Cell*, vol. 70, no. 4, pp. 745–756.e6, 2018.

[887] A. K. Ghosh, B. Chance, and E. K. Pye, "Metabolic Coupling and Synchronization of NADH Oscillations in Yeast Cell Populations," *Archives of biochemistry and biophysics*, vol. 145, pp. 319–331, 1971.

[888] S. Laxman, B. P. Tu, and A. Manuscript, "Systems approaches for the study of metabolic cycles in yeast," *Current Opinion in Genetics & Development*, vol. 20, pp. 599–604, dec 2011.

[889] E. Baake, M. Baake, and H. Wagner, "Ising quantum chain is equivalent to a model of biological evolution," *Physical Review Letters*, vol. 78, no. 3, p. 559, 1997.

[890] H. Morita and K. Kaneko, "Collective oscillation in a hamiltonian system," *Physical review letters*, vol. 96, no. 5, p. 050602, 2006.

[891] J. C. Ewald, A. Kuehne, N. Zamboni, and J. M. Skotheim, "The Yeast Cyclin-Dependent Kinase Routes Carbon Fluxes to Fuel Cell Cycle Progression," *Molecular Cell*, vol. 62, pp. 532–545, may 2016.

[892] M. Bömer, J. A. O'Brien, I. Pérez-Salamó, J. Krasauskas, P. Finch, A. Briones, A. Daudi, P. Souda, T.-L. Tsui, J. P. Whitelegge, G. Paul Bolwell, and A. Devoto, "COI1-dependent jasmonate signalling affects growth, metabolite production and cell wall protein composition in arabidopsis.," *Annals of Botany*, jun 2018.

[893] J. L. Bolen, O. Durán, and R. R. Klevecz, "Amplification and damping of deterministic noise in coupled cellular arrays," *Physica D: Nonlinear Phenomena*, vol. 67, no. 1-3, pp. 245–256, 1993.

[894] R. R. Klevecz, S. A. Kauffman, and R. M. Shymko, "Cellular Clocks and Oscillators," *International Review of Cytology*, vol. 86, pp. 97–128, jan 1984.

[895] H. Aoki, N. Tsuji, M. Eckstein, M. Kollar, T. Oka, and P. Werner, "Nonequilibrium dynamical mean-field theory and its applications," *Reviews of Modern Physics*, vol. 86, pp. 779–837, jun 2014.

[896] R. R. Klevecz, J. Bolen, and O. Duran, "Self-organization in biological tissues: Analysis of asynchronous and synchronous periodicity, turbulence and synchronous chaos emergent in coupled chaotic arrays," *International Journal of Bifurcation and Chaos*, vol. 02, pp. 941–953, dec 1992.

[897] E. Schwob, T. Bohm, M. D. Mendenhall, and K. Nasmyth, "The B-type cyclin kinase inhibitor p40SIC1 controls the G1 to S transition in S. cerevisiae," *Cell*, vol. 79, no. 2, pp. 233–244, 1994.

[898] J. C. Ewald, "How yeast coordinates metabolism, growth and division," *Current Opinion in Microbiology*, vol. 45, pp. 1–7, 2018.

[899] L. Shi and B. P. Tu, "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 of the United States of America*, vol. 110, no. 18, pp. 7318–7323, 2013.

[900] M. Tyers, G. Tokiwa, and B. Futcher, "Comparison of the Saccharomyces cerevisiae G1 cyclins: Cln3 may be an upstream activator of Cln1, Cln2 and other cyclins.," *The EMBO journal*, vol. 12, pp. 1955–68, may 1993.

[901] H. Wijnen, A. Landman, and B. Futcher, "The G1 cyclin Cln3 promotes cell cycle entry via the transcription factor Swi6," *Molecular and cellular biology*, vol. 22, pp. 4402–18, jun 2002.

[902] T. Miura and T. YANAGITA, "Cellular senescence in yeast caused by carbon-source starvation," *The Journal of Biochemistry*, vol. 72, no. 1, pp. 141–148, 1972.

[903] A. A. Petti, C. A. Crutchfield, J. D. Rabinowitz, and D. Botstein, "Survival of starving yeast is correlated with oxidative stress response and nonrespiratory mitochondrial function," *Proceedings of the National Academy of Sciences*, vol. 108, no. 45, pp. E1089–E1098, 2011.

[904] D. Laporte, A. Lebaudy, A. Sahin, B. Pinson, J. Ceschin, B. Daignan-Fornier, and I. Sagot, "Metabolic status rather than cell cycle signals control quiescence entry and exit," *Journal of Cell Biology*, vol. 192, pp. 949–957, mar 2011.

[905] K. G. Sabra, P. Roux, and W. A. Kuperman, "Emergence rate of the time-domain Green's function from the ambient noise cross-correlation function,"

The Journal of the Acoustical Society of America, vol. 118, pp. 3524–3531, dec 2005.

[906] M. Kaern, T. C. Elston, W. J. Blake, and J. J. Collins, "Stochasticity in gene expression: from theories to phenotypes," *Nature Reviews Genetics*, vol. 6, no. 6, pp. 451–464, 2005.

[907] A. Raj and A. van Oudenaarden, "Nature, nurture, or chance: stochastic gene expression and its consequences," *Cell*, vol. 135, no. 2, pp. 216–226, 2008.

[908] Y. Buganim, D. A. Faddah, A. W. Cheng, E. Itskovich, S. Markoulaki, K. Ganz, S. L. Klemm, A. van Oudenaarden, and R. Jaenisch, "Single-cell expression analyses during cellular reprogramming reveal an early stochastic and a late hierarchic phase," *Cell*, vol. 150, no. 6, pp. 1209–1222, 2012.

[909] J. M. Thevelein, "Signal transduction in yeast," *Yeast*, vol. 10, pp. 1753–1790, dec 1994.

[910] B. Smets, R. Ghillebert, P. De Snijder, M. Binda, E. Swinnen, C. De Virgilio, and J. Winderickx, "Life in the midst of scarcity: adaptations to nutrient availability in saccharomyces cerevisiae," *Current genetics*, vol. 56, no. 1, pp. 1–32, 2010.

[911] A. Levchenko and I. Nemenman, "Cellular noise and information transmission," *Current opinion in biotechnology*, vol. 28, pp. 156–164, 2014.

[912] M. D. Harton and E. Batchelor, "Determining the limitations and benefits of noise in gene regulation and signal transduction through single cell, microscopy-based analysis," *Journal of molecular biology*, vol. 429, no. 8, pp. 1143–1154, 2017.

[913] D. J. Thomson, "Criteria for the selection of stochastic models of particle trajectories in turbulent flows," *Journal of Fluid Mechanics*, vol. 180, pp. 529–556, jul 1987.

[914] F. Delvigne, A. Lejeune, J. Destain, and P. Thonart, "Stochastic models to study the impact of mixing on a fed-batch culture of Saccharomyces cerevisiae," in *Biotechnology Progress*, vol. 22, pp. 259–269, American Chemical Society (ACS), feb 2006.

[915] F. Delvigne, J. Destain, and P. Thonart, "A methodology for the design of scale-down bioreactors by the use of mixing and circulation stochastic models," *Biochemical Engineering Journal*, vol. 28, pp. 256–268, mar 2006.

[916] J. T. Mettetal, D. Muzzey, J. M. Pedraza, E. M. Ozbudak, and A. Van Oudenaarden, "Predicting stochastic gene expression dynamics in single cells," *Proceedings of the National Academy of Sciences*, vol. 103, no. 19, pp. 7304–7309, 2006.

[917] V. Shahrezaei and P. S. Swain, "Analytical distributions for stochastic gene expression," *Proceedings of the National Academy of Sciences*, vol. 105, no. 45, pp. 17256–17261, 2008.

[918] S. Vidal, M. Petitot, F. Boulier, F. Lemaire, and C. Kuttler, "Models of stochastic gene expression and weyl algebra," in *Algebraic and Numeric Biology*, pp. 76–97, Springer, 2012.

[919] T. Tian and K. Burrage, "Stochastic neural network models for gene regulatory networks," in *The 2003 Congress on Evolutionary Computation, 2003. CEC'03.*, vol. 1, pp. 162–169, IEEE, 2003.

[920] J. Nacher and T. Akutsu, "Sensitivity of the power-law exponent in gene expression distribution to mrna decay rate," *Physics Letters A*, vol. 360, no. 1, pp. 174–178, 2006.

[921] B.-S. Chen and W.-S. Wu, "Robust filtering circuit design for stochastic gene networks under intrinsic and extrinsic molecular noises," *Mathematical Biosciences*, vol. 211, no. 2, pp. 342–355, 2008.

[922] B.-S. Chen, Y.-T. Chang, and Y.-C. Wang, "Robust h-infinity-stabilization design in gene networks under stochastic molecular noises: Fuzzy-interpolation approach," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 38, no. 1, pp. 25–42, 2008.

[923] R. Durrett, *Stochastic calculus: a practical introduction*, vol. 6. CRC press, 1996.

[924] D. Nualart, "Analysis on wiener space and anticipating stochastic calculus," in *Lectures on probability theory and statistics*, pp. 123–220, Springer, 1998.

[925] G. E. P. Box and D. A. Pierce, "Distribution of Residual Autocorrelations in Autoregressive-Integrated Moving Average Time Series Models," *Journal of the American Statistical Association*, vol. 65, pp. 1509–1526, dec 1970.

[926] D. J. Bartholomew, "Time Series Analysis Forecasting and Control," *Journal of the Operational Research Society*, vol. 22, pp. 199–201, jun 1971.

[927] M. D. Ma, C. C. Chang, S. S. Jang, and D. S. H. Wong, "Mixed product run-to-run process control - An ANOVA model with ARIMA disturbance approach," *Journal of Process Control*, vol. 19, no. 4, pp. 604–614, 2009.

[928] H. Dym, "Stationary Measures for the Flow of a Linear Differential Equation Driven by White Noise," *Transactions of the American Mathematical Society*, vol. 123, p. 130, may 1966.

[929] R. E. Mortensen, "Mathematical problems of modeling stochastic nonlinear dynamic systems," *Journal of Statistical Physics*, vol. 1, no. 2, pp. 271–296, 1969.

[930] A. E. Permanasari, D. R. A. Rambli, and P. D. D. Dominic, "Forecasting method selection using ANOVA and Duncan multiple range tests on time series dataset," *Proceedings 2010 International Symposium on Information Technology - Engineering Technology, ITSim'10*, vol. 2, pp. 941–945, 2010.

[931] C. S. Peskin, G. M. Odell, and G. F. Oster, "Cellular motions and thermal fluctuations: the Brownian ratchet," *Biophysical Journal*, vol. 65, pp. 316–324, jul 1993.

[932] L. Wang, D. Jiang, and D. O'Regan, "The periodic solutions of a stochastic chemostat model with periodic washout rate," *Communications in Nonlinear Science and Numerical Simulation*, vol. 37, pp. 1–13, aug 2016.

[933] J. H. Kotecha and P. M. Djurić, "Gaussian sum particle filtering," *IEEE Transactions on Signal Processing*, vol. 51, pp. 2602–2612, oct 2003.

[934] S. Di Talia, J. M. Skotheim, J. M. Bean, E. D. Siggia, and F. R. Cross, "The effects of molecular noise and size control on variability in the budding yeast cell cycle," *Nature*, vol. 448, no. 7156, pp. 947–951, 2007.

[935] B. Lehner, "Conflict between noise and plasticity in yeast," *PLoS genetics*, vol. 6, no. 11, 2010.

[936] A. Llamosi, A. M. Gonzalez-Vargas, C. Versari, E. Cinquemani, G. Ferrari-Trecate, P. Hersen, and G. Batt, "What population reveals about individual cell identity: single-cell parameter estimation of models of gene expression in yeast," *PLoS computational biology*, vol. 12, no. 2, 2016.

[937] A. A. Kolodziejczyk, J. K. Kim, V. Svensson, J. C. Marioni, and S. A. Teichmann, "The technology and biology of single-cell rna sequencing," *Molecular cell*, vol. 58, no. 4, pp. 610–620, 2015.

[938] A. P. Gasch, F. B. Yu, J. Hose, L. E. Escalante, M. Place, R. Bacher, J. Kanbar, D. Ciobanu, L. Sandor, I. V. Grigoriev, C. Kendziorski, S. R. Quake, and M. N. McClean, "Single-cell RNA sequencing reveals intrinsic and extrinsic regulatory heterogeneity in yeast responding to stress," *PLoS Biology*, vol. 15, no. 12, pp. 10–12, 2017.

[939] J. R. Newman, S. Ghaemmaghami, J. Ihmels, D. K. Breslow, M. Noble, J. L. DeRisi, and J. S. Weissman, "Single-cell proteomic analysis of s. cerevisiae reveals the architecture of biological noise," *Nature*, vol. 441, no. 7095, pp. 840–846, 2006.

[940] K. R. Willison and D. R. Klug, "Quantitative single cell and single molecule proteomics for clinical studies," *Current opinion in biotechnology*, vol. 24, no. 4, pp. 745–751, 2013.

[941] H. H. McAdams and A. Arkin, "It'sa noisy business! genetic regulation at the nanomolar scale," *Trends in genetics*, vol. 15, no. 2, pp. 65–69, 1999.

[942] A. J. van Maris, D. A. Abbott, E. Bellissimi, J. van den Brink, M. Kuyper, M. A. Luttik, H. W. Wisselink, W. A. Scheffers, J. P. van Dijken, and J. T. Pronk, "Alcoholic fermentation of carbon sources in biomass hydrolysates by Saccharomyces cerevisiae: Current status," in *Antonie van Leeuwenhoek, International Journal of General and Molecular Microbiology*, vol. 90, pp. 391–418, Kluwer Academic Publishers, nov 2006.

[943] Q. Wang, Y. Zhang, C. Yang, H. Xiong, Y. Lin, J. Yao, H. Li, L. Xie, W. Zhao, Y. Yao, Z. B. Ning, R. Zeng, Y. Xiong, K. L. Guan, S. Zhao, and G. P. Zhao, "Acetylation of metabolic enzymes coordinates carbon source utilization and metabolic flux," *Science*, vol. 327, pp. 1004–1007, feb 2010.

[944] S. H. Lillie and J. R. Pringle, "Reserve carbohydrate metabolism in Saccharomyces cerevisiae: responses to nutrient limitation," *Journal of Bacteriology*, vol. 143, pp. 1384–1394, sep 1980.

[945] P. Daran-Lapujade, M. L. A. Jansen, J.-M. Daran, W. van Gulik, J. H. de Winde, and J. T. Pronk, "Role of Transcriptional Regulation in Controlling Fluxes in Central Carbon Metabolism of Saccharomyces cerevisiae," *Journal of Biological Chemistry*, vol. 279, pp. 9125–9138, mar 2004.

[946] M. S. Esposito, R. E. Esposito, M. Arnaud, and H. O. Halvorson, "Acetate utilization and macromolecular synthesis during sporulation of yeast.," *Journal of Bacteriology*, vol. 100, pp. 180–186, oct 1969.

[947] J. H. Waterborg, "Dynamics of histone acetylation in Saccharomyces cerevisiae," *Biochemistry*, vol. 40, no. 8, pp. 2599–2605, 2001.

[948] C. V. Dang, "MYC on the path to cancer," *Cell*, vol. 149, pp. 22–35, mar 2012.

[949] C. de Virgilio, "The essence of yeast quiescence," *FEMS Microbiology Reviews*, vol. 36, no. 2, pp. 306–339, 2012.

[950] J. R. Broach, "Nutritional control of growth and development in yeast," *Genetics*, vol. 192, pp. 73–105, sep 2012.

[951] B. P. Tu and J. S. Weissman, "Oxidative protein folding in eukaryotes: Mechanisms and consequences," *Journal of Cell Biology*, vol. 164, pp. 341–346, feb 2004.

[952] L. Stiburek and J. Zeman, "Assembly factors and ATP-dependent proteases in cytochrome c oxidase biogenesis," *Biochimica et Biophysica Acta - Bioenergetics*, vol. 1797, pp. 1149–1158, jun 2010.

[953] Y. Pan, "Mitochondria, reactive oxygen species, and chronological aging: A message from yeast," *Experimental Gerontology*, vol. 46, pp. 847–852, nov 2011.

[954] M. Schlackow and M. Gullerova, "Understanding non-coding DNA regions in yeast," *Biochemical Society transactions*, vol. 41, pp. 1654–9, dec 2013.

[955] M. H. Barros, B. Bandy, E. B. Tahara, and A. J. Kowaltowski, "Higher respiratory activity decreases mitochondrial reactive oxygen release and increases lifespan in Saccharomyces cerevisiae," *The Journal of Biological Chemistry*, vol. 279, no. 48, pp. 49883–49888, 2004.

[956] G. Barja, "Mitochondrial Oxygen Consumption and Reactive Oxygen Species Production are Independently Modulated: Implications for Aging Studies," *Rejuvenation Research*, vol. 10, pp. 215–224, jun 2007.

[957] D. K. Bricker, E. B. Taylor, J. C. Schell, T. Orsak, A. Boutron, Y. C. Chen, J. E. Cox, C. M. Cardon, J. G. Van Vranken, N. Dephoure, C. Redin, S. Boudina, S. P. Gygi, M. Brivet, C. S. Thummel, and J. Rutter, "A mitochondrial pyruvate carrier required for pyruvate uptake in yeast, Drosophila, and humans," *Science*, vol. 336, pp. 96–100, jul 2012.

[958] S. M. Schieke, D. Phillips, J. P. McCoy, A. M. Aponte, R. F. Shen, R. S. Balaban, and T. Finkel, "The mammalian target of rapamycin (mTOR) pathway regulates mitochondrial oxygen consumption and oxidative capacity," *Journal of Biological Chemistry*, vol. 281, no. 37, pp. 27643–27652, 2006.

[959] P. Richard, B. M. Bakker, B. Teusink, K. Van Dam, and H. V. Westerhoff, "Acetaldehyde mediates the synchronization of sustained glycolytic oscillations in populations of yeast cells," *European journal of biochemistry*, vol. 235, no. 1-2, pp. 238–241, 1996.

[960] B. P. Tu, *Ultradian metabolic cycles in yeast*, vol. 470. Elsevier Inc, 2 ed., 2010.

[961] J. S. Breunig, S. R. Hackett, J. D. Rabinowitz, and L. Kruglyak, "Genetic Basis of Metabolome Variation in Yeast," *PLoS Genetics*, vol. 10, p. e1004142, mar 2014.

[962] R. F. Voss, "Random Fractal Forgeries," in *Fundamental Algorithms for Computer Graphics*, pp. 805–835, Berlin, Heidelberg: Springer Berlin Heidelberg, 1985.

[963] M. F. Shlesinger, "Fractal Time and 1/f Noise in Complex Systems," *Annals of the New York Academy of Sciences*, vol. 504, pp. 214–228, jul 1987.

[964] T. Hida, H.-H. Kuo, J. J. Potthoff, and L. Streit, *White Noise: An Infinite Dimensional Calculus*. New York: Springer, 1993.

[965] G. Tremberger Jr, T. Holden, E. Cheung, S. Dehipawala, N. Gadura, U. Golebiewska, K. Valentin, M. Smulczeski, W. Satizabal, P. Schneider, *et al.*, "Cyanobacteria gene and protein sequences in diurnal oscillation metabolic processes," in *Instruments, Methods, and Missions for Astrobiology XIII*, vol. 7819, p. 78190U, International Society for Optics and Photonics, 2010.

[966] B. Da, A. Gupta, Y. S. Ong, and L. Feng, "The boon of gene-culture interaction for effective evolutionary multitasking," in *Australasian Conference on Artificial Life and Computational Intelligence*, pp. 54–65, Springer, 2016.

[967] A. V. Oppenheim, R. W. Schafer, and T. G. Stockham, "Nonlinear Filtering of Multiplied and Convolved Signals," *Proceedings of the IEEE*, vol. 56, pp. 1264–1291, sep 1968.

[968] D. Childers and M.-T. Pao, "Complex demodulation for transient wavelet detection and extraction," *IEEE Transactions on Audio and Electroacoustics*, vol. 20, no. 4, pp. 295–308, 1972.

[969] A. Tangirala, S. Shah, and N. Thornhill, "Pscmap: A new tool for plant-wide oscillation detection," *Journal of Process Control*, vol. 15, no. 8, pp. 931–941, 2005.

[970] A. Hymans and J. Lait, "Analysis of a frequency-modulated continuous-wave ranging system," *Proceedings of the IEE Part B: Electronic and Communication Engineering*, vol. 107, no. 34, p. 365, 1960.

[971] G. D. Bergland, "A guided tour of the fast Fourier Transform," *IEEE Spectrum*, vol. 6, pp. 41–52, jul 1969.

[972] S. G. Johnson and M. Frigo, "A modified split-radix FFT with fewer arithmetic operations," *IEEE Transactions on Signal Processing*, vol. 55, no. 1, pp. 111–119, 2007.

[973] A. G. J. Macfarlane, "Perspectives The Development of Frequency-Response Methods in Automatic Control," *IEEE Transactions on Automatic Control*, vol. 24, no. 2, pp. 250–265, 1979.

[974] L. Eyer and P. Bartholdi, "Variable stars: Which Nyquist frequency?," *Astron. Astrophys. Suppl. Ser.*, vol. 135, no. 1, pp. 1–3, 1999.

[975] D. N. Fittinghoff, K. W. DeLong, R. Trebino, and C. L. Ladera, "Noise sensitivity in frequency-resolved optical-gating measurements of ultrashort pulses," *Josa B*, vol. 12, p. 1955, oct 1995.

[976] J. V. Gray, G. a. Petsko, G. C. Johnston, D. Ringe, R. a. Singer, and M. Werner-washburne, "" Sleeping Beauty ": Quiescence in Saccharomyces cerevisiae," *Microbiology and molecular biology reviews*, vol. 68, no. 2, pp. 187–206, 2004.

[977] M. M. Klosinska, C. A. Crutchfield, P. H. Bradley, J. D. Rabinowitz, and J. R. Broach, "Yeast cells can access distinct quiescent states," *Genes and Development*, vol. 25, no. 4, pp. 336–349, 2011.

[978] R. I. Joh, J. S. Khanduja, I. A. Calvo, M. Mistry, C. M. Palmieri, A. J. Savol, S. J. Ho Sui, R. I. Sadreyev, M. J. Aryee, and M. Motamedi, "Survival in Quiescence Requires the Euchromatic Deployment of Clr4/SUV39H by Argonaute-Associated Small RNAs," *Molecular Cell*, vol. 64, pp. 1088–1101, dec 2016.

[979] H. Frickmann, A. E. Zautner, A. Moter, J. Kikhney, R. M. Hagen, H. Stender, and S. Poppert, "Fluorescence in situ hybridization (FISH) in the microbiological diagnostic routine laboratory: a review," *Critical Reviews in Microbiology*, vol. 43, pp. 263–293, may 2017.

[980] H. L. Smith and P. Waltman, *The theory of the chemostat : dynamics of microbial competition*. Cambridge University Press, 1995.

[981] H. Guo and L. Chen, "Periodic solution of a chemostat model with Monod growth rate and impulsive state feedback control," *Journal of Theoretical Biology*, vol. 260, pp. 502–509, oct 2009.

[982] M. Théry and M. Bornens, "Cell shape and cell division," *Current Opinion in Cell Biology*, vol. 18, pp. 648–657, dec 2006.

[983] M. Piel and P. T. Tran, "Cell Shape and Cell Division in Fission Yeast," *Current Biology*, vol. 19, pp. R823–R827, sep 2009.

[984] M. R. Spiegel, "The Dirac delta-function and the summation of fourier series," *Journal of Applied Physics*, vol. 23, pp. 906–909, aug 1952.

[985] O. M. Nikodým, "Dirac's Delta-function," in *The Mathematical Apparatus for Quantum-Theories*, pp. 724–760, Berlin, Heidelberg: Springer Berlin Heidelberg, 1966.

[986] A. Krivoruchko, Y. Zhang, V. Siewers, Y. Chen, and J. Nielsen, "Microbial acetyl-CoA metabolism and metabolic engineering," *Metabolic Engineering*, vol. 28, pp. 28–42, mar 2015.

[987] G. B. Overal, B. Teusink, F. J. Bruggeman, J. Hulshof, and R. Planqué, "Understanding start-up problems in yeast glycolysis," *Mathematical Biosciences*, vol. 299, pp. 117–126, may 2018.

[988] J. Lian, T. Si, N. U. Nair, and H. Zhao, "Design and construction of acetyl-CoA overproducing Saccharomyces cerevisiae strains," in *Food, Pharmaceutical and Bioengineering Division 2014 - Core Programming Area at the 2014 AIChE Annual Meeting*, vol. 2, pp. 750–760, Academic Press, jul 2014.

[989] L. Shi and B. P. Tu, "Acetyl-CoA and the regulation of metabolism: Mechanisms and consequences," *Current Opinion in Cell Biology*, vol. 33, pp. 125–131, 2015.

[990] P. De Jong-Gubbels, M. A. Van Den Berg, H. Y. Steensma, J. P. Van Dijken, and J. T. Pronk, "The Saccharomyces cerevisiae acetyl-coenzyme a synthetase encoded by the a ACS1 gene, but not the a ACS2-encoded enzyme, is subject to glucose catabolite inactivation," *FEMS Microbiology Letters*, vol. 153, pp. 75–81, jan 1997.

[991] J. T. Pronk, H. Y. Steensma, and J. P. Van Dijken, "Pyruvate metabolism in Saccharomyces cerevisiae," *Yeast*, vol. 12, pp. 1607–1633, dec 1996.

[992] S. G. Miran, J. E. Lawson, and L. J. Reed, "Characterization of PDH beta 1, the structural gene for the pyruvate dehydrogenase beta subunit from Saccharomyces cerevisiae.," *Proceedings of the National Academy of Sciences*, vol. 90, pp. 1252–1256, feb 1993.

[993] J. E. Lawson, R. H. Behal, and L. J. Reed, "Disruption and Mutagenesis of the Saccharomyces cerevisiae PDX1 Gene Encoding the Protein X Component of the Pyruvate Dehydrogenase Complex," *Biochemistry*, vol. 30, pp. 2834–2839, mar 1991.

[994] R. Edgar, M. Domrachev, and A. E. Lash, "Gene expression omnibus: Ncbi gene expression and hybridization array data repository," *Nucleic acids research*, vol. 30, no. 1, pp. 207–210, 2002.

[995] T. Barrett and R. Edgar, "Gene expression omnibus: Microarray data storage, submission, retrieval, and analysis," *Methods in enzymology*, vol. 411, pp. 352–369, 2006.

[996] T. Barrett, S. E. Wilhite, P. Ledoux, C. Evangelista, I. F. Kim, M. Tomashevsky, K. A. Marshall, K. H. Phillippy, P. M. Sherman, M. Holko, *et al.*, "Ncbi geo: archive for functional genomics data sets—update," *Nucleic acids research*, vol. 41, no. D1, pp. D991–D995, 2012.

[997] E. Clough and T. Barrett, "The gene expression omnibus database," in *Statistical Genomics*, pp. 93–110, Springer, 2016.

[998] M. Unser, A. Aldroubi, and M. Eden, "B-spline signal processing. ii. efficiency design and applications," *IEEE transactions on signal processing*, vol. 41, no. 2, pp. 834–848, 1993.

[999] K. W. Mulder, X. Wang, C. Escriu, Y. Ito, R. F. Schwarz, J. Gillis, G. Sirokmany, G. Donati, S. Uribe-Lewis, P. Pavlidis, *et al.*, "Diverse epigenetic strategies interact to control epidermal differentiation," *Nature cell biology*, vol. 14, no. 7, p. 753, 2012.

[1000] D. E. Schones, K. Cui, and S. Cuddapah, *Genome-Wide Approaches to Studying Yeast Chromatin Modifications*. Totowa, NJ: Humana Press, 2011.

[1001] D. Park, Y. Lee, G. Bhupindersingh, and V. R. Iyer, "Widespread misinterpretable chip-seq bias in yeast," *PloS one*, vol. 8, pp. e83506–e83506, 12 2013.

[1002] A. Weiner, T. H. S. Hsieh, A. Appleboim, H. V. Chen, A. Rahat, I. Amit, O. J. Rando, and N. Friedman, "High-resolution chromatin dynamics during a yeast stress response," *Molecular Cell*, vol. 58, no. 2, pp. 371–386, 2015.

[1003] G. D. Gilfillan, T. Hughes, Y. Sheng, H. S. Hjorthaug, T. Straub, K. Gervin, J. R. Harris, D. E. Undlien, and R. Lyle, "Limitations and possibilities of low cell number chip-seq," *BMC genomics*, vol. 13, pp. 645–645, 11 2012.

[1004] S. R. Head, H. K. Komori, S. A. LaMere, T. Whisenant, F. Van Nieuwerburgh, D. R. Salomon, and P. Ordoukhanian, "Library construction for next-generation sequencing: overviews and challenges," *Biotechniques*, vol. 56, no. 2, pp. 61–77, 2014.

[1005] J. T. Wade, "Mapping transcription regulatory networks with chip-seq and rna-seq," in *Prokaryotic systems biology*, pp. 119–134, Springer, 2015.

[1006] C. Trapnell, B. A. Williams, G. Pertea, A. Mortazavi, G. Kwan, M. J. Van Baren, S. L. Salzberg, B. J. Wold, and L. Pachter, "Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation," *Nature Biotechnology*, vol. 28, pp. 511–515, may 2010.

[1007] S. Anders and W. Huber, "Differential expression analysis for sequence count data," *Genome Biology*, vol. 11, p. R106, oct 2010.

[1008] B. Langmead, K. D. Hansen, and J. T. Leek, "Cloud-scale RNA-sequencing differential expression analysis with Myrna," *Genome Biology*, vol. 11, p. R83, aug 2010.

[1009] M. J. Clark, R. R. Chen, H. Y. K. Lam, K. J. Karczewski, R. R. Chen, G. Euskirchen, A. J. Butte, and M. Snyder, "Performance comparison of exome DNA sequencing technologies," *Nature Biotechnology*, vol. 29, pp. 908–914, oct 2011.

[1010] S. Jin, H. Fu, Q. Zhou, S. Sun, S. Jiang, Y. Xiong, Y. Gong, H. Qiao, and W. Zhang, "Transcriptome Analysis of Androgenic Gland for Discovery of Novel Genes from the Oriental River Prawn, Macrobrachium nipponense, Using Illumina Hiseq 2000," *PLoS ONE*, vol. 8, p. e76840, oct 2013.

[1011] J. C. Marioni, C. E. Mason, S. M. Mane, M. Stephens, and Y. Gilad, "RNA-seq: An assessment of technical reproducibility and comparison with gene expression arrays," *Genome Research*, vol. 18, pp. 1509–1517, jul 2008.

[1012] J. M. Kim, S. Vanguri, J. D. Boeke, A. Gabriel, and D. F. Voytas, "Transposable elements and genome organization: A comprehensive survey of retrotransposons revealed by the complete Saccharomyces cerevisiae genome sequence," *Genome Research*, vol. 8, pp. 464–478, may 1998.

[1013] E. H. Hurowitz and P. O. Brown, "Genome-wide analysis of mRNA lengths in Saccharomyces cerevisiae," *Genome biology*, vol. 5, no. 1, p. R2, 2003.

[1014] D. M. Gelperin, M. A. White, M. L. Wilkinson, Y. Kon, L. A. Kung, K. J. Wise, N. Lopez-Hoyo, L. Jiang, S. Piccirillo, H. Yu, M. Gerstein, M. E. Dumont, E. M. Phizicky, M. Snyder, and E. J. Grayhack, "Biochemical and genetic analysis of the yeast proteome with a movable ORF collection," *Genes and Development*, vol. 19, pp. 2816–2826, dec 2005.

[1015] L. Peña-Castillo and T. R. Hughes, "Why are there still over 1000 uncharacterized yeast genes?," *Genetics*, vol. 176, pp. 7–14, may 2007.

[1016] W. Fischle, Y. Wang, and C. D. Allis, "Histone and chromatin cross-talk," *Current opinion in cell biology*, vol. 15, no. 2, pp. 172–183, 2003.

[1017] D. Schübeler, D. M. MacAlpine, D. Scalzo, C. Wirbelauer, C. Kooperberg, F. van Leeuwen, D. E. Gottschling, L. P. O'Neill, B. M. Turner, J. Delrow, *et al.*, "The histone modification pattern of active genes revealed through genome-wide chromatin analysis of a higher eukaryote," *Genes & development*, vol. 18, no. 11, pp. 1263–1271, 2004.

[1018] H. Xing, Y. Mo, W. Liao, and M. Q. Zhang, "Genome-wide localization of protein-dna binding and histone modification by a bayesian change-point method with chip-seq data," *PLoS computational biology*, vol. 8, no. 7, 2012.

[1019] J. Feng, T. Liu, B. Qin, Y. Zhang, and X. S. Liu, "Identifying chip-seq enrichment using macs," *Nature protocols*, vol. 7, no. 9, p. 1728, 2012.

[1020] William Bober, C.-T. Tsai, and O. Masory, *Numerical and Analytical Methods With Matlab*, vol. 20093026 of *Computational Mechanics and Applied Analysis*. CRC Press, aug 2009.

[1021] K.-L. Du and M. Swamy, "Simulated annealing," in *Search and Optimization by Metaheuristics*, pp. 29–36, Springer, 2016.

[1022] R. Rebolledo, "Central limit theorems for local martingales," *Zeitschrift fur Wahrscheinlichkeitstheorie und Verwandte Gebiete*, vol. 11, no. 3, pp. 4871–4895, 2017.

[1023] J. Jacod, "On continuous conditional Gaussian martingales and stable convergence in law," *Séminaire de probabilités de Strasbourg XXXI*, vol. 31, pp. 232–246, 1997.

[1024] M. L. Esquivel, "A Conditional Gaussian Martingale Algorithm for Global Optimization," *ICCSA*, vol. 3982, pp. 841–851, 2006.

[1025] S. Kirkpatrick, J. R. Gelatt, C. D. Gelatt, and M. P. Vecchi, "Optimization by simulated annealing," *Science*, vol. 220, p. 671{\textendash}680, may 1983.

[1026] P. J. M. van Laarhoven and E. H. L. Aarts, "Simulated annealing," in *Simulated Annealing: Theory and Applications*, pp. 7–15, Dordrecht: Springer Netherlands, 1987.

[1027] S. A. Goudsmit, "A Time-of-Flight Mass Spectrometer," *Physical review*, vol. 74, pp. 622–623, sep 1948.

[1028] J. A. Hipple, J. A. Thomas, "A Time-of-Flight Mass Spectrometer with varying field," *Physical Review*, vol. 75, pp. 1616–1616, may 1949.

[1029] W. C. Wiley and I. H. McLaren, "Time-of-flight mass spectrometer with improved resolution," *Review of Scientific Instruments*, vol. 26, pp. 1150–1157, dec 1955.

[1030] L. Mondello, P. Q. Tranchida, P. Dugo, and G. Dugo, "Comprehensive two-dimensional gas chromatography-mass spectrometry: A review," *Mass spectrometry reviews*, vol. 27, no. 2, pp. 101–124, 2008.

[1031] A. Staub, J. Schappler, S. Rudaz, and J.-L. Veuthey, "Ce-tof/ms: fundamental concepts, instrumental considerations and applications," *Electrophoresis*, vol. 30, no. 10, pp. 1610–1623, 2009.

[1032] G. T. Booth, I. X. Wang, V. G. Cheung, and J. T. Lis, "Divergence of a conserved elongation factor and transcription regulation in budding and fission yeast," *Genome Research*, vol. 26, pp. 799–811, 2016.

[1033] L. Warfield, S. Ramachandran, T. Baptista, D. Devys, L. Tora, and S. Hahn, "Transcription of Nearly All Yeast RNA Polymerase II- Transcribed Genes Is Dependent on Transcription Article Transcription of Nearly All Yeast RNA Polymerase II-Transcribed Genes Is Dependent on Transcription Factor TFIID," *Molecular Cell*, vol. 68, pp. 1–12, 2017.

[1034] F. Schirmaier and P. Philippsen, "Identification of two genes coding for the translation elongation factor EF-1 alpha of S. cerevisiae.," *The EMBO journal*, vol. 3, no. 13, pp. 3311–5, 1984.

[1035] A. Carr-Schmid, N. Durko, J. Cavallius, W. C. Merrick, and T. G. Kinzy, "Mutations in a GTP-binding motif of eukaryotic elongation factor 1A reduce both translational fidelity and the requirement for nucleotide exchange," *Journal of Biological Chemistry*, vol. 274, pp. 30297–30302, oct 1999.

[1036] A. Murthi, H. H. Shaheen, H.-Y. Huang, M. A. Preston, T.-P. Lai, E. M. Phizicky, and A. K. Hopper, "Regulation of tRNA Bidirectional Nuclear-Cytoplasmic Trafficking in Saccharomyces cerevisiae," *Molecular Biology of the Cell*, vol. 21, pp. 639–649, feb 2010.

[1037] S. Mukherji and A. Van Oudenaarden, "Synthetic biology: understanding biological design from synthetic circuits," *Nature Reviews Genetics*, vol. 10, no. 12, pp. 859–871, 2009.

[1038] S. B. Haase and C. Wittenberg, "Topology and control of the cell-cycle-regulated transcriptional circuitry," *Genetics*, vol. 196, no. 1, pp. 65–90, 2014.

[1039] H. Firczuk, S. Kannambath, J. Pahle, A. Claydon, R. Beynon, J. Duncan, H. Westerhoff, P. Mendes, and J. E. McCarthy, "An in vivo control map for the eukaryotic mrna translation machinery," *Molecular systems biology*, vol. 9, no. 1, 2013.

[1040] T. von der Haar and L. S. Valášek, "mrna translation: fungal variations on a eukaryotic theme," in *Fungal RNA Biology*, pp. 113–134, Springer, 2014.

[1041] C. G. De Boer and T. R. Hughes, "YeTFaSCo: A database of evaluated yeast transcription factor sequence specificities," *Nucleic Acids Research*, vol. 40, no. D1, pp. 169–179, 2012.

[1042] L. E. Maquat, "Nonsense-mediated mRNA decay: Splicing, translation and mRNP dynamics," *Nature Reviews Molecular Cell Biology*, vol. 5, no. 2, pp. 89–99, 2004.

[1043] J. J. Kim and J. J. Kim, "KEM1 is involved in filamentous growth of Saccharomyces cerevisiae," *FEMS Microbiology Letters*, vol. 216, no. 1, pp. 33–38, 2002.

[1044] U. Sheth, "Decapping and Decay of Messenger RNA Occur in Cytoplasmic Processing Bodies," *Science*, vol. 300, no. 5620, pp. 805–808, 2003.

[1045] F. W. Larimer and A. Stevens, "Disruption of the gene XRN1, coding for a 5'-3' exoribonuclease, restricts yeast cell growth," *Gene*, vol. 95, no. 1, pp. 85–90, 1990.

[1046] J. a. A. Solinger, D. Pascolini, and W. D. D. Heyer, "Active-site mutations in the Xrn1p exoribonuclease of Saccharomyces cerevisiae reveal a specific role in meiosis," *Molecular and Cellular Biology*, vol. 19, no. 9, p. 5930, 1999.

[1047] F. Sinturel, D. Bréchemier-baey, M. Kiledjian, C. Condon, and L. Bénard, "Dcs1 is essential for mitochondrial function in yeast," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 109, pp. 8264–8269, 2012.

[1048] R. D. Klausner, T. A. Rouault, and J. B. Harford, "Regulating the fate of mRNA: The control of cellular iron metabolism," *Cell*, vol. 72, pp. 19–28, jan 1993.

[1049] C. Y. A. Chen and A. B. Shyu, "AU-rich elements: characterization and importance in mRNA degradation," *Trends in Biochemical Sciences*, vol. 20, pp. 465–470, nov 1995.

[1050] R. Binder, J. A. Horowitz, J. P. Basilion, D. M. Koeller, R. D. Klausner, and J. B. Harford, "Evidence that the pathway of transferrin receptor mRNA degradation involves an endonucleolytic cleavage within the 3' UTR and does not involve poly(A) tail shortening.," *The EMBO journal*, vol. 13, pp. 1969–80, apr 1994.

[1051] S. Puig, E. Askeland, and D. J. Thiele, "Coordinated remodeling of cellular metabolism during iron deficiency through targeted mRNA degradation," *Cell*, vol. 120, pp. 99–110, jan 2005.

[1052] C. K. Cohen, T. C. Fox, D. F. Garvin, and L. V. Kochian, "The Role of Iron-Deficiency Stress Responses in Stimulating Heavy-Metal Transport in Plants," *Plant Physiology*, vol. 116, pp. 1063–1072, mar 1998.

[1053] J. Feltham and S. Xi, "FeSO4 concentration in the media increases the YMC period from 2h to 3h," 2017. Personal communication concerning unpublished data.

[1054] K. J. Vetter and F. Gorn, "Kinetics of layer formation and corrosion processes of passive iron in acid solutions," *Electrochimica Acta*, vol. 18, no. 4, pp. 321–326, 1973.

[1055] B. S. Boyanov, "Interaction of (NH4)2SO4 and FeSO4 with metal oxides and ferrites," *Thermochimica acta*, vol. 240, pp. 225–233, 1994.

[1056] R. S. Bedwal and A. Bahuguna, "Zinc, copper and selenium in reproduction," *Experientia*, vol. 50, pp. 626–640, jul 1994.

[1057] C. Askwith and J. Kaplan, "Iron and copper transport in yeast and its relevance to human disease," *Trends in biochemical sciencesiochemical Sciences*, vol. 23, pp. 135–138, apr 1998.

[1058] J. C. Rutherford and A. J. Bird, "Metal-Responsive Transcription Factors That Regulate Iron, Zinc, and Copper Homeostasis in Eukaryotic Cells," *Eukaryotic Cell*, vol. 3, pp. 1–13, feb 2004.

[1059] U. Krämer and S. Clemens, *Functions and homeostasis of zinc, copper, and nickel in plants*, vol. 14. Springer, Heidelberg, 2006.

[1060] C. H. Chio, S. K. Sharma, and D. W. Muenow, "The hydrates and deuterates of ferrous sulfate (FeSO4): a Raman spectroscopic study," *Journal of Raman Spectroscopy*, vol. 38, no. April, pp. 88–99, 2007.

[1061] O. S. Chen, R. J. Crisp, M. Valachovic, M. Bard, D. R. Winge, and J. Kaplan, "Transcription of the yeast iron regulon does not respond directly to iron but rather to iron-sulfur cluster biosynthesis," *Journal of Biological Chemistry*, vol. 279, no. 28, pp. 29513–29518, 2004.

[1062] R. Lill and U. Mühlenhoff, "Iron-sulfur-protein biogenesis in eukaryotes," *Trends in Biochemical Sciences*, vol. 30, no. 3, pp. 133–141, 2005.

[1063] D. C. Johnson, D. R. Dean, A. D. Smith, and M. K. Johnson, "Structure, Function, and Formation of Biological Iron-Sulfur Clusters," *Annual Review of Biochemistry*, vol. 74, no. 1, pp. 247–281, 2005.

[1064] S. A. Garland, K. Hoff, L. E. Vickery, and V. C. Culotta, "Saccharomyces cerevisiae ISU1 and ISU2: Members of a well-conserved gene family for iron-sulfur cluster assembly," *Journal of Molecular Biology*, vol. 294, no. 4, pp. 897–907, 1999.

[1065] B. Schilke, C. Voisine, H. Beinert, and E. Craig, "Evidence for a conserved system for iron metabolism in the mitochondria of Saccharomyces cerevisiae.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 96, no. 18, pp. 10206–11, 1999.

[1066] T. A. Rouault and R. D. Klausner, "Iron-sulfur clusters as biosensors of oxidants and iron," *Trends in Biochemical Sciences*, vol. 21, pp. 174–177, may 1996.

[1067] W.-H. Tong, G. N. L. Jameson, B. H. Huynh, and T. A. Rouault, "Subcellular compartmentalization of human Nfu, an iron-sulfur cluster scaffold protein, and its ability to assemble a [4Fe-4S] cluster," *Proceedings of the National Academy of Sciences*, vol. 100, no. 17, pp. 9762–9767, 2003.

[1068] R. H. Houtkooper, C. Cantó, R. J. Wanders, and J. Auwerx, "The secret life of NAD+: An old metabolite controlling new metabolic signaling pathways," *Endocrine Reviews*, vol. 31, no. 2, pp. 194–223, 2010.

[1069] C. Cantó and J. Auwerx, "NAD + as a signaling molecule modulating metabolism," *Critical Reviews in Biochemistry and Molecular Biology*, vol. 48, no. 4, pp. 291–298, 2013.

[1070] Z. Zhang, A. Milias-Argeitis, and M. Heinemann, "Dynamic single-cell NAD(P)H measurement reveals oscillatory metabolism throughout the E. coli cell division cycle," *Scientific Reports*, vol. 8, no. 1, p. 2162, 2018.

[1071] J. M. Cherry, C. Adler, C. Ball, S. A. Chervitz, S. S. Dwight, E. T. Hester, Y. Jia, G. Juvik, T. Roe, M. Schroeder, S. Weng, and D. Botstein, "SGD: Saccharomyces genome database," *Nucleic Acids Research*, vol. 26, pp. 73–79, jan 1998.

[1072] S. Weng, Q. Dong, R. Balakrishnan, K. Christie, M. Costanzo, K. Dolinski, S. S. Dwight, S. Engel, D. G. Fisk, E. Hong, L. Issel-Tarver, A. Sethuraman, C. Theesfeld, R. Andrada, G. Binkley, C. Lane, M. Schroeder, D. Botstein, and J. M. Cherry, "Saccharomyces Genome Database (SGD) provides biochemical and structural information for budding yeast proteins," *Nucleic Acids Research*, vol. 31, pp. 216–218, jan 2003.

[1073] K. R. Christie, "Saccharomyces Genome Database (SGD) provides tools to identify and analyze sequences from Saccharomyces cerevisiae and related sequences from other organisms," *Nucleic Acids Research*, vol. 32, pp. 311D–314, jan 2004.

[1074] J. M. Cherry, "The saccharomyces genome database: A tool for discovery," *Cold Spring Harbor Protocols*, vol. 2015, pp. 1033–1036, dec 2015.

[1075] S. T. Hellerstedt, R. S. Nash, S. Weng, K. M. Paskov, E. D. Wong, K. Karra, S. R. Engel, and J. M. Cherry, "Curated protein information in the Saccharomyces genome database," *Database*, vol. 2017, jan 2017.

[1076] O. W. Lang, R. S. Nash, S. T. Hellerstedt, and S. R. Engel, "An introduction to the saccharomyces genome database (SGD)," in *Methods in Molecular Biology*, vol. 1757, pp. 21–30, Humana Press, New York, NY, 2018.

[1077] A. Singh, N. Kaur, and D. J. Kosman, "The metalloreductase Fre6p in Fe-efflux from the yeast vacuole.," *The Journal of biological chemistry*, vol. 282, pp. 28619–26, sep 2007.

[1078] L. Li and J. Kaplan, "Characterization of two homologous yeast genes that encode mitochondrial iron transporters," *Journal of Biological Chemistry*, vol. 272, pp. 28485–28493, nov 1997.

[1079] F. Foury, A. Pastore, and M. Trincal, "Acidic residues of yeast frataxin have an essential role in Fe-S cluster assembly," *EMBO reports*, vol. 8, pp. 194–9, feb 2007.

[1080] E. Peiter, M. Fischer, K. Sidaway, S. K. Roberts, and D. Sanders, "The Saccharomyces cerevisiae Ca2+channel Cch1pMid1p is essential for tolerance to cold stress and iron toxicity," *FEBS Letters*, vol. 579, pp. 5697–5703, oct 2005.

[1081] J. R. Greene, N. H. Brown, B. J. DiDomenico, J. Kaplan, and D. J. Eide, "The GEF1 gene of Saccharomyces cerevisiae encodes an integral membrane protein; mutations in which have effects on respiration and iron-limited growth," *MGG Molecular & General Genetics*, vol. 241, pp. 542–553, dec 1993.

[1082] L. Batista-Nascimento, M. B. Toledano, D. J. Thiele, and C. Rodrigues-Pousada, "Yeast protective response to arsenate involves the repression of the high affinity iron uptake system," *Biochimica et Biophysica Acta - Molecular Cell Research*, vol. 1833, pp. 997–1005, may 2013.

[1083] R. Miao, M. M. M. Martinho, J. G. Morales, H. Kim, E. A. Ellis, R. Lill, M. P. Hendrich, E. Munck, P. A. Lindahl, E. Muńck, P. A. Lindahl, E. Munck, and P. A. Lindahl, "EPR and Mossbauer Spectroscopy of Intact Mitochondria Isolated from Yah1 Depleted S. cerevisiae," *Biochemistry*, vol. 47, pp. 9888–9899, sep 2008.

[1084] G. Karthikeyan, L. K. Lewis, M. A. Resnick, K. Lewis, M. A. Resnick, L. K. Lewis, and M. A. Resnick, "The mitochondrial protein frataxin prevents nuclear damage," *Human Molecular Genetics*, vol. 11, pp. 1351–1362, may 2002.



Figure 3.7 enlarged - Raw sequencing data for gene YLN122W (*DBP2*) using the IGV display.

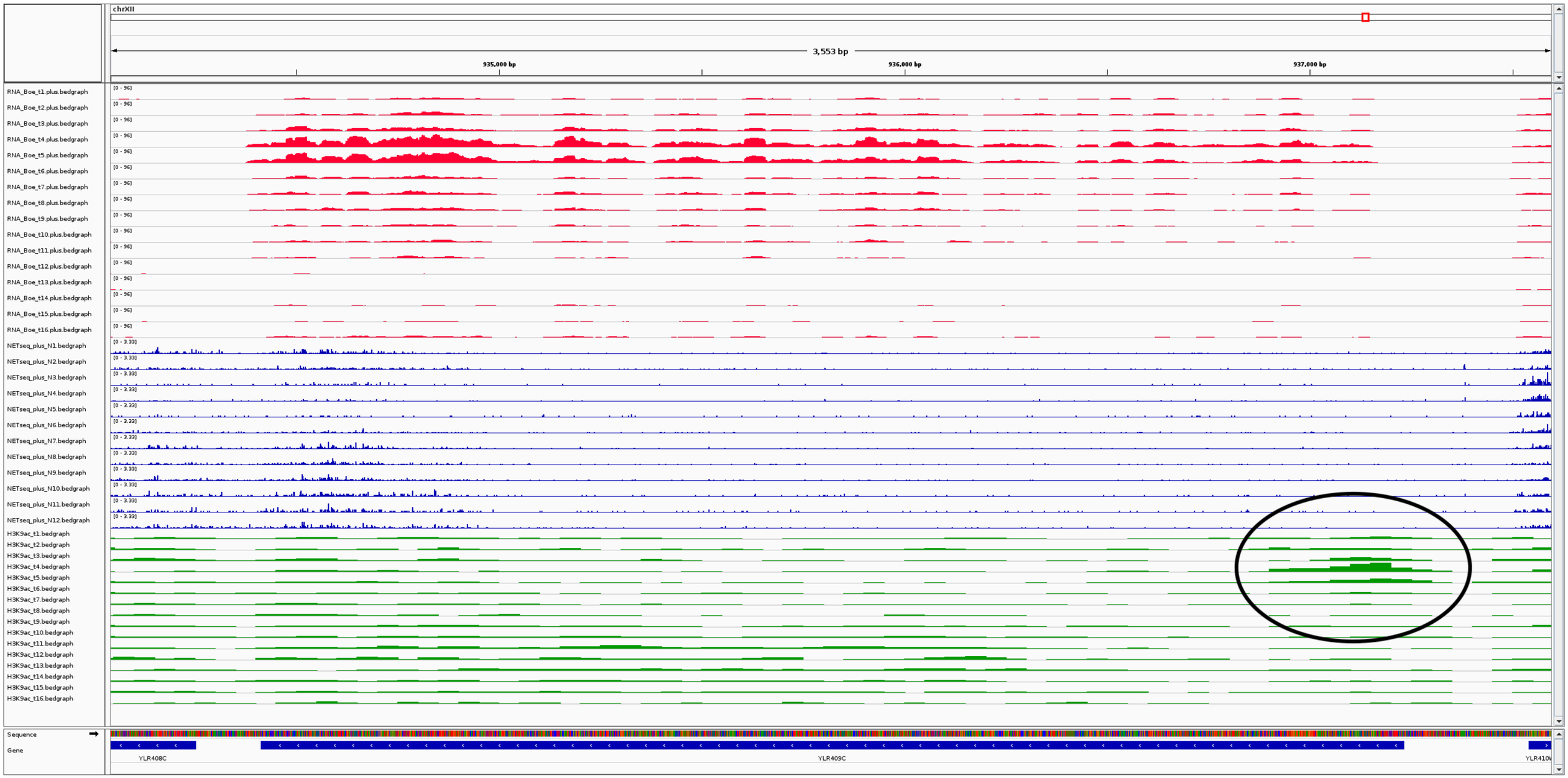


Figure 3.8 enlarged - Raw sequencing data for gene YLR409C (*UTP21*) using the IGV display.

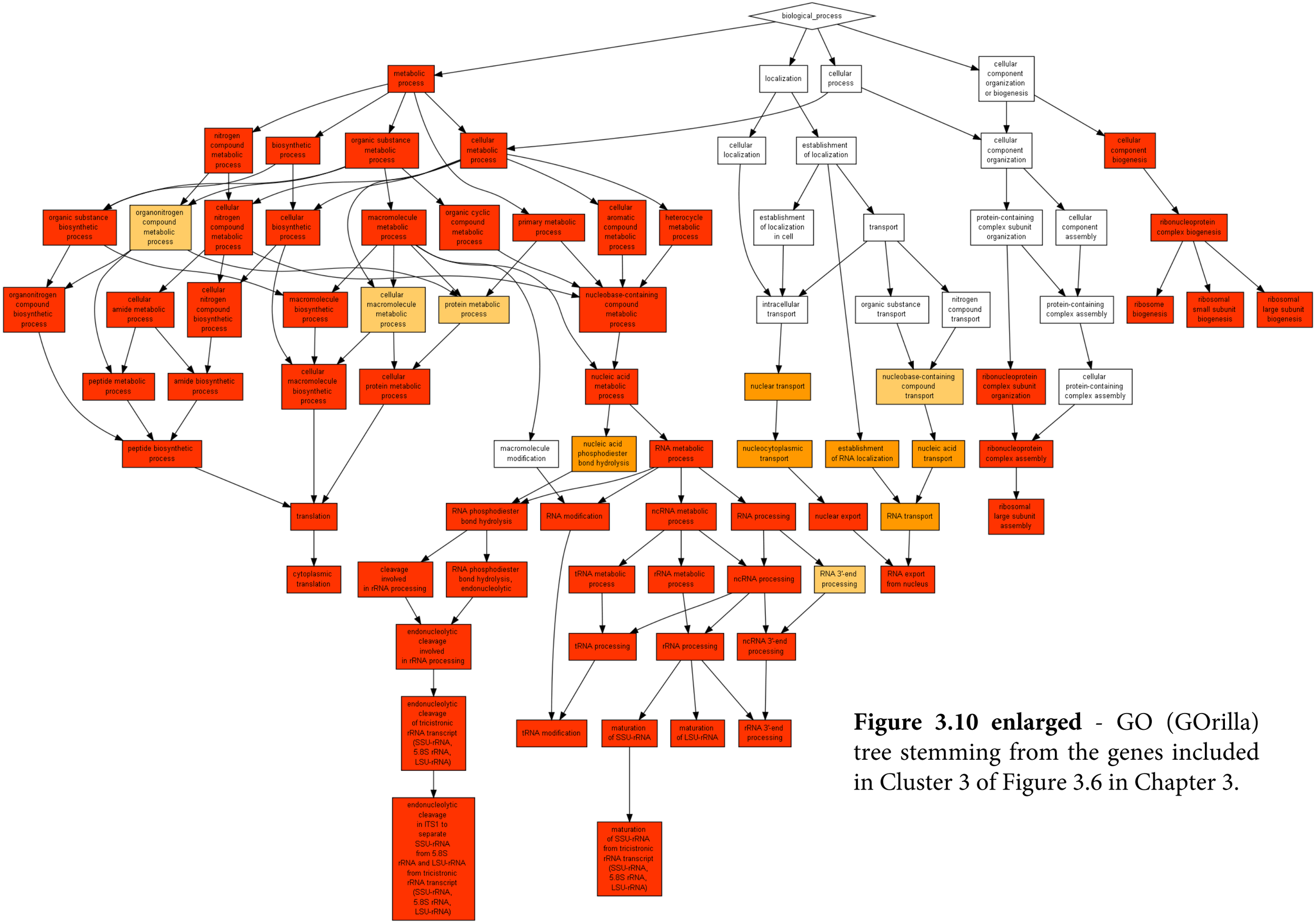


Figure 3.10 enlarged - GO (GOrilla) tree stemming from the genes included in Cluster 3 of Figure 3.6 in Chapter 3.

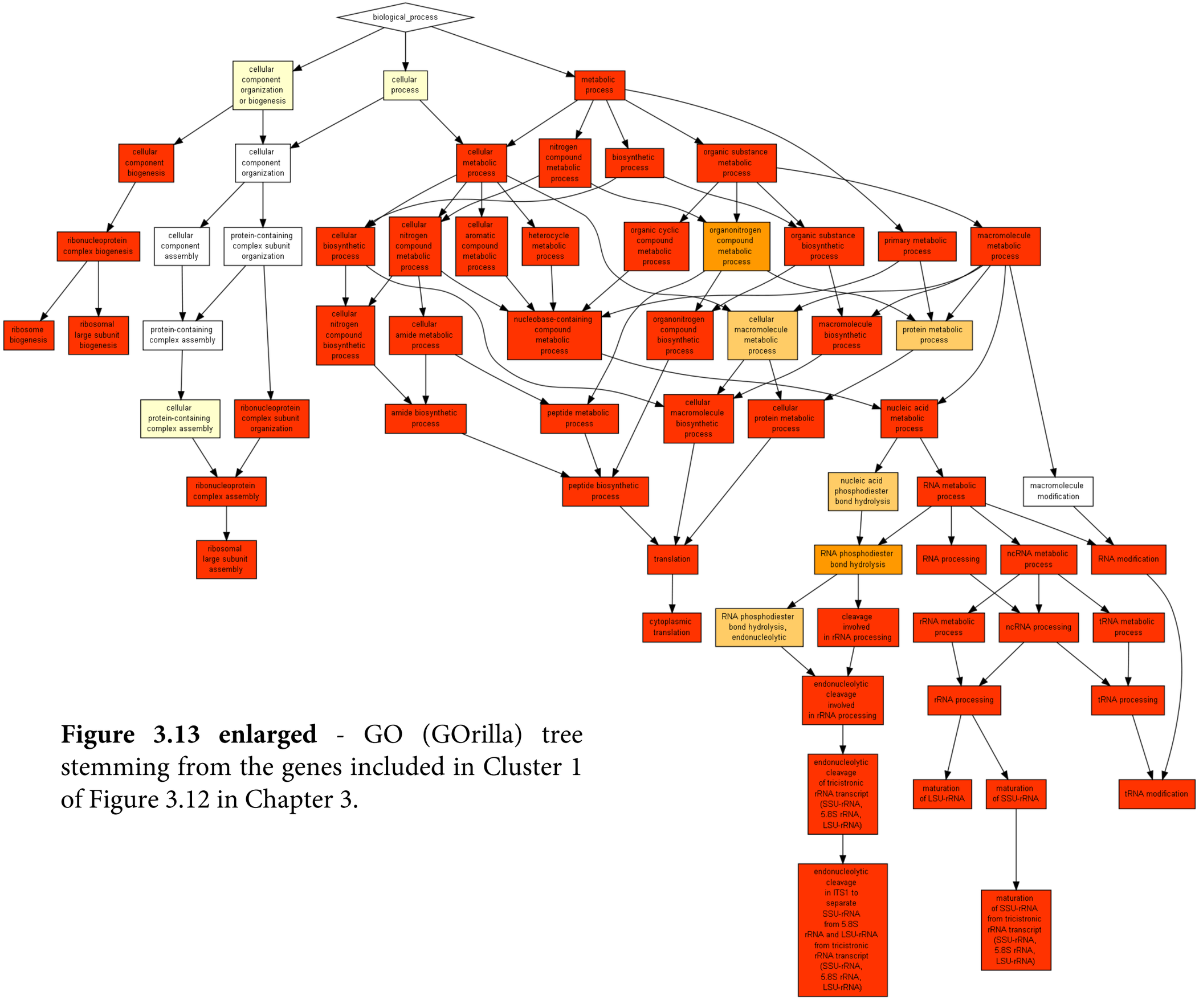


Figure 3.13 enlarged - GO (GOrilla) tree stemming from the genes included in Cluster 1 of Figure 3.12 in Chapter 3.